AN EM ALGORITHM FOR MAXIMUM LIKELIHOOD ESTIMATION OF PROCESS FACTOR ANALYSIS MODELS

Taehun Lee

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology (Quantitative).

Chapel Hill
2010

Approved by

Robert C. MacCallum, Ph.D.

Stephen H. C. du Toit, Ph.D.

David M. Thissen, Ph.D.

Kenneth A. Bollen, Ph.D.

Eric A. Youngstrom, Ph.D.

ABSTRACT

Taehun Lee: AN EM ALGORITHM FOR MAXIMUM LIKELIHOOD ESTIMATION OF PROCESS FACTOR ANALYSIS MODELS
(Under the direction of Robert C. MacCallum and Stephen H. C. du Toit)

In this dissertation, an Expectation-Maximization (EM) algorithm is developed and implemented to obtain maximum likelihood estimates of the parameters and the associated standard error estimates characterizing temporal flows for the latent variable time series following stationary vector ARMA processes, as well as the parameters defining the relationship between the latent stochastic vector and the observed scores taking measurement errors into account. Such models have been known as Process Factor Analysis (PFA) models (Browne & Nesselroade, 2005). In the *E-step*, the complete-data expected log-likelihood, the so-called *Q-function*, which is the joint likelihood of the manifest variables and the latent time series process variables, is constructed by supposing the latent process variables are observed. In the *M-step*, the Newton-Raphson algorithm is employed in order to update the parameter estimates. The closed form expressions for the gradient vector and the Hessian matrix of the target function are derived for implementing the *M-step* of the EM algorithm. Methods for obtaining the associated standard error estimates are developed and implemented. The proposed EM algorithm employs the covariance structure derived by du Toit and Browne (2007) where the influence of the time series prior to the first observation has remained stable and unchanged when the first observations are made. Thus, unlike other conventional structural equation modeling (SEM) software, model implied covariance matrices satisfy the stability condition and are Block-Toeplitz matrices. The proposed algorithm is applied to simulated data in order to ascertain its viability. Specifically, the recovery of the population parameter values of the proposed EM algorithm is studied with simulated data, which is generated so as to follow a PFA model. The performance of the developing method for standard error estimation is evaluated in the simulation study. The results of the simulation study show that the proposed methods for obtaining parameter estimates and the associated standard error estimates for PFA models can be effectively employed both to single-subject time-series analysis and to repeated time-series analysis. Remaining methodological issues for future research are discussed.

ACKNOWLEDGEMENTS

First and foremost, I am especially grateful to my mentor, Dr. Robert, C. MacCallum, whose encouragement, supervision and support enabled me to successfully complete five years of graduate careers in L. L. Thurstone Psychometric Laboratory. I wish to acknowledge all of my committee members for their insightful feedback on my dissertation. It would have been impossible to write this dissertation without your help and guidance.

I especially wish to thank Dr. Stephen, H. C. du Toit, whose encouragement and guidance from the preliminary to the concluding level enabled me to develop an understanding of the subject. I am indebted to him who supported the project with a grant from NIAAA, SBIR Phase 2 grant (5R44AA014999-03).

I am thankful to Dr. Cheongtag Kim, who introduced me to the field of Psychometrics. I have vivid memories of the exciting moments when I first encountered this dynamic, challenging, and exciting field. I also wish to thank student colleagues of L. L. Thurstone Psychometric Laboratory for their intellectual challenge, support, and friendship.

Lastly, I am heartily thankful to my wife, always being there for me. You are a wonderful wife to me and a wonderful mother to our son, Sean Lee.

To my parents, Sanghwan Lee and Sunjin Sohn,

To my son, Sean Lee,

And

To my wife, Sunhee Noh.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In psychological research, it is common for researchers to administer a battery of measures that allow inferences about unobserved latent factors that are presumed to underlie the observed measures. In cross-sectional data analysis, when the central theme of research is to investigate correlational or causal relationships among latent factors, latent variable modeling such as factor analysis, structural equation modeling (SEM), and item response theory (IRT) have been routinely employed and commercial software is readily available to applied researchers.

However, when the central theme of research is to investigate psychological processes of latent factors, (e.g. whether conscientiousness shows any dramatic change over time, how the change in conscientiousness is related to the change in emotional stability, etc.) a new development dealing with a sequence of measurements taking measurement error into account is necessary in terms of model specification and parameter estimation.

A sequence of measurements on multiple variables at equally spaced time intervals is called a multivariate time series. In psychology, multivariate time series data may occur in a context where a moderate number of measurements is obtained on each individual in a large sample (Browne & du Toit, 1991; Browne & Zhang, 2007b) or in the situation where a large number of measurements is made on a single individual (Browne & Zhang, 2007a; Hamaker, Dolan, & Molenaar, 2005). Often times, the latter is called a single time series and the former is called a repeated time series in psychological literature (Browne & Nesselroade, 2005). Since structural equation modeling (SEM) was employed by Jöreskog (1970) to fit a nonstationary first order autoregressive (AR) process model, known as the Simplex model, SEM methodology has been used frequently for modeling time series data in psychology (Browne & Zhang, 2007b).

For example, in the context of a univariate time series observed in a single subject, van Buuren (1997) investigated how the stationary autoregressive moving average (ARMA) model can be formulated and fitted as a structural equation model using a lagged covariance matrix. In a slightly different context where the number of measurements T is intermediate, say, $T = 30$ and the sample size is small, say, $N = 5$, Hamaker, Dolan, and Molenaar (2003) showed how the raw data maximum likelihood method, which is conventionally adopted for modeling incomplete data in SEM software, can be effectively employed to estimate autoregressive and moving average coefficients in a univariate ARMA model. Recently, du Toit and Browne (2007) developed the model specification for the covariance structure of a vector ARMA time series compatible with conventional SEM software. The derived expression is suitable both for the covariance structure of a single-subject time series and for the covariance structure of a repeated time series in a random sample of subjects.

The idea for fitting a (vector) ARMA model in SEM software is as follows. In a repeated time series, where a sequence of observations on k variables at T time points for each of $n = 1, \ldots, N$ subjects, which amounts to $N \times T \times k$ observations in all, are obtained and assumed to follow the same vector ARMA (VARMA) process, the usual sample covariance matrix, S can be calculated and a structural equation modeling program such as LISREL may be used to fit a VARMA covariance structure.

In other situations, where a single time series, a sequence of observations on multiple variables on a single subject at regular intervals over time, is being analyzed, there are two alternative approaches that may be used (du Toit & Browne, 2007). One is to calculate autocovariance matrices and analyze them using structural equation modeling software (Hamaker, Dolan, & Molenaar, 2002; Nesselroade, McArdle, Aggen, & Meyers, 2002; van

Buuren, 1997). The other alternative is to maximize the log-likelihood function based on a single set of repeated measurements. Some structural equation modeling programs providing full information maximum likelihood estimates by considering subjects one at a time can be adapted to estimating a single subject times series model. An alternative is to use the Kalman filter in conjunction with the prediction error decomposition of the log-likelihood function of an ARMA model (Engle & Watson, 1981), but this approach can not be implemented with available SEM software.

An advantage of the aforementioned approaches is that some of them can be readily implemented with available SEM software. Those available SEM programs, however, do not provide facilities for imposing complicated stability constraints on the autoregressive weight matrices and inequality constraints on the moving average weight matrices for identification (Brockwell & Davis, 2002; du Toit & Browne, 2007). More importantly, none of the available SEM software can accommodate the specification of the covariance structure of a VARMA(p, q) times series model with the stationarity condition imposed (du Toit & Browne, 2007) although, in the case of a VARMA(1,1), this might be accomplished in LISREL by imposing complicated constraints on the time series parameters[1]. In other words, when the psychological process has started in the distant past and the parameters have remained constant throughout so that the stability conditions are satisfied, there is no way of implementing this constraint in estimating parameters using available SEM software. In particular, although the autocorrelation approach must assume that a sample of observations is generated by a stationary underlying process, conventional SEM software does not provide an easy way of constructing appropriate model-implied covariance structure matrices of Block-Toepliz form with the stationariy condition imposed.

The autocorrelation approach has an additional danger because the independence assumption required for the autocovariance matrices to follow the Wishart distribution

[1]By personal communication with Dr. Stephen du Toit

is routinely violated. Consequently standard error estimates and test statistics produced by these programs have no theoretical foundation. A least squares estimation method combined with bootstrapping has been proposed (Zhang, 2006) based on the autocorrelation matrix but the resultant estimates do not have the desirable statistical properties of maximum likelihood estimates.

Above all, the current approaches are restricted to modeling univariate time series data of manifest variables disregarding measurement error, which plays a crucial role in psychometric methodology such as factor analysis and latent variable modeling in psychological research. In other words, when the observed scores at time t, say, y_t may be decomposed into the three components of a deterministic trend (μ_t), a latent stochastic component (f_t), and a measurement error (ε_t), and one desires to model the dynamic processes of the latent component (f_t) using a VARMA time series, new developments are necessary in terms of model specifications and parameter estimations. Furthermore, when dynamic processes are assumed to be stationary, to the best of the author's knowledge, no existing SEM software possesses the facility of estimating AR and MA coefficients for the latent variable VARMA models with the nonlinear constraints of the stationarity condition imposed. In addition, as Hamaker et al. (2003) pointed out, the parameter estimates obtained from SEM software based on autocovariance matrix are not true maximum likelihood estimates.

Chapter 2

Objectives

The goal of this dissertation is to develop and implement an Expectation-Maximization (EM) algorithm to obtain maximum likelihood estimates (MLEs) of the AR and MA coefficients characterizing the latent time series following stationary VARMA processes as well as the parameters defining the relationship between the latent stochastic vector and the observed scores taking measurement errors into account. Such models have been known as Process Factor Analysis (PFA) models in psychological research (Browne & Nesselroade, 2005). The closed form expressions of the gradient vector and the Hessian matrix for implementing the *M-step* of the EM algorithm will be derived in such a way that they can be readily implemented in SEM software such as LISREL (Jöreskog & Sörbom, 2006).

In order to ascertain that the proposed EM algorithm and the associated gradient vector and Hessian matrix are suitable to maximize the expected complete data log-likelihood to obtain the MLEs, a simulation study will be conducted with known values of parameters. The proposed algorithm will be implemented in the general scientific computing program, R (R Development Core Team, 2009) and resulting parameter estimates will be compared to the specified population parameter values.

In practice, getting maximum likelihood estimates is not the final end of the model fitting and statistical inference process. One of the early criticisms of the EM approach is that the EM algorithm does not automatically produce an estimate of the covariance matrix of the maximum likelihood estimates. The proposed algorithm is not free from this criticism, either. In this dissertation project, a method for standard error estimation is developed and its performance is evaluated by comparing the estimates and the actual variability of the parameter estimates obtained in the simulation study.

The remainder of this dissertation is organized as follows. First, after a brief review of maximum likelihood estimation and the EM algorithm, the PFA model is introduced with the associated covariance structure originally derived by du Toit and Browne (2007) briefly reproduced. Next, the *E-step* and the *M-step* are described to implement the EM algorithm for obtaining maximum likelihood estimates for PFA models. Details of derivation of the gradient vector and the Hessian matrix are described. Then, a method for standard error estimation is developed. The performances of the proposed EM algorithm and standard error estimation method are then investigated by application to simulated data. Finally, relevant methodological issues and future research plans are discussed.

Chapter 3

Method of maximum likelihood and the EM algorithm

Suppose that random vector, Y has the density function $f(Y|\gamma)$, where $Y = (Y_1, \ldots, Y_n)'$ and $\gamma = (\gamma_1, \ldots, \gamma_q)'$. Given the observed data $y = (y_1, \ldots, y_n)'$ and a statistical model parameterized by a $q \times 1$ unknown vector of values, γ, the likelihood function $L(\gamma|y)$ is any function of γ proportional to $f(y|\gamma)$. The likelihood function provides us with a measure of relative preferences for various parameter values and the maximum likelihood estimate (MLE), denoted by $\hat{\gamma}$, provides a point estimate for a vector of parameters of interest that makes the observed data most likely.

Under certain regularity conditions, maximum likelihood estimators possess so-called asymptotically optimal properties (Bickel & Doksum, 2001; van der Vaart, 1998; Lehmann & Casella, 1998). That is, as sample size approaches infinity, the bias of MLE tends to zero (asymptotically unbiased), and the variance of the MLE tends to the inverse Fisher information, achieveing the Cramér-Rao lower bound (asymptotic efficiency). In addition, as the sample size tends to infinity, the distribution of MLE converges to the normal distribution with mean equal to the true value of the parameter and covariance matrix equal to the inverse of the Fisher-information matrix . This is practically a very important result because we may treat $\hat{\gamma}$ as a normal variate with mean γ and with covariance that can be calculated from a knowledge of the postulated density $f(Y|\gamma)$.

In most situations where exponential families are involved, it is more convenient to work with the natural logarithm of $L(\gamma|y)$, which is termed the log-likelihood function and denoted by $\ell(\gamma|y)$. Then, for a given sample, $y = (y_1, \ldots, y_n)'$, the MLE of γ, is the value of γ that maximizes the log-likelihood function, $\ell(\gamma|y)$. If $\ell(\gamma|y)$ has partial derivatives with respect to $\gamma_1, \ldots, \gamma_q$, then $\hat{\gamma}$ can be obtained by solving the likelihood equations,

$$\frac{\partial \ell(\gamma|y)}{\partial \gamma_r} = 0, \qquad\qquad r = 1, \ldots, q \tag{3.1}$$

In general situations, one may use an iterative numerical method to locate a mode of the likelihood function, *i.e.* MLE. Let $\gamma^{(0)}$ denote an initial guess of $\hat{\gamma}$ and let $\gamma^{(i)}$ denote the guess at the $\imath^{th}$ iteration. Then, one such algorithm, the Newton-Raphson algorithm, is given as

$$\gamma^{(\imath+1)} = \gamma^{(\imath)} + \left[-\frac{\partial^2 \ell(\gamma|Y)}{\partial \gamma\, \partial \gamma'} \right]^{-1}_{\gamma=\gamma^{(\imath)}} \left[\frac{\partial \ell(\gamma|Y)}{\partial \gamma} \right]_{\gamma=\gamma^{(\imath)}} \tag{3.2}$$

The vector of first-order partial derivatives of a function is called the gradient vector and the symmetric matrix of second-order partial derivatives of a function is called the Hessian matrix. In the present setting, the gradient vector and the Hessian matrix are given as $\frac{\partial \ell(\gamma|Y)}{\partial \gamma}$ and $\frac{\partial^2 \ell(\gamma|Y)}{\partial \gamma\, \partial \gamma'}$, respectively. If the log-likelihood function is quadratic, this algorithm converges in one iteration from any starting value. For general functions, this method gives quadratic convergence near the minimum. As can be seen in equation (3.2), the Newton-Raphson algorithm requires the gradient vector and the Hessian matrix of the log-likelihood function, which may not be a trivial task in highly parameterized models. However, when converged, this algorithm provides the asymptotic variance-covariance matrix associated with the parameter estimates as its natural by-product. The expectation of the negative Hessian matrix is called the Fisher information matrix and the asymptotic variance-covariance matrix is given by the inverse of the Fisher information matrix evaluated at MLE *i.e.* $-\mathrm{E}\left[\frac{\partial^2 \ell(\gamma|Y)}{\partial \gamma\, \partial \gamma'}\right]^{-1}_{\gamma=\hat{\gamma}}$. The Fisher-Scoring algorithm, an alternative approach to the Newton-Raphson algorithm, replaces the Hessian matrix by its expectation, which is given by

$$\gamma^{(\imath+1)} = \gamma^{(\imath)} + \left[-\mathrm{E}\left(\frac{\partial^2 \ell(\gamma|Y)}{\partial \gamma\, \partial \gamma'} \right) \right]^{-1}_{\gamma=\gamma^{(\imath)}} \left[\frac{\partial \ell(\gamma|Y)}{\partial \gamma} \right]_{\gamma=\gamma^{(\imath)}} \tag{3.3}$$

The Newton-Raphson and Fisher-Scoring algorithms have been important tools for finding MLE in the context of factor analysis and structural equation modeling (Jöreskog, 1966, 1967, 1969, 1971; Bock & Bargmann, 1966; du Toit & du Toit, 2008).

3.1 The EM algorithm as a method to obtain MLE

The Newton-Raphson algorithm is a calculus-based method that can be employed for finding the zeroes (or roots) of a real-valued function in general. On the other hand, the EM algorithm is a statistically motivated algorithm where missing data are involved and the analysis of the likelihood function based on the observed data is somewhat complicated. The notion of 'missingness' does not have to involve actual missing data but any incomplete information. For example, unobservable latent variables can be treated as missing (Dempster, Laird, & Rubin, 1977).

The EM algorithm is a type of data augmentation algorithm in which one augments the observed data Y with the unobserved *missing* or *latent data* Z. The augmented data $X = (Y, Z)$ are called the *complete data*, while $p(X|\gamma)$, the associated likelihood function of X, is termed the complete data likelihood function. Specifically, while the observed data likelihood function $p(Y|\gamma)$ is difficult to maximize with respect to γ, the augmented or the complete data likelihood function $p(Y, Z|\gamma)$ is simple to maximize with respect to γ. The EM algorithm makes use of this simplicity in maximizing the observed data likelihood function (Schafer, 1997).

Dempster et al. (1977) in their seminal paper showed that the EM algorithm converges to a (local) maximum of the observed data likelihood function without explicitly manipulating it. The EM algorithm, since first proposed by Dempster et al. (1977) has attracted a great deal of interest and stimulated a considerable amount of research. In psychometric literature, Rubin and Thayer (1982) applied the EM algorithm to the exploratory factor analysis model and Bock and Aitkin (1981) to the item response theory model. Lee and Poon (1998) and Liang and Bentler (2004) applied the EM algorithm to the multilevel structural equation model.

The EM algorithm consists of two steps called the expectation step (*E-step*) and the maximization step (*M-step*). In the *E-step*, the conditional expectations of the missing data are computed given the observed data and the current estimates of the parameters. These expected values are then substituted for the missing data and used to complete the data. In the *M-step*, maximum likelihood estimation of the parameters is performed in the usual manner using the completed data. The estimated parameters are then used to reestimate the missing data, which in turn lead to new parameter estimates. These two steps define an iterative procedure, which is repeated until convergence is achieved.

Formally stated, the EM algorithm starts with an initial guess of the parameters, $\gamma^{(0)}$ and alternates the following two steps at $\ell = 0, 1, 2, \ldots$: The E-step and the M-step. In the most general setting, the E-step computes the target function

$$Q(\gamma|\gamma^{(\ell)}) = \int \ln p(Y, Z|\gamma) p(Z|\gamma^{(\ell)}, Y) \, \mathrm{d}Z \tag{3.4}$$

i.e. the expectation of $\ln p(Y, Z|\gamma)$ with respect to $p(Z|\gamma^{(\ell)}, Y)$. In the M-step, the Q function is maximized with respect to γ to obtain $\gamma^{(\ell+1)}$. The algorithm is iterated until $\|\gamma^{(\ell+1)} - \gamma^{(\ell)}\|$ or $|Q(\gamma^{(\ell+1)}|\gamma^{(\ell)}) - Q(\gamma^{(\ell)}|\gamma^{(\ell)})|$ is sufficiently small.

3.2 The monotonicity of the EM algorithm

One of the most important properties of the EM algorithm, known as monotonicity, is that it always increases the observed likelihood:

$$p(Y|\gamma^{(\ell+1)}) \geq p(Y|\gamma^{(\ell)}) \tag{3.5}$$

The monotonicity of the EM algorithm in (3.5) can be shown as follows. Notice that $p(Y, Z|\gamma) = p(Y|\gamma)p(Z|, Y\gamma)$, which implies

$$-\ln p(Y|\gamma) = -\ln p(Y, Z|\gamma) + \ln p(Z|Y, \gamma) \tag{3.6}$$

By taking the conditional expectation of (3.6) with respect to the density $p(Z|Y, \gamma^{(\ell)})$ yields

$$h(\gamma|\gamma^{(\ell)}) = Q(\gamma|\gamma^{(\ell)}) - \ln p(Y|\gamma) \tag{3.7}$$

where $h(\gamma|\gamma^{(\ell)}) = \int \ln p(Z|Y, \gamma)p(Z|Y, \gamma^{(\ell)})\,\mathrm{d}Z$. Then, by the Kullback-Leibler information inequality, it can be shown that

$$h(\gamma|\gamma^{(\ell)}) \leq h(\gamma^{(\ell)}|\gamma^{(\ell)}) \tag{3.8}$$

Substituting (3.7) into (3.8) and rearranging the terms, we have

$$\ln p(Y|\gamma^{(\ell+1)}) - \ln p(Y|\gamma^{(\ell)}) \geq Q(\gamma^{(\ell+1)}|\gamma^{(\ell)}) - Q(\gamma^{(\ell)}|\gamma^{(\ell)}) \geq 0 \tag{3.9}$$

By definition of $\gamma^{(\ell+1)}$ as the maximizer of $Q(\gamma|\gamma^{(\ell)})$, the last inequality of (3.9) holds. Thus, the monotinicity of the EM algorithm holds. In other words, the EM algorithm guarantees the increment of the observed likelihood at each iteration.

Chapter 4

Introduction to Process Factor Analysis

The process factor analysis (PFA) model is a type of dynamic factor model. There are various definitions of a dynamic process but one that is adopted in this dissertation is "any natural process in which each succeeding state is a function of the preceding states plus a non-forecastable change" (Browne & Zhang, 2007b). Dynamic factor analysis models include variables representing the forces that cause change (random-shock variables), variables on which the change (outcomes) is actually manifested (process variables), and parameters that define, at least to some extent, a temporal flow in the relationships among and between the two kinds of variables (autoregressive and moving average coefficients) (Browne & Nesselroade, 2005). The random variable providing the dynamic force that changes the process variable is the so-called "white noise" variable or "random shock" variable. The values of random shock are unpredictable and produce sudden changes in the process variables (Browne & Nesselroade, 2005).

One notable characteristic of PFA models is that the process variables are treated as latent variables measured with manifest indicators. For example, in a PFA model, *state of food deprivation*, treated as an unobserved variable and assumed to be measurable by observable indicators such as blood sugar level and self-reported feelings of hunger, can be represented as a common factor that drives the two indicators (Nesselroade et al., 2002, p.245). Thus, in PFA models, process variables are latent common factors.

4.1 The model specification

In PFA models, the manifest variables, Y_t are assumed to satisfy a factor analysis model

$$Y_t = \mu_t + \lambda_t \eta_t + \varepsilon_t, \qquad t = 1, \ldots, T \tag{4.1}$$

where Y_t represents a $k \times 1$ random vector of manifest variables at time t, μ_t is a $k \times 1$ mean vector, which can be constant, $\mu_t = \mu$ or may vary systematically with time, η_t is an $m \times 1$ random vector representing latent common factors at time t, λ_t is a constant $k \times m$ factor loading matrix at time t and ε_t is a $k \times 1$ random vector representing unique factors at time t. Unique factors are assumed to be independent of η_t for all t. Common factors, η_t are regarded as process variables and they are assumed to follow a VARMA(p, q) process, vector autoregressive of order p and moving average of order q process of the form:

$$\eta_t = A_1\eta_{t-1} + A_2\eta_{t-2} + \cdots + A_p\eta_{t-p} + z_t + B_1 z_{t-1} + B_2 z_{t-2} + \cdots + B_q z_{t-q} \tag{4.2}$$

where $\mathrm{E}(z_t) = 0$ and $\mathrm{cov}(z_t, z_t') = \Psi$. The elements of z_t are m random shocks that drive change in the m common factors.

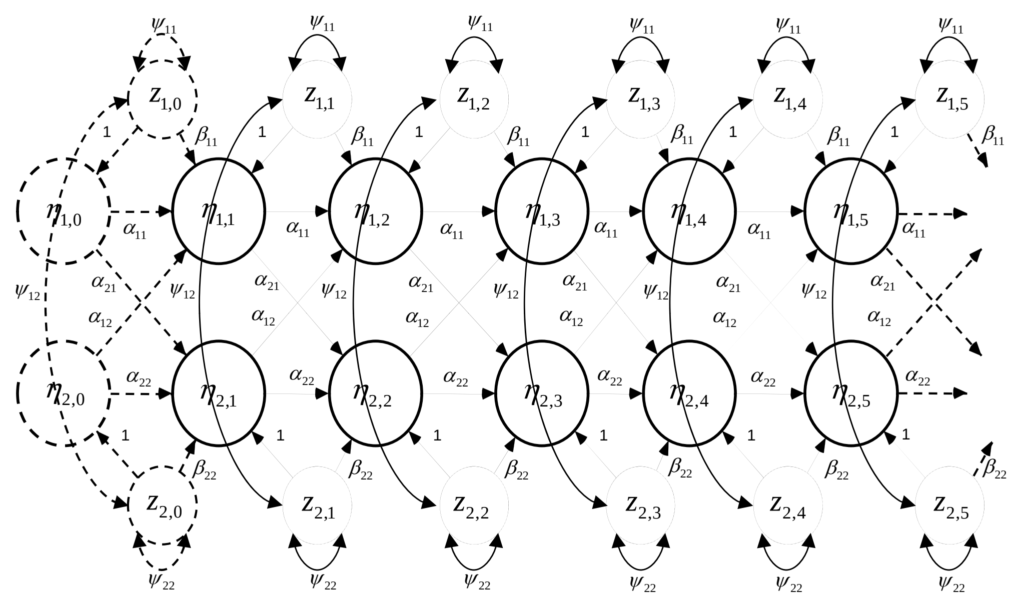
The PFA model is given by Equation (4.1) and (4.2). In particular, (4.1) shows that the principal role of the observed variables is that of the indicator variables for the common factors of the form represented by Equation (4.2), which is one of major differences from previous attempts to specify and fit (vector) ARMA models where the process variables were manifest variables rather than latent variables.

Figure 4.1 shows a path diagram for a vector of two common factors following a stationary VARMA(1,1) time series, which can be given in the following scalar form equations:

$$\eta_{1,t} = \alpha_{11}\eta_{1,t-1} + \alpha_{12}\eta_{2,t-1} + z_{1,t} + \beta_{11} z_{1,t-1}$$

$$\eta_{2,t} = \alpha_{21}\eta_{1,t-1} + \alpha_{22}\eta_{2,t-1} + z_{2,t} + \beta_{22} z_{2,t-1}$$

Figure 4.1: VARMA(1,1) Time Series, $m = 2, T = 5$



where $t = 1, 2, \ldots, 5$ and $\Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix}$. Dashed circles represent the unseen process common factors and random shock variables before Time 1. Dashed arrows are used to represent that the VARMA$(1, 1)$ process has started before Time 1 and may continue after Time 5.

In most general settings, the random vectors Y_t and η_t can follow any distributions other than normal but, in this dissertation, only normally distributed Y_t and η_t will be discussed. The primary purpose of such restriction is to prove the point that the proposed EM algorithm can be employed to yield MLE for PFA models. Theoretically, it is indeed a trivial extension to incorporate non-normal distributions in model specifications. Issues related to the extensions beyond normality, e.g. numerical integration, will be briefly discussed in the final chapter and will be further investigated in the future research.

4.2 Relationships to other dynamic factor analysis models

There is another type of dynamic factor analysis model known as shock factor analysis (SFA) models (Geweke & Singleton, 1981; Molenaar, 1985; Browne & Nesselroade, 2005), also known as white noise factor score (WNFS) models (Nesselroade et al., 2002). In these models, the process variables are no longer latent and they are manifested on the observed variables. Thus, the observed variables such as blood sugar level and self-reported hunger level are represented as the process variables, whose changes over time are driven by the current and earlier random shocks or white noise. The model specification of a SFA model is given as

$$Y_t = \mu_t + \Lambda_0 z_t + \Lambda_1 z_{t-1} + \cdots + \Lambda_q z_{t-q} + \varepsilon_t \tag{4.3}$$

where Y_t, μ_t, ε_t, and z_t are defined as before. In this model, the process variables are Y_t and the z_t, possibly regarded as underlying factors, are actually unpredictable random shocks or white noise variables, being identically and independently distributed between any two time periods.

The choice between which model to fit should be made based on substantive consideration (Nesselroade et al., 2002, p.254). For example, if the underlying common factors, such as state of food deprivation, are regarded as unpredictable shocks to the system of manifest variables such as blood sugar or self-reported hunger level, a SFA model can reflect such substantive considerations more adequately. In contrast, if the state of food deprivation is considered to be relatively stable and predictable by earlier state with changes being caused by unobserved, unpredictable shocks acting on the factors, with measures of blood sugar and self-reported hunger levels being multiple indicators of the state of food deprivation, a PFA model can be a more accurate representation of such substantive considerations.

Mathematical relationships between PFA models and SFA models are explained in

Browne and Nesselroade (2005, p.447) and Nesselroade et al. (2002, p.254). In this dissertation, the procedures for obtaining MLEs for only PFA models are to be considered. The estimation of SFA or WNFS models will be investigated in the future.

4.3 The covariance structure of the latent stationary VARMA(p, q) model

Now the goal of this dissertation can be restated as follows: given the observed scores on the vector of manifest variables, $y_1, \ldots, y_T$, our problem is to find the maximum likelihood estimates of the unknown parameters, μ_t, λ_t, A_t, and B_t, in equations (4.1) and (4.2). The equations in (4.1) and (4.2) are often referred to as the *data model* in psychometric literature. It is clear that equations (4.1) and (4.2) represent a severely under-identified system of equations because the number of known quantities in y_t is far out-numbered by the number of unknown quantities so that it is not possible to estimate all of these values simultaneously.

However, the data model implies a testable model for the population covariance matrix of the manifest variables. Such a model-implied covariance matrix is known as a *covariance structure* in psychometric literature and this covariance structure can be derived from the data model and reflects a hypotheses concerning the variances and covariances among $y_1, \ldots, y_T$. A motivation of deriving a covariance structure from the data model is in the reduced number of unknown quantities to be estimated by modeling variance and covariances among the common factors (η_t), unique factors (u_t), and random shock variables (μ_t), instead of estimating their actual scores by fitting the data model. The closed form expression for the covariance structure of a random vector following a VARMA(p, q) times series was derived by du Toit and Browne (2007).

In this section, steps for deriving the covariance structure of the latent common factors following a VARMA (p, q) time series will be briefly reproduced using the same notation used by du Toit and Browne (2007). Then, based on the model-implied covariance matrix, the complete-data log-likelihood function will be derived, whereby the EM algorithm begins.

First, consider an infinite VARMA(p, q) Gaussian time series:

$$\eta_t = \sum_{i=1}^{p} A_i \eta_{t-i} + u_t + \sum_{j=1}^{q} B_j u_{t-j}, \qquad t = 0, \pm 1, \pm 2, \ldots \tag{4.4}$$

where the $m \times 1$ vector variate η_t are the process vectors and the u_t are the $m \times 1$ random shock vectors mutually independently distributed as

$$u_t \sim \mathcal{N}_m(0, \Psi) \tag{4.5}$$

where Ψ is a $m \times m$ positive definite random shock covariance matrix, $A_1, \ldots, A_p$ are $m \times m$ autoregression weight matrices and $B_1, \ldots B_q$ are $m \times m$ moving average weight matrices. Here the process vector, η_t represents latent variables underlying the manifest variables. Then the covariance structure of η_t following VARMA(p, q) with $t = 1, \ldots, T$ can be derived as follows: First, let $s = \max(p, q)$ and also let $I_{T|s}$ be a matrix formed by the first $m \times s$ columns of the $(m \times T) \times (m \times T)$ identity matrix. Then, the random vector $\eta = (\eta_1, \ldots, \eta_T)'$ can be written as

$$\eta = \mathrm{T}_{-\mathrm{A}}^{-1}(I_{T|s}\mathrm{x}_1 + \mathrm{T}_{\mathrm{B}}\mathrm{u}) \tag{4.6}$$

where

$$\mathrm{x}_1 = \begin{pmatrix} \mathbf{x}_{11} \\ \mathbf{x}_{21} \\ \vdots \\ \mathbf{x}_{s1} \end{pmatrix} = \begin{pmatrix} A_s^{[1]} & \cdots & A_2^{[1]} & A_1^{[1]} \\ & A_s^{[2]} & \cdots & A_2^{[2]} \\ & & \ddots & \vdots \\ & & & A_s^{[s]} \end{pmatrix} \begin{pmatrix} \eta_{-(s-1)} \\ \vdots \\ \eta_{-1} \\ \eta_0 \end{pmatrix}$$

$$+ \begin{pmatrix} B_s^{[1]} & \cdots & B_2^{[1]} & B_1^{[1]} \\ & B_s^{[2]} & \cdots & B_2^{[2]} \\ & & \ddots & \vdots \\ & & & B_s^{[s]} \end{pmatrix} \begin{pmatrix} u_{-(s-1)} \\ \vdots \\ u_{-1} \\ u_0 \end{pmatrix} \tag{4.7}$$

$$
\mathrm{T_{-A}} = \begin{pmatrix} I_m & 0 & & \cdots & & 0 \\ -A_1 & I_m & \ddots & & & \\ \vdots & \ddots & \ddots & 0 & & \vdots \\ -A_s & & -A_1 & I_m & \ddots & \\ 0 & \ddots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & -A_s & & -A_1 & I_m \end{pmatrix} \tag{4.8}
$$

$$
\mathrm{T_{B}} = \begin{pmatrix} I_m & 0 & & \cdots & & 0 \\ B_1 & I_m & \ddots & & & \\ \vdots & \ddots & \ddots & 0 & & \vdots \\ B_s & & B_1 & I_m & \ddots & \\ 0 & \ddots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & B_s & & B_1 & I_m \end{pmatrix} \tag{4.9}
$$

Then, the covariance structure of η, $\Sigma(\tau)$ is given by

$$
\mathrm{cov}\,(\eta, \eta') = \mathrm{T}_{-A}^{-1} \left(I_{T|s} \Theta I'_{T|s} + \mathrm{T}_B (I_T \otimes \Psi) \mathrm{T}'_B \right) \mathrm{T}_{-A}^{-1'} \tag{4.10}
$$

where $\Theta = \mathrm{cov}(\mathrm{x}_1, \mathrm{x}'_1)$. In the case where the stationary VARMA process is assumed, there is no need of superscripts in (4.7). Moreover, du Toit and Bronwe (2007) showed that Θ is a function of autoregressive and moving average parameters, A_i's and B_j's . More specifically, they showed that

$$
\mathrm{vec}(\Theta) = (I - A \otimes A)^{-1} \mathrm{vec}(G \Psi G') \tag{4.11}
$$

where

$$A = \begin{pmatrix} A_1 & I_m & 0 & & 0 \\ A_2 & 0 & I_m & & \\ \vdots & & & \ddots & \\ A_{s-1} & 0 & 0 & & I_m \\ A_s & 0 & 0 & & 0 \end{pmatrix} \tag{4.12}$$

$$G = \begin{pmatrix} A_1 + B_1 \\ A_2 + B_2 \\ A_3 + B_3 \\ \vdots \\ A_s + B_s \end{pmatrix} \tag{4.13}$$

Then, the covariance structure of a random vector following a stationary VARMA(p, q) process is given by substituting (4.11) into (4.10), resulting in the form of a Block-Toeplitz matrix. This covariance structure of the latent process variables will be the building blocks of the EM algorithm developed in the next chapter for maximum likelihood estimation for PFA models.

Chapter 5

Maximum likelihood estimation for the PFA model by the EM algorithm

The EM algorithm is a method for finding maximum likelihood estimates of parameters by alternating two steps, that is, the *E-step* and the *M-step*. In the *E-step*, an expectation of the complete-data log-likelihood is computed given the observed data and the current estimates of the parameters. In the *M-step*, the parameter estimates are updated by maximizing the expected complete-data log-likelihood. Thus, it is clear that the EM algorithm begins with the complete-data log-likelihood function. In the next section, the construction of the complete-data log-likelihood function for PFA(p, q) models and its expectation for completing the *E-step* will be described. And then, the details of the *M-step* will be followed.

5.1 The general expression

Let $Y = (Y_1, Y_2, \ldots, Y_T)'$ and $\eta = (\eta_1, \eta_2, \ldots, \eta_T)'$ be random vectors of the observed variables and the latent process variables, respectively. By supposing η is observed, the complete data log-likelihood function, $\ln p(Y, \eta|\xi)$ can be written as

$$\ln p(Y, \eta|\xi) = \ln g(Y|\eta, \xi_1) + \ln h(\eta|\xi_2) \tag{5.1}$$

where ξ_1 and ξ_2 represent the collection of parameters governing the generation of the observed data and the latent time series data, respectively. The separation of the complete data log-likelihood into two terms in this way will play a crucial role in the context of PFA modeling because the two terms on the right hand side represent the general expressions of the measurement model log-likelihood and the latent time-series model log-likelihood function, immediately suggesting that the current algorithm can readily incorporate various types of measurement models well developed in item response theory (IRT) as well as the general time series models by treating η as observed.

The target function can be obtained by taking conditional expectation of the unknown η given the observed data, Y and the current estimates of the parameters, $\xi^{(\ell)}$. That is,

$$Q(\xi|\xi^{(\ell)}) = \int \Big\{ \ln g(Y|\eta, \xi_1) + \ln h(\eta|\xi_2) \Big\} p(\eta|Y, \xi^{(\ell)})\, \mathrm{d}\eta \tag{5.2}$$

In most general settings, $\ln g(Y|\eta, \xi_1)$ and $\ln h(\eta|\xi_2)$ can be assumed to take any distributional forms, but in the next section, only normally distributed Y and η are considered for the purpose of proving the feasibility of the proposed algorithm.

5.2 The conditional likelihood function of Y given η

Let Y_t be a $k \times 1$ normally distributed random vector for manifest variables or indicator variables at time t and η_t be a $m \times 1$ normally distributed random vector for unobserved latent common factors at time t. And let λ_t be a constant $k \times m$ factor loading matrix at time t, where m is smaller than k. Then, the conditional distribution of $Y = (Y_1, Y_2, \ldots, Y_T)'$ given $\eta = (\eta_1, \eta_2, \ldots, \eta_T)'$ is again normal with mean $\mu + \Lambda\eta$ and covariance matrix Φ, where $\mu = (\mu_1, \ldots, \mu_T)'$, Λ and Φ are given by,

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & \lambda_T \end{pmatrix} \tag{5.3}$$

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1T} \\ \phi_{21} & \phi_{22} & & \\ \vdots & & \ddots & \\ \phi_{T1} & \phi_{T2} & \cdots & \phi_{TT} \end{pmatrix} \tag{5.4}$$

where $\phi_{t,t'}$ represents the $k \times k$ covariance matrix between unique factor vectors u_t and $u_{t'}$ where $t = 1, \ldots, T$ and $t' = 1, \ldots, T$. In the most general settings, Φ is allowed to be any symmetric matrix, but a reasonable way of imposing constraints in the current setting of PFA model fitting would be to restrict $\phi_{t,t'}$ as a diagonal matrix so that all the unique factors are uncorrelated with each other within time and the covariances of specific factors are allowed to be non-zero across time . In a similar way, any other types of constraints can be imposed on Φ. It is often appropriate to incorporate deterministic trends for the mean over time, in which case the mean vector can be expressed as a function of time t and a parameter vector γ i.e. $\mu_t = \mu(t, \gamma)$. In the present context, however, where the primary goal is in obtaining maximum likelihood estimates for AR and MA coefficients explaining the temporal flow of latent common factors, the mean vector of μ is effectively a nuisance parameter so that it will be set to be a zero vector in the population. Thus, the observed-data log-likelihood function conditional on η is given proportional to

$$\ln g(Y|\eta, \Lambda, \Phi) = -\frac{kmT}{2} \ln 2\pi - \frac{1}{2} \ln |\Phi| - \frac{1}{2} (Y - \Lambda\eta)' \, \Phi^{-1} \, (Y - \Lambda\eta) \tag{5.5}$$

5.3 The marginal likelihood function of η

The marginal distribution of $\eta = (\eta_1, \ldots, \eta_T)'$ following a VARMA(p, q) process is assumed to be normal with zero mean and covariance matrix, $\Sigma(\tau)$, where τ is the parameter vector representing AR and MA coefficients and the variance-covariance matrix of the initial status vector, Θ, and the random shock vector, Ψ. Then, the likelihood function of η is proportional to

$$\ln h(\eta|\tau) = -\frac{mT}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma(\tau)| - \frac{1}{2}\eta'\Sigma(\tau)^{-1}\eta \tag{5.6}$$

The general expression of the covariance structure of a stationary latent time series, $\Sigma(\tau)$ was derived by du Toit and Browne (2007), which was reproduced in the previous section in equation (4.10).

5.4 The complete-data log-likelihood of Y and η

The EM algorithm uses the complete-data likelihood, which is the joint likelihood of Y and η constructed by supposing η is observed. Let ξ be the collection of the model parameters specified in Λ, Φ, and τ. Then the joint log-likelihood is proportional to

$$\begin{aligned}\ln f(Y,\eta|\xi) &= \ln g(Y|\eta,\Lambda,\Phi) + \ln h(\eta|\tau)\\ &= -\frac{mT(k+1)}{2}\ln 2\pi - \frac{1}{2}\ln|\Phi| - \frac{1}{2}\ln|\Sigma(\tau)|\\ &\quad - \frac{1}{2}\mathrm{tr}\big\{\Phi^{-1}YY' - \Phi^{-1}\Lambda\eta Y' - \Phi^{-1}Y\eta'\Lambda' + \Phi^{-1}\Lambda\eta\eta'\Lambda'\big\}\\ &\quad - \frac{1}{2}\mathrm{tr}\big\{\Sigma(\tau)^{-1}\eta\eta'\big\}\end{aligned} \tag{5.7}$$

The complete-data expected log-likelihood or Q-function is

$$\begin{aligned}Q(\xi|\xi^{(\ell)}) &= \mathrm{E}[\ln f(Y,\eta|\xi)|Y,\xi^{(\ell)}]\\ &= -\frac{mT(k+1)}{2}\ln 2\pi - \frac{1}{2}\ln|\Phi| - \frac{1}{2}\ln|\Sigma(\tau)|\\ &\quad - \frac{1}{2}\mathrm{tr}\left[\Phi^{-1}\big\{YY' - \Lambda\mathrm{E}[\eta|Y,\xi^{(\ell)}]Y' - Y\mathrm{E}[\eta|Y,\xi^{(\ell)}]'\Lambda' + \Lambda E[\eta\eta'|Y,\xi^{(\ell)}]\Lambda'\big\}\right]\\ &\quad - \frac{1}{2}\mathrm{tr}\big\{\Sigma(\tau)^{-1}E[\eta\eta'|Y,\xi^{(\ell)}]\big\}\end{aligned} \tag{5.8}$$

where the $\xi^{(\ell)}$ indicates the current estimates of parameters. As proved in Appendix A,

$$
\begin{aligned}
\mathrm{E}[\eta|Y,\xi^{(\ell)}] &= \overbrace{\Sigma(\tau)\Lambda'(\Phi+\Lambda\Sigma(\tau)\Lambda')^{-1}}^{(\ell)}Y \\
&= \underbrace{\left(\Lambda'\Phi^{-1}\Lambda+\Sigma(\tau)^{-1}\right)^{-1}\Lambda'\Phi^{-1}}_{(\ell)}Y
\end{aligned}
\tag{5.9}
$$

$$
\begin{aligned}
\mathrm{cov}[\eta,\eta'|Y,\xi^{(\ell)}] &= \overbrace{\Sigma(\tau)-\Sigma(\tau)\Lambda'(\Phi+\Lambda\Sigma(\tau)\Lambda')^{-1}\Lambda\Sigma(\tau)}^{(\ell)} \\
&= \underbrace{\left(\Lambda'\Phi^{-1}\Lambda+\Sigma(\tau)^{-1}\right)^{-1}}_{(\ell)}
\end{aligned}
\tag{5.10}
$$

where the overbraced- and the underbraced- quantities with ℓ represent that parameters in Λ, Φ and $\Sigma(\tau)$ matrices are replaced by current estimates of those parameters. Notice that

$$
\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}] = \mathrm{E}[\eta|Y,\xi^{(\ell)}]\mathrm{E}[\eta|Y,\xi^{(\ell)}]' + \mathrm{cov}[\eta,\eta'|Y,\xi^{(\ell)}] \tag{5.11}
$$

Therefore, the E step is completed by inserting (5.9) and (5.11) into (5.8). The next M step then maximizes this expected complete log-likelihood function of (5.8) to obtain the next iterate $\xi^{(\ell+1)}$. The Newton-Raphson or Fisher Scoring subroutine can be employed. In order to implement these subroutines, the analytic gradient vector and the Hessian matrix of Q-function are required, which will be derived in the next section.

5.5 The analytic gradient vector and Hessian matrix of Q-function

Let γ denote a vector of parameters for the latent VARMA(p,q) process model. Then the typical element of the gradient vector of the Q-function with respect to an element in γ can be calculated by using the following facts of matrix calculus

$$\frac{\partial y}{\partial x} = \mathrm{tr}\left[\frac{\partial y}{\partial Z}\frac{\partial Z'}{\partial x}\right] = \mathrm{tr}\left[\frac{\partial y}{\partial Z'}\frac{\partial Z}{\partial x}\right] \tag{5.12}$$

$$\frac{\partial Z^{-1}}{\partial x} = -Z^{-1}\frac{\partial Z}{\partial x}Z^{-1} \tag{5.13}$$

$$\frac{\partial \ln |Z|}{\partial Z} = Z^{-1'} \tag{5.14}$$

where y is a scalar function of the elements of a $p \times q$ nonsingular Z matrix and the elements of Z, in turn, are functions of the scalar variable, x. The equation (5.14) holds when Z is a nonsingular $p \times p$ matrix. Therefore, the the gradient vector is given as

$$\begin{aligned}
\frac{\partial Q}{\partial \gamma_r} &= -\frac{1}{2}\mathrm{tr}\left\{\Sigma_\tau^{-1}\frac{\partial \Sigma_\tau}{\partial \gamma_r}\right\} + \frac{1}{2}\mathrm{tr}\left\{\Sigma_\tau^{-1}\frac{\partial \Sigma_\tau}{\partial \gamma_r}\Sigma_\tau^{-1}\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right\} \\
&= -\frac{1}{2}\mathrm{tr}\left\{\Sigma_\tau^{-1}\left(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\Sigma_\tau^{-1}\frac{\partial \Sigma_\tau}{\partial \gamma_r}\right\} \\
&= -\frac{1}{2}\mathrm{vec}'\left\{\Sigma_\tau^{-1}\left(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\Sigma_\tau^{-1}\right\}\left(\frac{\partial \mathrm{vec}\Sigma_\tau}{\partial \gamma_r}\right) \\
&= -\frac{1}{2}\left\{\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)\mathrm{vec}\left(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\right\}'\left(\frac{\partial \mathrm{vec}\Sigma_\tau}{\partial \gamma_r}\right) \\
&= -\frac{1}{2}\left\{\mathrm{vec}'\left(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)\right\}\left(\frac{\partial \mathrm{vec}\Sigma_\tau}{\partial \gamma_r}\right)
\end{aligned} \tag{5.15}$$

where γ_r represents the r^{th} element in γ. In matrix notation,

$$\frac{\partial Q}{\partial \gamma} = -\frac{1}{2}\triangle'(\gamma)\left\{\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)\mathrm{vec}\left(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\right\} \tag{5.16}$$

where $\triangle(\gamma) \equiv \frac{\partial \mathrm{vec}\Sigma_\tau}{\partial \gamma'}$.

Also note the following fact regarding the derivative of a product of matrices with respect to a matrix,

$$\frac{\partial XY}{\partial Z} = \frac{\partial X}{\partial Z}(I_q \otimes Y) + (I_p \otimes X)\frac{\partial Y}{\partial Z} \tag{5.17}$$

where X, Y, and Z are matrices of order $m \times n$, $n \times v$ and $p \times q$, respectively. Then, the Hessian matrix can be derived using the above fact in conjunction with other properties of the matrix calculus including equations (5.12) to (5.14), which is given as

$$\begin{aligned}
\frac{\partial^2 Q}{\partial \gamma \, \partial \gamma'} &= \frac{\partial}{\partial \gamma'} \left[-\frac{1}{2} \triangle(\gamma)' W \right] \\
&= -\frac{1}{2} \left[\frac{\partial \triangle(\gamma)'}{\partial \gamma'} (I_q \otimes W) + (I_1 \otimes \triangle(\gamma)') \frac{\partial W}{\partial \gamma'} \right] \qquad (5.18)
\end{aligned}$$

where I_q is a $q \times q$ identity matrix, $I_1 = 1$ and $W = (\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1}) \operatorname{vec}\left(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}]\right)$. Notice that

$$\begin{aligned}
W &= \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1}\right) \operatorname{vec}\left(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}]\right) \\
&= \operatorname{vec}\{\Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1}\} \qquad (5.19)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial W}{\partial \gamma'} &= \frac{\partial \operatorname{vec}\{\Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1}\}}{\partial \gamma'} \\
&= \sum_{s=1}^{q} \frac{\partial \operatorname{vec}\{\Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1}\}}{\partial \gamma_s} E_{1q} \\
&= \sum_{s=1}^{q} \operatorname{vec}\Bigg\{ -\Sigma_\tau^{-1} \left(\frac{\partial \Sigma_\tau}{\partial \gamma_s}\right) \Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1} \\
&\qquad + \Sigma_\tau^{-1} \left(\frac{\partial \Sigma_\tau}{\partial \gamma_s}\right) \Sigma_\tau^{-1} - \Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1} \left(\frac{\partial \Sigma_\tau}{\partial \gamma_s}\right) \Sigma_\tau^{-1} \Bigg\} E_{1q} \\
&= \sum_{s=1}^{q} \Bigg\{ -\left(\Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1}\right) \operatorname{vec}\left(\frac{\partial \Sigma_\tau}{\partial \gamma_s}\right) \\
&\qquad + \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1}\right) \operatorname{vec}\left(\frac{\partial \Sigma_\tau}{\partial \gamma_s}\right) - \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1}\right) \operatorname{vec}\left(\frac{\partial \Sigma_\tau}{\partial \gamma_s}\right) \Bigg\} E_{1q} \\
&= -\left(\Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1}\right) \triangle(\gamma) + \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1}\right) \triangle(\gamma) \\
&\qquad - \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1}(\Sigma_\tau - \mathrm{E}[\eta\eta'|Y, \xi^{(\ell)}])\Sigma_\tau^{-1}\right) \triangle(\gamma) \qquad (5.20)
\end{aligned}$$

where E_{1q} is a $1 \times q$ zero vector with only non-zero element, 1, positioned at $(1, q)$. It follows that the Hessian matrix of (5.18) becomes

$$
\begin{aligned}
\frac{\partial^2 Q}{\partial\gamma\,\partial\gamma'} = -\frac{1}{2}&\left\{\frac{\partial\triangle(\gamma)'}{\partial\gamma'}\left(I_q\otimes\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)\mathrm{vec}\left(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\right)\right\}\\
&-\frac{1}{2}\triangle(\gamma)'\left\{\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)-\left(\Sigma_\tau^{-1}(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}])\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)\right.\\
&\left.-\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}])\Sigma_\tau^{-1}\right)\right\}\triangle(\gamma) \qquad (5.21)
\end{aligned}
$$

Using the following relationship,

$$
\begin{aligned}
&\triangle(\gamma)'\left(\Sigma_\tau^{-1}(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}])\Sigma_\tau^{-1}\right)\otimes\Sigma_\tau^{-1}\triangle(\gamma)\\
&=\triangle(\gamma)'N'\left(\Sigma_\tau^{-1}(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}])\Sigma_\tau^{-1}\right)\otimes\Sigma_\tau^{-1}N\triangle(\gamma)\\
&=\triangle(\gamma)'N'\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}])\Sigma_\tau^{-1}\right)N\triangle(\gamma)\\
&=\triangle(\gamma)'\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}])\Sigma_\tau^{-1}\right)\triangle(\gamma) \qquad (5.22)
\end{aligned}
$$

where N is a matrix such that $N\mathrm{vec}(A)=\frac{1}{2}\mathrm{vec}(A+A')$, the final form of Hessian matrix is given as

$$
\begin{aligned}
\frac{\partial^2 Q}{\partial\gamma\,\partial\gamma'} &= -\frac{1}{2}\left\{\frac{\partial\triangle(\gamma)'}{\partial\gamma'}\left(I_q\otimes\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)\mathrm{vec}\left(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\right)\right\}\\
&\quad-\frac{1}{2}\triangle(\gamma)'\left\{\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)-2\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}])\Sigma_\tau^{-1}\right)\right\}\triangle(\gamma)\\
&= -\frac{1}{2}\left\{\frac{\partial\triangle(\gamma)'}{\partial\gamma'}\left(I_q\otimes\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)\mathrm{vec}\left(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\right)\right\}\\
&\quad+\triangle(\gamma)'\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}(\Sigma_\tau-\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}])\Sigma_\tau^{-1}\right)\triangle(\gamma)\\
&\quad-\frac{1}{2}\triangle(\gamma)'\left(\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\right)\triangle(\gamma) \qquad (5.23)
\end{aligned}
$$

In order to complete the analytic derivation of the gradient vector and the Hessian matrix, the Jacobian matrix, denoted by $\triangle(\gamma)$, whose typical element is $\frac{\partial \Sigma_\tau}{\partial \gamma_r}$, is required.

$$\begin{aligned}
\frac{\partial \Sigma_\tau}{\partial \gamma_r} &= \frac{\partial}{\partial \gamma_r}\Big\{\mathrm{T}_{-A}^{-1}\left(I_{T|s}\Theta I_{T|s}' + \mathrm{T}_B(I_T \otimes \Psi)\mathrm{T}_B'\right)\mathrm{T}_{-A}^{-1'}\Big\} \\
&= U + U' + T_{-A}^{-1}\left\{I_{T|s}\left(\frac{\partial \Theta}{\partial \gamma_r}\right)I_{T|s}' + V + T_B\left[I_T \otimes \left(\frac{\partial \Psi}{\partial \gamma_r}\right)\right]T_B' + V'\right\}T_{-A}^{-1'} \\
&= U + U' + T_{-A}^{-1}(V+V')T_{-A}^{-1'} + T_{-A}^{-1}I_{T|s}\left(\frac{\partial \Theta}{\partial \gamma_r}\right)I_{T|s}'T_{-A}^{-1'} \\
&\quad + T_{-A}^{-1}T_B\left[I_T \otimes \left(\frac{\partial \Psi}{\partial \gamma_r}\right)\right]T_B'T_{-A}^{-1'}
\end{aligned} \tag{5.24}$$

where $U = -T_{-A}^{-1}\left(\frac{\partial T_{-A}}{\partial \gamma_r}\right)\Sigma_\tau$, $V = \left(\frac{\partial T_B}{\partial \gamma_r}\right)(I_T \otimes \Psi)\,T_B'$ with

$$\frac{\partial T_{-A}}{\partial \gamma_r} = \begin{cases} -\sum_{c=1}^{T-j} J_{1\ u+(j+c-1)m, v+(c-1)m}, & \gamma_r = A_{j(u,v)} \\ 0, & \gamma_r = B_{j(u,v)} \\ 0, & \gamma_r = \Psi_{(u,v)} \end{cases} \tag{5.25}$$

$$\frac{\partial T_B}{\partial \gamma_r} = \begin{cases} 0, & \gamma_r = A_{j(u,v)} \\ \sum_{c=1}^{T-j} J_{1\ u+(j+c-1)m, v+(c-1)m}, & \gamma_r = B_{j(u,v)} \\ 0, & \gamma_r = \Psi_{(u,v)} \end{cases} \tag{5.26}$$

$$\frac{\partial \Psi}{\partial \gamma_r} = \begin{cases} 0, & \gamma_r = A_{j(u,v)} \\ 0, & \gamma_r = B_{j(u,v)} \\ \frac{2-\delta_{uv}}{2}(J_{2\ u,v} + J_{2\ v,u}), & \gamma_r = \Psi_{(u,v)} \end{cases} \tag{5.27}$$

where $j = 1, 2, \ldots, s = \max(p, q)$ and $J_{1\ u,v}$ is the $mT \times mT$ zero matrix that has its only nonzero element, a one, in the $(u,v)^{th}$ position and $J_{2\ u,v}$ is the $m \times m$ zero matrix with its nonzero element, a one, in the $(u,v)^{th}$ position. Finally, vec $\left(\frac{\partial \Theta}{\partial \gamma_r}\right)$ is give by

$$\mathrm{vec}\left(\frac{\partial \Theta}{\partial \gamma_r}\right) = (I - A \otimes A)^{-1}\mathrm{vec}(G_{\gamma_s}^*) \tag{5.28}$$

where

$$G_{\gamma_s}^* = \left(\frac{\partial A}{\partial \gamma_r}\right)\Theta A' + A\Theta\left(\frac{\partial A'}{\partial \gamma_r}\right) + \left(\frac{\partial G}{\partial \gamma_r}\right)\Psi G' + G\left(\frac{\partial \Psi}{\partial \gamma_r}\right)G' + G\Psi\left(\frac{\partial G'}{\partial \gamma_r}\right) \tag{5.29}$$

and

$$\frac{\partial A}{\partial \gamma_r} = \begin{cases} J_{3\ (j-1)m+u,v}, & \gamma_r = A_{j(u,v)} \\ 0, & \gamma_r = B_{j(u,v)} \\ 0, & \gamma_r = \Psi_{(u,v)} \end{cases} \tag{5.30}$$

$$\frac{\partial G}{\partial \gamma_r} = \begin{cases} J_{4\ (j-1)m+u,v}, & \gamma_r = A_{j(u,v)} \\ J_{4\ (j-1)m+u,v}, & \gamma_r = B_{j(u,v)} \\ 0, & \gamma_r = \Psi_{(u,v)} \end{cases} \tag{5.31}$$

where $J_{3\ u,v}$ is the $ms \times ms$ zero matrix that has its only nonzero element, a one, in the $(u,v)^{th}$ position and $J_{4\ u,v}$ is the $ms \times m$ zero matrix with its nonzero element, a one, in the $(u,v)^{th}$ position. Again, $s = \max(p,q)$ and $j = 1, 2, \ldots, s$.

In addition, the typical element for the second order partial derivatives of Σ_τ, whose typical element can be represented as $\frac{\partial^2 \Sigma_\tau}{\partial \gamma_r \partial \gamma_s}$, needs to be derived for obtaining $\frac{\partial \triangle(\tau)'}{\partial \gamma'}$. Notice that

$$\begin{aligned} \frac{\partial^2 T_{-A}}{\partial \gamma_r \partial \gamma_s} = \frac{\partial^2 T_B}{\partial \gamma_r \partial \gamma_s} = \frac{\partial^2 \Psi}{\partial \gamma_r \partial \gamma_s} \\ = \frac{\partial^2 A}{\partial \gamma_r \partial \gamma_s} = \frac{\partial^2 G}{\partial \gamma_r \partial \gamma_s} = 0 \end{aligned} \tag{5.32}$$

Therefore,

$$\frac{\partial U}{\partial \gamma_s} = T_{-A}^{-1} \left(\frac{\partial T_{-A}}{\partial \gamma_s} \right) T_{-A}^{-1} \left(\partial \frac{T_{-A}}{\partial \gamma_r} \right) \Sigma_\gamma - T_{-A}^{-1} \left(\frac{\partial T_{-A}}{\partial \gamma_r} \right) \left(\frac{\partial \Sigma_\tau}{\partial \gamma_s} \right) \tag{5.33}$$

$$\frac{\partial V}{\partial \gamma_s} = \left(\frac{\partial T_B}{\partial \gamma_r} \right) \left(I_T \otimes \frac{\partial \Psi}{\partial \gamma_s} \right) T_B' + \left(\frac{\partial T_B}{\partial \gamma_r} \right) (I_T \otimes \Psi) \left(\frac{\partial T_B}{\partial \gamma_s} \right)' \tag{5.34}$$

Then, $\frac{\partial^2 \Sigma_\tau}{\partial \gamma_r \partial \gamma_s}$ is given as

$$\begin{aligned} \frac{\partial^2 \Sigma_\tau}{\partial \gamma_r \partial \gamma_s} = {} & \frac{\partial U}{\partial \gamma_s} + \frac{\partial U'}{\partial \gamma_s} + W + W' + T_{-A}^{-1} \left(\frac{\partial V}{\partial \gamma_s} + \frac{\partial V'}{\partial \gamma_s} \right) T_{-A}^{-1'} \\ & + Z + Z' + T_{-A}^{-1} T_{T|s} \left(\frac{\partial^2 \Theta}{\partial \gamma_r \partial \gamma_s} \right) I_{T|s}' T_{-A}^{-1'} \\ & + X + X' + Y + Y' \end{aligned} \tag{5.35}$$

where

$$W = -T_{-A}^{-1}\left(\frac{\partial T_{-A}}{\partial \gamma_s}\right)T_{-A}^{-1}(V+V')T_{-A}^{-1'} \tag{5.36}$$

$$Z = -T_{-A}^{-1}\left(\frac{\partial T_{-A}}{\partial \gamma_s}\right)T_{-A}^{-1}I_{T|s}\left(\frac{\partial \Theta}{\partial \gamma_r}\right)I_{T|s}'T_{-A}^{-1'} \tag{5.37}$$

$$X = -T_{-A}^{-1}\left(\frac{\partial T_{-A}}{\partial \gamma_s}\right)T_{-A}^{-1}T_B\left[I_T \otimes \left(\frac{\partial \Psi}{\partial \gamma_r}\right)\right]T_B'T_{-A}^{-1'} \tag{5.38}$$

$$Y = T_{-A}^{-1}\left(\frac{\partial T_B}{\partial \gamma_s}\right)\left[I_T \otimes \left(\frac{\partial \Psi}{\partial \gamma_r}\right)\right]T_B'T_{-A}^{-1'} \tag{5.39}$$

The $\left(\frac{\partial^2 \Theta}{\partial \gamma_r \partial \gamma_s}\right)$ can obtained by taking partial derivatives of the following equations derived by du Toit and Browne (2007).

$$\Theta - A\Theta A' = G\Theta G' \tag{5.40}$$

The first order partial derivatives are given as

$$\frac{\partial \Theta}{\partial \gamma_r} - \left\{\left(\frac{\partial A}{\partial \gamma_r}\right)\Theta A' + A\left(\frac{\partial \Theta}{\partial \gamma_r}\right)A' + A\Theta\left(\frac{\partial A'}{\partial \gamma_r}\right)\right\}$$
$$= \left(\frac{\partial G}{\partial \gamma_r}\right)\Psi G' + G\left(\frac{\partial \Psi}{\partial \gamma_r}\right)G' + G\Psi\left(\frac{\partial G'}{\partial \gamma_r}\right) \tag{5.41}$$

$$\tag{5.42}$$

And the second order partial derivatives are given as

$$\frac{\partial^2 \Theta}{\partial \gamma_r \partial \gamma_s} - \Bigg\{ \left(\frac{\partial A}{\partial \gamma_r}\right)\left(\frac{\partial \Theta}{\partial \gamma_s}\right)A' + \left(\frac{\partial A}{\partial \gamma_r}\right)\Theta\left(\frac{\partial A'}{\partial \gamma_s}\right)$$
$$\left(\frac{\partial A}{\partial \gamma_s}\right)\left(\frac{\partial \Theta}{\partial \gamma_r}\right)A' + A\left(\frac{\partial^2 \Theta}{\partial \gamma_r \partial \gamma_s}\right)A' + A\left(\frac{\partial \Theta}{\partial \gamma_r}\right)\left(\frac{\partial A'}{\partial \gamma_s}\right)$$
$$\left(\frac{\partial A}{\partial \gamma_s}\right)\Theta\left(\frac{\partial A'}{\partial \gamma_r}\right) + A\left(\frac{\partial \Theta}{\partial \gamma_s}\right)\left(\frac{\partial A'}{\partial \gamma_r}\right)\Bigg\}$$
$$=$$
$$\left(\frac{\partial G}{\partial \gamma_r}\right)\left(\frac{\partial \Psi}{\partial \gamma_s}\right)G' + \left(\frac{\partial G}{\partial \gamma_r}\right)\Psi\left(\frac{\partial G'}{\partial \gamma_s}\right)$$
$$\left(\frac{\partial G}{\partial \gamma_s}\right)\left(\frac{\partial \Psi}{\partial \gamma_r}\right)G' + G\left(\frac{\partial \Psi}{\partial \gamma_r}\right)\left(\frac{\partial G'}{\partial \gamma_s}\right)$$
$$\left(\frac{\partial G}{\partial \gamma_s}\right)\Psi\left(\frac{\partial G'}{\partial \gamma_r}\right) + G\left(\frac{\partial \Psi}{\partial \gamma_s}\right)\left(\frac{\partial G'}{\partial \gamma_r}\right) \tag{5.43}$$

Therefore,

$$\text{vec}\left(\frac{\partial^2\Theta}{\partial\gamma_r\partial\gamma_s}\right) = (I - A\otimes A)^{-1}\text{vec}\left(\ddot{G}\right) \tag{5.44}$$

where

$$\begin{aligned}\ddot{G} = &\left(\frac{\partial A}{\partial\gamma_r}\right)\left(\frac{\partial\Theta}{\partial\gamma_s}\right)A' + \left(\frac{\partial A}{\partial\gamma_r}\right)\Theta\left(\frac{\partial A'}{\partial\gamma_s}\right)\\ &\left(\frac{\partial A}{\partial\gamma_s}\right)\left(\frac{\partial\Theta}{\partial\gamma_r}\right)A' + A\left(\frac{\partial\Theta}{\partial\gamma_r}\right)\left(\frac{\partial A'}{\partial\gamma_s}\right)\\ &\left(\frac{\partial A}{\partial\gamma_s}\right)\Theta\left(\frac{\partial A'}{\partial\gamma_r}\right) + A\left(\frac{\partial\Theta}{\partial\gamma_s}\right)\left(\frac{\partial A'}{\partial\gamma_r}\right)\\ &\left(\frac{\partial G}{\partial\gamma_r}\right)\left(\frac{\partial\Psi}{\partial\gamma_s}\right)G' + \left(\frac{\partial G}{\partial\gamma_r}\right)\Psi\left(\frac{\partial G'}{\partial\gamma_s}\right)\\ &\left(\frac{\partial G}{\partial\gamma_s}\right)\left(\frac{\partial\Psi}{\partial\gamma_r}\right)G' + G\left(\frac{\partial\Psi}{\partial\gamma_r}\right)\left(\frac{\partial G'}{\partial\gamma_s}\right)\\ &\left(\frac{\partial G}{\partial\gamma_s}\right)\Psi\left(\frac{\partial G'}{\partial\gamma_r}\right) + G\left(\frac{\partial\Psi}{\partial\gamma_s}\right)\left(\frac{\partial G'}{\partial\gamma_r}\right)\end{aligned} \tag{5.45}$$

Finally, $\frac{\partial\triangle(\tau)}{\partial\gamma}$ can be calculated by using the following properties well known in the matrix algebra (Abadir and Magnus, 2005),

$$A = \sum_j A_{.j}e'_j \tag{5.46}$$

$$K_{pm}(A\otimes b) = b\otimes A \tag{5.47}$$

$$K_{mn}(dc'\otimes A) = c'\otimes A\otimes d \tag{5.48}$$

where $A_{.j}$ indicates j^{th} column of a $m\times n$ A matrix, e_j represents the j^{th} column of the $n\times n$ identity matrix, and b, c and d is a $p\times 1$, $q\times 1$, and $n\times 1$ vector, respectively. K_{mn} is the so-called communication matrix such that $K_{mn}\text{vec}(A) = \text{vec}(A')$.

$$
\begin{aligned}
\frac{\partial \triangle(\tau)}{\partial \gamma} &= \sum_{s=1}^{q} e_s \otimes \left(\frac{\partial \triangle(\tau)}{\partial \gamma_s} \right) \\
&= \sum_{s=1}^{q} e_s \otimes \frac{\partial}{\partial \gamma_s} \left(\frac{\partial \mathrm{vec}\Sigma_\tau}{\partial \gamma'} \right) \\
&= \sum_{s=1}^{q} e_s \otimes \frac{\partial}{\partial \gamma_s} \left[\sum_{r=1}^{s} \left(\frac{\partial \mathrm{vec}\Sigma_\tau}{\partial \gamma_r} \right) e_r' \right] \\
&= \sum_{r=1}^{q} \sum_{s=1}^{q} e_s \otimes (a_{rs} e_r') \\
&= \sum_{r=1}^{q} \sum_{s=1}^{q} K_{qp^2} (a_{rs} e_r' \otimes e_s) \\
&= \sum_{r=1}^{q} \sum_{s=1}^{q} (e_r' \otimes e_s \otimes a_{rs}) \qquad (5.49)
\end{aligned}
$$

where $a_{rs} = \frac{\partial^2 \mathrm{vec}\Sigma_\tau}{\partial \gamma_r \partial \gamma_s}$

In sum, the gradient vector and the Hessian matrix of the Q-function with respect to γ are given by

$$
\frac{\partial Q}{\partial \gamma} = -\frac{1}{2} \triangle(\gamma)' \Big\{ \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1} \right) \mathrm{vec} \left(\Sigma_\tau - \mathrm{E}[\eta\eta' | Y, \xi^{(\ell)}] \right) \Big\} \qquad (5.50)
$$

where $\triangle(\gamma) = \frac{\partial \mathrm{vec}\Sigma_\tau}{\partial \gamma'}$. The Hessian matrix is given as

$$
\begin{aligned}
\frac{\partial^2 Q}{\partial \gamma \, \partial \gamma'} = -\frac{1}{2} \Bigg\{ & \frac{\partial \triangle(\gamma)'}{\partial \gamma'} \left(I_q \otimes \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1} \right) \mathrm{vec} \left(\Sigma_\tau - \mathrm{E}[\eta\eta' | Y, \xi^{(\ell)}] \right) \right) \Bigg\} \\
& + \triangle(\gamma)' \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1} (\Sigma_\tau - \mathrm{E}[\eta\eta' | Y, \xi^{(\ell)}]) \Sigma_\tau^{-1} \right) \triangle(\gamma) \\
& - \frac{1}{2} \triangle(\gamma)' \left(\Sigma_\tau^{-1} \otimes \Sigma_\tau^{-1} \right) \triangle(\gamma) \qquad (5.51)
\end{aligned}
$$

Now, the gradient vector and the Hessian matrix of the Q function with respect to free parameters in Λ and Φ need to be derived. Note the following facts from matrix calculus,

$$\mathrm{tr}[ABCD] = \mathrm{vec}'(B')(A' \otimes C)\mathrm{vec}(D) \tag{5.52}$$

where A, B, C, and D are square matrices of appropriate order.

Using the above fact in conjunction with equations (5.12) to (5.14), the typical element in the gradient vector of the the Q-function with respect to parameters in the measurement model, λ and Φ is given as follows:

$$\begin{aligned}
\frac{\partial Q}{\partial \lambda_{(r,s)}} &= -\frac{1}{2}\mathrm{tr}\left\{-2\mathrm{E}[\eta|Y,\xi^{(\ell)}]Y'\Phi^{-1}\left(\frac{\partial \Lambda}{\partial \lambda_{(r,s)}}\right) + 2\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\Lambda'\Phi^{-1}\left(\frac{\partial \Lambda}{\partial \lambda_{(r,s)}}\right)\right\} \\
&= \mathrm{vec}'\left(\Phi^{-1}Y\mathrm{E}[\eta|Y,\xi^{(\ell)}]' - \Phi^{-1}\Lambda\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\right)\mathrm{vec}\left(\frac{\partial \Lambda}{\partial \lambda_{(r,s)}}\right)
\end{aligned} \tag{5.53}$$

$$\begin{aligned}
\frac{\partial Q}{\partial \Phi_{(r,s)}} &= -\frac{1}{2}\frac{\partial \ln|\Phi|}{\partial \Phi_{(r,s)}} - \frac{1}{2}\mathrm{tr}\left\{\left(\frac{\partial \Phi^{-1}}{\partial \Phi_{(r,s)}}\right)\hat{e}\right\} \\
&= -\frac{1}{2}\mathrm{tr}\left\{\frac{\partial \ln|\Phi|}{\partial \Phi'}\left(\frac{\partial \Phi}{\partial \Phi_{(r,s)}}\right)\right\} + \frac{1}{2}\mathrm{tr}\left\{\Phi^{-1}\left(\frac{\partial \Phi}{\partial \Phi_{(r,s)}}\right)\Phi^{-1}\hat{e}\right\} \\
&= -\frac{1}{2}\mathrm{tr}\left\{\left(\Phi^{-1} - \Phi^{-1}\hat{e}\Phi^{-1}\right)\left(\frac{\partial \Phi}{\partial \Phi_{(r,s)}}\right)\right\} \\
&= \frac{1}{2}\mathrm{vec}'\left(\Phi^{-1}(\hat{e} - \Phi)\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial \Phi}{\partial \Phi_{(r,s)}}\right)
\end{aligned} \tag{5.54}$$

where $\hat{e} = YY' - \Lambda\mathrm{E}[\eta|Y,\xi^{(\ell)}]Y' - Y\mathrm{E}[\eta|Y,\xi^{(\ell)}]'\Lambda' + \Lambda\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\Lambda'$. In matrix notation,

$$\begin{aligned}
\frac{\partial Q}{\partial \mathrm{vec}(\lambda)} &= \triangle'(\lambda)\left[\left(\mathrm{E}[\eta|Y,\xi^{(\ell)}]Y' \otimes \Phi^{-1}\right)\mathrm{vec}(I) - \left(\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}] \otimes \Phi^{-1}\right)\mathrm{vec}(\Lambda)\right] \\
\frac{\partial Q}{\partial \mathrm{vec}(\phi)} &= \frac{1}{2}\triangle'(\phi)\left[\left(\Phi^{-1} \otimes \Phi^{-1}\right)\mathrm{vec}\left(\hat{e} - \Phi\right)\right]
\end{aligned} \tag{5.55}$$

where $\triangle(\lambda) = \frac{\partial \mathrm{vec}\Lambda}{\partial \mathrm{vec}'(\lambda)}$ and $\triangle(\phi) = \frac{\partial \mathrm{vec}\Phi}{\partial \mathrm{vec}'(\phi)}$. Here, λ and ϕ represent the vector of free parameters in Λ and Φ, respectively.

The typical elements in the Hessian matrix of the Q-function with respect to λ, Φ are derived as follows:

$$\begin{aligned}
\frac{\partial^2 Q}{\partial\lambda_{(r,s)}\partial\lambda_{(u,v)}} &= -\frac{1}{2}\mathrm{tr}\left\{2\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\left(\frac{\partial\Lambda}{\partial\lambda_{(u,v)}}\right)'\Phi^{-1}\left(\frac{\partial\Lambda}{\partial\lambda_{(r,s)}}\right)\right\}\\
&= -\mathrm{vec}'\left(\frac{\partial\Lambda}{\partial\lambda_{(u,v)}}\right)\left(\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\otimes\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Lambda}{\partial\lambda_{(r,s)}}\right) \qquad (5.56)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 Q}{\partial\Phi_{(r,s)}\partial\Phi_{(u,v)}} &= -\frac{1}{2}\mathrm{tr}\left\{-\Phi^{-1}\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\Phi^{-1}\left(\frac{\partial\Phi}{\partial\Phi_{(r,s)}}\right)\right\}\\
&+\frac{1}{2}2\mathrm{tr}\left\{-\Phi^{-1}\hat{e}\Phi^{-1}\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\Phi^{-1}\left(\frac{\partial\Phi}{\partial\Phi_{(r,s)}}\right)\right\}\\
&= \frac{1}{2}\mathrm{vec}'\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\left(\Phi^{-1}\otimes\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\Phi_{(r,s)}}\right)\\
&- \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\left(\Phi^{-1}\hat{e}\Phi^{-1}\otimes\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\Phi_{(r,s)}}\right)\\
&= \frac{1}{2}\mathrm{vec}'\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\left[\left(\Phi^{-1}(\Phi-2\hat{e})\Phi^{-1}\right)\otimes\Phi^{-1}\right]\mathrm{vec}\left(\frac{\partial\Phi}{\partial\Phi_{(r,s)}}\right) \qquad (5.57)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 Q}{\partial\lambda_{(r,s)}\partial\Phi_{(u,v)}} &= -\mathrm{tr}\left\{\mathrm{E}[\eta|Y,\xi^{(\ell)}]Y'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\Phi^{-1}\left(\frac{\partial\Lambda}{\partial\lambda_{(r,s)}}\right)\right\}\\
&+\mathrm{tr}\left\{\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\Phi^{-1}\left(\frac{\partial\Lambda}{\partial\lambda_{(r,s)}}\right)\right\}\\
&= -\mathrm{vec}'\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\left(\Phi^{-1}Y\mathrm{E}[\eta|Y,\xi^{(\ell)}]'\otimes\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Lambda}{\partial\lambda_{(r,s)}}\right)\\
&+\mathrm{vec}'\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\left(\Phi^{-1}\Lambda\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\otimes\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Lambda}{\partial\lambda_{(r,s)}}\right)\\
&= \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\Phi_{(u,v)}}\right)\Bigg[\left(\Phi^{-1}\Lambda\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]-\Phi^{-1}Y\mathrm{E}[\eta|Y,\xi^{(\ell)}]'\right)\\
&\qquad\otimes\Phi^{-1}\Bigg]\mathrm{vec}\left(\frac{\partial\Lambda}{\partial\lambda_{(r,s)}}\right) \qquad (5.58)
\end{aligned}$$

where $\hat{e} = YY' - \Lambda\mathrm{E}[\eta|Y,\xi^{(\ell)}]Y' - Y\mathrm{E}[\eta|Y,\xi^{(\ell)}]'\Lambda' + \Lambda\mathrm{E}[\eta\eta'|Y,\xi^{(\ell)}]\Lambda'$. In matrix notation,

$$\frac{\partial^2 Q}{\partial \text{vec}(\lambda)\partial \text{vec}'(\lambda)} = -\triangle'(\lambda)\left(\text{E}[\eta\eta'|Y,\xi^{(\ell)}] \otimes \Phi^{-1}\right)\triangle(\lambda) \tag{5.59}$$

$$\frac{\partial^2 Q}{\partial \text{vec}(\phi)\partial \text{vec}'(\phi)} = \frac{1}{2}\triangle'(\phi)\left[\left(\Phi^{-1}(\Phi - 2\hat{e})\Phi^{-1}\right) \otimes \Phi^{-1}\right]\triangle(\phi) \tag{5.60}$$

$$\frac{\partial^2 Q}{\partial \text{vec}(\lambda)\partial \text{vec}'(\phi)} = \triangle'(\lambda)\Bigg[\left(\text{E}[\eta\eta'|Y,\xi^{(\ell)}]\Lambda'\Phi^{-1} - \text{E}[\eta|Y,\xi^{(\ell)}]Y'\Phi^{-1}\right) \otimes \Phi^{-1}\Bigg]\triangle(\phi) \tag{5.61}$$

In sum, the gradient vector and the Hessian matrix of the target function in (5.8) with respect to a vector of parameters of the latent VARMA(p, q) process, γ, are given in (5.50) and (5.51), respectively. The corresponding gradient vector and the Hessian matrix with respect to free parameters in Λ and Φ of the measurement model are given in (5.55) and (5.59)-(5.61), respectively. These vectors and matrices constitute the gradient vector and the Hessian matrix of the target function with respect to ξ, the vector of all free parameters in the model, i.e. Λ, Φ, and τ.

Specifically, let $\frac{\partial Q(\xi|\xi^{(\ell)})}{\partial \xi}$ and $\frac{\partial^2 Q(\xi|\xi^{(\ell)})}{\partial \xi \partial \xi'}$ denote the gradient vector and the Hessian matrix of the target function, respectively. Then, they are given as

$$\frac{\partial Q(\xi|\xi^{(\ell)})}{\partial \xi} = \begin{pmatrix} \frac{\partial Q(\xi|\xi^{(\ell)})}{\partial \gamma} \\ \frac{\partial Q(\xi|\xi^{(\ell)})}{\partial \text{vec}(\lambda)} \\ \frac{\partial Q(\xi|\xi^{(\ell)})}{\partial \text{vec}(\phi)} \end{pmatrix} \tag{5.62}$$

$$\frac{\partial^2 Q(\xi|\xi^{(\ell)})}{\partial \xi \partial \xi'} = \begin{pmatrix} \frac{\partial^2 Q(\xi|\xi^{(\ell)})}{\partial \gamma \partial \gamma'} & 0 & 0 \\ 0 & \frac{\partial^2 Q(\xi|\xi^{(\ell)})}{\partial \text{vec}(\lambda)\partial \text{vec}(\lambda)'} & \frac{\partial^2 Q(\xi|\xi^{(\ell)})}{\partial \text{vec}(\lambda)\partial \text{vec}(\phi)'} \\ 0 & \frac{\partial^2 Q(\xi|\xi^{(\ell)})}{\partial \text{vec}(\phi)\partial \text{vec}(\lambda)'} & \frac{\partial^2 Q(\xi|\xi^{(\ell)})}{\partial \text{vec}(\phi)\partial \text{vec}(\phi)'} \end{pmatrix} \tag{5.63}$$

where the sub-vectors and sub-matrices in (5.62) and (5.63) are given by equations (5.50), (5.51), (5.55), and (5.59)-(5.61).

Finally, the the M-step, the updating of the current parameter estimates, $\xi^{(\ell)}$ by maximizing the target function in (5.8), is carried out by calculating

$$\xi^{(\ell+1)} = \xi^{(\ell)} - \left[\frac{\partial^2 Q(\xi|\xi^{(\ell)})}{\partial\xi\partial\xi'}\right]^{-1} \frac{\partial Q(\xi|\xi^{(\ell)})}{\partial\xi} \tag{5.64}$$

The updated parameter estimates, $\xi^{(\ell+1)}$ are then regarded as the current parameter estimates at the next iteration and used to re-estimate the unknown quantities in (5.9), (5.10), and (5.11), which, in turn, will be substituted in the target function to complete the subsequent E-step. These two steps, i.e. the E-step and M-step, alternate until some convergence criterion is satisfied. A standard criterion for deterministic EM algorithm is to stop iterations when the relative change in the parameter estimates (or target function values) from successive iteration is smaller than a pre-specified value.

Chapter 6

Estimation of Standard Errors for Parameter Estimates

In practice, getting maximum likelihood estimates is not the final end of the model fitting and statistical inference process. The sampling variability of the parameter estimates needs to be estimated for statistical inferences and one way of addressing this issue is to compute the asymptotic covariance matrix of the parameter estimates. One of the early criticisms of the EM approach is that, unlike the Newton-Raphson and related methods directly maximizing the observed-data log-likelihood function, the EM algorithm does not automatically produce an estimate of the covariance matrix of the maximum likelihood estimates. The proposed algorithm is not free from this criticism, either. In this chapter, a new method to obtain standard error estimates is proposed. Before describing the details of the proposed method, the relevant theory and previous investigations are briefly described.

6.1 The principle of missing information

Using the same notation as the previous chapters, let (Y, η) be the complete data constructed by supposing η is observed, where Y and η represent random vectors of observed variables and unobserved common factors, respectively. Notice that $p(Y, \eta|\xi) = p(Y|\xi)p(\eta|Y, \xi)$, which implies that

$$-\frac{\partial^2}{\partial\xi\partial\xi'} \ln p(Y|\xi) = -\frac{\partial^2}{\partial\xi\partial\xi'} \ln p(Y, \eta|\xi) + \frac{\partial^2}{\partial\xi\partial\xi'} \ln p(\eta|Y, \xi) \tag{6.1}$$

where ξ represents the vector of model parameters indexing the log-likelihood function. Integrating both sides with respect to $p(\eta|Y, \hat{\xi})$, we can obtain

$$-\frac{\partial^2}{\partial\xi\partial\xi'} \ln p(Y|\xi) = -\frac{\partial^2}{\partial\xi\partial\xi'} Q(\xi|\hat{\xi}) + \frac{\partial^2}{\partial\xi\partial\xi'} H(\xi|\hat{\xi}) \tag{6.2}$$

where

$$Q(\xi|\hat{\xi}) = \int \ln p(Y, \eta|\xi) p(\eta|Y, \hat{\xi}) \, \mathrm{d}\eta \tag{6.3}$$

$$H(\xi|\hat{\xi}) = \int \ln p(\eta|Y, \xi) p(\eta|Y, \hat{\xi}) \, \mathrm{d}\eta \tag{6.4}$$

The equation (6.2) is known as the Missing Information Principle (Louis, 1982; Orchard & Woodbury, 1972). Let

$$\mathcal{I}(\xi|Y) = -\frac{\partial^2 \ln p(Y|\xi)}{\partial\xi\partial\xi'}, \; \mathcal{I}_c(\xi|Y) = -\frac{\partial^2 Q(\xi|\hat{\xi})}{\partial\xi\partial\xi'}, \; \mathcal{I}_m(\xi|Y) = -\frac{\partial^2}{\partial\xi\partial\xi'} H(\xi|\hat{\xi}) \tag{6.5}$$

be the observed data information, the complete data information and the missing information matrix, respectively. Then, the Missing Information Principle can be expressed as

$$\begin{aligned} \mathcal{I}(\xi|Y) &= \mathcal{I}_c(\xi|Y) - \mathcal{I}_m(\xi|Y) \\ &= \{I_q - \nabla(\xi)\} \, \mathcal{I}_c(\xi|Y) \end{aligned} \tag{6.6}$$

where $\nabla(\xi) = \mathcal{I}_m(\xi|Y)\mathcal{I}_c^{-1}(\xi|Y)$, known as the fraction of missing information (Dempster et al., 1977) and I_q is a $q \times q$ identity matrix. Intuitively, this principle means that the information contained in the complete data is greater than the information in the observed data by the amount of missing information. In other words, the observed information can be computed by subtracting the missing information from the complete data information. Alternatively, the observed information can be obtained by adjusting out the fraction of missing information from the compete data information.

6.2 Previous studies on standard error estimation

The large-sample theory explains that the inverse of the observed information matrix provides the asymptotic variance-covariance matrix associated with the MLEs. Unlike the results obtained with the Newton-Raphson or Fisher Scoring algorithm, however, this matrix is not given as an automatic by-product, but must be derived separately under the EM algorithm. In some situations, arriving at the observed information by directly evaluating the Hessian matrix of the observed data log-likelihood may be very difficult for a given problem and hence alternative approaches have been proposed. Among others, major contributions were made by Louis (1982), Meng and Rubin (1991) and Oakes (1999). The idea behind these approaches is to find simpler formulae to compute the observed data information using the quantities available as a by-product of the EM algorithm. A key result obtained by Louis (1982) showed that

$$\mathcal{I}_m(\hat{\xi}|Y) = \text{var}\left(\frac{\partial \ln p(Y,\eta|\xi)}{\partial \xi}\middle| Y, \hat{\xi}\right)\Bigg|_{\xi=\hat{\xi}} \tag{6.7}$$

and, by treating both ξ and $\hat{\xi}$ as free variables, Oakes (1999) showed that,

$$\mathcal{I}_m(\hat{\xi}|Y) = -\frac{\partial^2 Q(\xi|\hat{\xi})}{\partial \xi \partial \hat{\xi}\prime}\Bigg|_{\xi=\hat{\xi}} \tag{6.8}$$

In other words, Louis (1982)'s method computes the missing information by calculating the conditional variance of the score function of the complete data log-likelihood, while Oakes (1999)'s method requires calculating the mixed second order derivative of $Q(\xi|\hat{\xi})$ with respect to ξ and $\hat{\xi}$ by considering both of them to be free variables.

Inspired by the result due to Dempster et al. (1977) that, in the neighborhood of $\hat{\xi}$, the Jacobian matrix of the *EM* map, $\xi \mapsto M(\xi)$, is equal to the faction of missing information,

$$\nabla(\hat{\xi}) = \frac{\partial M(\xi)}{\partial \xi}\Bigg|_{\xi=\hat{\xi}} \tag{6.9}$$

Meng and Rubin (1991) proposed the so-called Supplemented *EM* algorithm that approximates $\left.\frac{\partial M(\xi)}{\partial \xi}\right|_{\xi=\hat{\xi}}$ by numerically differentiating the *EM* map, thereby augments the original *EM* algorithm by computing the observed data information matrix.

These methods, proposed to provide general frameworks for calculating observed data information matrix, commonly require computing the complete data information matrix. The key difference lies in the way the missing information is calculated, suggesting that the choice of a particular method largely depends on the relative simplicity of deriving and implementing different formulae for a given problem.

In the current context of PFA modeling, computing the conditional variance in (6.7) or the mixed second order derivative in (6.8) is not as simple a problem as the original authors suggested due in large part to the complexity of multidimensional optimization problem. The method by Meng and Rubin (1991) can be of use in such cases, but as the original authors pointed out (Meng & Rubin, 1991, p.903), the estimated observed data information matrix may be asymmetric, which is not an acceptable property for a covariance matrix.

In the next section, yet another alternative method for computing the missing information is proposed. This method takes a full advantage of the fact that the complete data information matrix is already available from the proposed EM algorithm. Only the missing information matrix is computed by directly taking the second order derivatives of $H(\xi|\hat{\xi})$ in (6.2). Surprisingly, for PFA models, this method turns out to be simpler than the aforementioned Louis' and Oakes' method in computing the missing information matrix. This method, unlike the Supplemented EM method, guarantees a symmetric observed data information matrix.

6.3 The derivation of the missing information matrix

The large-sample variance-covariance matrix of the parameter estimates can be calculated by inverting Equation (6.2), whose square root of diagonal elements provide the associated standard error estimates. Note that the first term of the right-hand-side in (6.2) was already derived in equations (5.50), (5.59), (5.60), (5.61), and (5.64). Therefore, the second term needs to be obtained. In the current context,

$$\ln p(\eta|Y,\xi) = -\frac{mT}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega(\xi)| - \frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)^{-1}G(\xi)\right\} \tag{6.10}$$

where $G(\xi) = (\eta - \mu(\xi))(\eta - \mu(\xi))'$, $\mu(\xi) = \mathrm{E}[\eta|Y,\xi] = \Omega(\xi)\Lambda'\Phi^{-1}Y$, and $\Omega(\xi) = (\Lambda\Phi^{-1}\Lambda + \Sigma(\tau)^{-1})^{-1}$. Notice that

$$\begin{aligned}
\frac{\partial \ln p(\eta|Y,\xi)}{\partial \xi_r} &= -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)^{-1}\left(\frac{\partial\Omega(\xi)}{\partial\xi_r}\right) + \left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_r}\right)G(\xi) + \Omega^{-1}(\xi)\left(\frac{\partial G(\xi)}{\partial\xi_r}\right)\right\} \\
&= \frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_r}\right)\right\} - \frac{1}{2}\mathrm{tr}\left\{G(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_r}\right)\right\} \\
&\quad + \mathrm{tr}\left\{\Omega^{-1}(\xi)(\eta-\mu(\xi))\left(\frac{\partial\mu(\xi)}{\partial\xi_r}\right)'\right\}
\end{aligned} \tag{6.11}$$

and

$$\begin{aligned}
\frac{\partial^2 \ln p(\eta|Y,\xi)}{\partial \xi_r\partial\xi_s} &= -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_s}\right)\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_r}\right)\right\} + \frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left(\frac{\partial^2\Omega^{-1}(\xi)}{\partial\xi_r\partial\xi_s}\right)\right\} \\
&\quad - \frac{1}{2}\mathrm{tr}\left\{\left(\frac{\partial G(\xi)}{\partial\xi_s}\right)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_r}\right)\right\} - \frac{1}{2}\mathrm{tr}\left\{G(\xi)\left(\frac{\partial^2\Omega^{-1}(\xi)}{\partial\xi_r\partial\xi_s}\right)\right\} \\
&\quad + \mathrm{tr}\left\{\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_s}\right)(\eta-\mu(\xi))\left(\frac{\partial\mu(\xi)}{\partial\xi_r}\right)'\right\} \\
&\quad - \mathrm{tr}\left\{\Omega^{-1}(\xi)\left(\frac{\partial\mu(\xi)}{\partial\xi_s}\right)\left(\frac{\partial\mu(\xi)}{\partial\xi_r}\right)'\right\} \\
&\quad + \mathrm{tr}\left\{\Omega^{-1}(\xi)(\eta-\mu(\xi))\left(\frac{\partial^2\mu(\xi)}{\partial\xi_r\partial\xi_s}\right)'\right\}
\end{aligned} \tag{6.12}$$

Therefore, taking expectations of (6.12) with respect to $p(\eta|Y,\hat{\xi})$ yields the second term in (6.2). Because $\mathrm{E}[(G(\xi)|Y,\hat{\xi})] = \Omega(\hat{\xi})$, where expectations are taken with respect to $p(\eta|Y,\hat{\xi})$, the typical element of the second term in (6.2) is given as,

$$\begin{aligned}
\frac{\partial^2}{\partial\xi_r\partial\xi_s}H(\xi|\hat{\xi}) &= -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_s}\right)\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\xi_r}\right)\right\} \\
&\quad - \mathrm{tr}\left\{\Omega^{-1}(\xi)\left(\frac{\partial\mu(\xi)}{\partial\xi_s}\right)\left(\frac{\partial\mu(\xi)}{\partial\xi_r}\right)'\right\}
\end{aligned} \tag{6.13}$$

Notice that

$$
\begin{aligned}
\frac{\partial \Omega^{-1}(\xi)}{\partial \lambda_r} &= \left(\frac{\partial \Lambda}{\partial \lambda_r}\right)' \Phi^{-1}\Lambda + \Lambda'\Phi^{-1}\left(\frac{\partial \Lambda}{\partial \lambda_r}\right) \\
\frac{\partial \Omega^{-1}(\xi)}{\partial \phi_r} &= -\Lambda'\Phi^{-1}\left(\frac{\partial \Phi}{\partial \phi_r}\right)\Phi^{-1}\Lambda \\
\frac{\partial \Omega^{-1}(\xi)}{\partial \gamma_r} &= -\Sigma_\tau^{-1}\left(\frac{\partial \Sigma_\tau}{\partial \gamma_r}\right)\Sigma_\tau^{-1} \\
\frac{\partial \mu(\xi)}{\partial \lambda_r} &= -\Omega(\xi)\left(\frac{\partial \Omega^{-1}}{\partial \lambda_r}\right)\Omega(\xi)\Lambda'\Phi^{-1}Y + \Omega(\xi)\left(\frac{\partial \Lambda}{\partial \lambda_r}\right)\Phi^{-1}Y \\
\frac{\partial \mu(\xi)}{\partial \phi_r} &= \Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial \Phi}{\partial \phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}Y - \Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial \Phi}{\partial \phi_r}\right)\Phi^{-1}Y \\
\frac{\partial \mu(\xi)}{\partial \gamma_r} &= \Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial \Sigma_\tau}{\partial \gamma_r}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}Y \qquad (6.14)
\end{aligned}
$$

Inserting (6.14) into (6.13) yields the closed form expressions of typical elements of the missing information matrix, the details of which are given in the Appendix B. Putting these together, the large-sample variance-covariance matrix of the MLEs can be calculated by substituting (5.51) and (6.13) into (6.2) and inverting.

Chapter 7

A study of parameter recovery of the proposed EM Algorithm

In order to ascertain that the proposed EM algorithm and the associated gradient vector and Hessian matrix are suitable in maximizing the expected complete data log-likelihood with its maximum achieved at MLE, a simulation study was conducted with the known values of parameters. The procedure was as follows: a sample of size equal to N was generated from the model-implied population covariance structure of a PFA(p, q) model in which multivariate latent process variables follow a stationary VARMA(p, q) process with each latent process variable being measured by multiple indicators. Then, a PFA(p, q) model was fitted to the simulated data-set using the proposed methods for obtaining MLEs and the associated standard error estimates. This procedure was repeated 1,000 times.

Because both the conditional distribution of Y given η and the marginal distribution of η are assumed to be multivariate normal, the marginal distribution of Y is also multivariate normal, whose mean and covariance matrix can be easily calculated as follows:

$$\mathrm{E}\left[Y\right] = \mathrm{E}\left[\mathrm{E}(Y|\eta)\right] = \mathrm{E}\left[\Lambda\eta\right] = \Lambda\mathrm{E}\left[\eta\right] = 0 \tag{7.1}$$

$$\begin{aligned} \mathrm{cov}\left[Y\right] &= \mathrm{cov}\left[\mathrm{E}(Y|\eta)\right] + \mathrm{E}\left[\mathrm{cov}(Y|\eta)\right] \\ &= \mathrm{cov}\left[\Lambda\eta\right] + \mathrm{E}\left[\Phi\right] \\ &= \Lambda\Sigma(\tau)\Lambda' + \Phi \end{aligned} \tag{7.2}$$

A sample of size equal to N was generated by Cholesky decomposition of the above model-implied population covariance matrix of Y.

7.1 The first case: repeated time-series with $N = 500$ and $T = 5$

In this case, the PFA model was specified setting $p = q = 1$, $m = 2$, $N = 500$, and $T = 5$. Each of the two factors had 4 variables as its indicators so that there were a total of 40 manifest variables in the model. Thus, a sample of $N = 500$ was generated by Cholesky decomposition of the 40×40 population covariance matrix of Y. The population values of model parameters were specified as follows:

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \begin{pmatrix} 1 & 0 \\ .8 & 0 \\ .8 & 0 \\ .8 & 0 \\ 0 & 1 \\ 0 & .7 \\ 0 & .5 \\ 0 & .5 \end{pmatrix} \tag{7.3}$$

$$\phi_{11} = \phi_{22} = \phi_{33} = \phi_{44} = \phi_{55} = \begin{pmatrix} .8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .8 \end{pmatrix} \tag{7.4}$$

The off-diagonal values of Φ were specified as follows[1]:

[1] Φ is a symmetric matrix in which off-diagonal elements in the upper triangular part have the same pattern and values corresponding to the lower triangular part specified in (7.5) to (7.8).

$$\Phi_{(9,2)} = \Phi_{(17,10)} = \Phi_{(25,18)} = \Phi_{(33,26)} = .8 \times \frac{1}{4} \tag{7.5}$$

$$\Phi_{(17,2)} = \Phi_{(25,10)} = \Phi_{(33,18)} = .8 \times \left(\frac{1}{4}\right)^2 \tag{7.6}$$

$$\Phi_{(25,2)} = \Phi_{(33,10)} = .8 \times \left(\frac{1}{4}\right)^3 \tag{7.7}$$

$$\Phi_{(33,2)} = .8 \times \left(\frac{1}{4}\right)^4 \tag{7.8}$$

The autoregressive coefficients, moving average coefficients and random-shock covariance matrices in the latent time series model were specified as

$$A_1 = \begin{pmatrix} .7 & .3 \\ .2 & .4 \end{pmatrix} \tag{7.9}$$

$$B_1 = \begin{pmatrix} .5 & 0 \\ 0 & .2 \end{pmatrix} \tag{7.10}$$

$$\Psi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{7.11}$$

The Newton-Raphson algorithm was employed for the *M-step*, in which the derived analytic expressions of the gradient vector and the Hessian matrix was used. The PFA(1,1) model was fitted to a sample of size equal to 500, which was repeated 1,000 times. For model identification, the factor loadings of the first items in each measurement occasion are fixed to unity, e.g. $\Lambda_{(1,1)} = \Lambda_{(9,3)} = \Lambda_{(17,5)} = \Lambda_{(25,7)} = \Lambda_{(32,9)} = 1.$

Based on the definition of the random-shock, the cross-lagged effect of random shock was constrained to be zero, e.g. $B_{1(1,2)} = B_{1(2,1)} = 0$. The equality constraints were also imposed such that measurement invariance held across time for all factor loadings. The unique variances were also set to be equal across time and covariances among specific factors were constrained in such a way that the equality relations specified in equations (7.5) to (7.8) held. These constraints may not be realistic in practical situations but the goal of imposing these constraints was to make sure that the proposed EM algorithm was suitable for optimization with such constraints.

The Tables 7.1 and 7.2 show the results obtained by the proposed EM algorithm. The first and second column show the name of parameters and the associated population values. A random sample of size N equal to 500 was generated and fit a PFA(1,1) model to the generated sample data using the proposed EM algorithm in order to obtain maximum likelihood estimates. The same procedure was repeated 1,000 times. The third to eighth columns show the minimum (Min.), the 25th percentile (Q_1), median, mean, the 75th percentile (Q_3), and maximum (Max.) of the parameter estimates obtained from 1,000 replications, respectively.

Across all parameter estimates, there is a hint of bias in that the means of 1,000 replications are not exactly equal to the corresponding population values. However, the large-sample theory explains that the maximum likelihood estimator is unbiased only asymptotically i.e. as sample size tends to infinity. Because the sample size of the current simulation study is only finite i.e. $N = 500$, biased estimates can be expected. In fact, it can be seen that the central tendencies reflected in the mean and median of the parameter estimates from the 1,000 replications closely match the population values. The dispersions reflected in the interquartile range ($Q_3 - Q_1$) are also very small. The maximum and minimum of parameter estimates among 1,000 replications do not seem to show any erratic behavior of the proposed algorithm.

The Tables 7.3 and 7.4 show the performance of the proposed standard error estimation method. The number in each cell from the third to eighth column represents the minimum (Min.), the 25th percentile (Q_1), median, mean, the 75th percentile (Q_3), and maximum (Max.) of the standard error estimates over the 1,000 replications, respectively. The target values to be compared are the corresponding standard deviations (SD) in the second column, which shows the actual variability of the parameter estimates. It can be seen that the means and medians of standard error estimates are almost identical with the corresponding standard deviations to the third decimal point across all parameters. The dispersions of the estimates reflected in the interquartile range ($Q_3 - Q_1$) are also small across all estimates. Although the respective maximums of standard error estimates for $\beta_{1,11}$ (0.382) and ψ_{11} (0.269) appear to be unusually large, this may simply reflect the fact that relatively larger sampling errors may have been involved in particular replications than in others. This argument can be supported by the fact that those two maximums came from the same replication (the 671^{th} replication out of 1,000), which yielded the respective maximum value (.881) of MLE for $\beta_{1,11}$ (Table 7.1) and the minimum value (.712) of MLE for ψ_{11} (Table 7.2), respectively. Except for this particular replication, in general, the maximum and minimum of standard error estimates among 1,000 replications did not show any erratic behavior of the proposed standard error estimation method. The benefit of having standard error estimates is clear. They provide a measure of uncertainty associated with the parameter estimates and allow for the construction of asymptotic confidence intervals.

7.2 The second case: repeated time-series with $N = 10$ and $T = 50$

In this study, the PFA model was specified setting $p = q = 1$, $m = 2$, $N = 10$, and $T = 50$. The population parameter values were the same as the first case, except for off-diagonal elements in Φ constrained to be zero. Such constraints can be easily relaxed if necessary. The procedures to generate data were the same as the first case.

Tables 7.5, 7.6, 7.7, and 7.8 show the performances of the proposed methods for obtaining MLEs and the associated standard error estimates. For simpler presentations, the parameters constrained to be equal across measurement occasions were presented as single parameters. For example, $\lambda_{1,km} = \ldots = \lambda_{50,km}$ are represented as λ_{km}, and $\phi_{1,kk} = \ldots, = \phi_{50,kk}$ are represented as ϕ_{kk} ($m = 1, 2$ and $k = 1, \ldots, 8$).

As the previous results, the summary statistics for MLEs and standard error estimates do not show any erratic behaviors of maximum likelihood estimators. The mean and median of MLEs obtained from 1,000 replications reasonably well match the population values, and the standard error estimates also appear to adequately reflect the actual variability of the parameter estimates over repeated sampling. Again, a hint of bias in parameter estimates and standard error estimates can be seen, but such results are expected under finite sample size, i.e. $N = 10$ in this case. Across all parameter estimates and the standard error estimates, the interquartile ranges and the ranges across all parameter estimates and the standard error estimates are reasonably narrow. As inferred in the first case, very deviant-looking maximum or minimum values may simply reflect relatively large sampling errors of a particular replication.

7.3 The third case: single-subject time series with $N = 1$ and $T = 100$

In this study, the PFA model was specified setting $p = q = 1$, $m = 2$, $N = 1$, and $T = 100$. The population parameter values are the same as the first case, except for Φ constrained to be a diagonal matrix as in the second case. The procedures to generate data were the same as the previous cases.

Tables 7.9, 7.10, 7.11, and 7.12 show the performances of of the proposed methods for obtaining MLEs and the associated standard error estimates. The mean and median of MLEs obtained from 1,000 replications are reasonably close to the population values, and the standard error estimates seem to reasonably well reflect the actual variability of the parameter estimates over repeated sampling. In this study, there is a clearer sign of bias in parameter estimates and standard error estimates, however, such results are expected under the condition of small finite sample size, i.e. $N = 1$ in this case. The interquartile ranges and the ranges are also reasonably narrow across all parameter estimates and the associated standard error estimates. Again, very deviant-looking maximum or minimum values may simply reflect the large sampling errors of a particular replication, contributing to the bias of parameter and standard error estimates.

Thus, the results of this study clearly show that the proposed EM algorithm and the proposed method for standard error estimation can be effectively employed for single-subject time-series data analysis.

Table 7.1: Summary of the maximum likelihood estimates of free parameters in Λ and Φ obtained from the simulation study (1,000 replications, T=5,N=500 PFA(1,1)).

	Value	Min.	Q_1	Median	Mean	Q_3	Max.
$\lambda_{1,11} = \ldots = \lambda_{5,11}$	1.000	-[a]	-	-	-	-	-
$\lambda_{1,21} = \ldots = \lambda_{5,21}$	0.800	0.763	0.792	0.800	0.800	0.807	0.833
$\lambda_{1,31} = \ldots = \lambda_{5,31}$	0.800	0.769	0.793	0.800	0.800	0.807	0.831
$\lambda_{1,41} = \ldots = \lambda_{5,41}$	0.800	0.769	0.792	0.800	0.799	0.807	0.833
$\lambda_{1,52} = \ldots = \lambda_{5,52}$	1.000	-	-	-	-	-	-
$\lambda_{1,62} = \ldots = \lambda_{5,62}$	0.700	0.648	0.687	0.700	0.700	0.711	0.754
$\lambda_{1,72} = \ldots = \lambda_{5,72}$	0.500	0.457	0.490	0.500	0.501	0.511	0.548
$\lambda_{1,82} = \ldots = \lambda_{5,82}$	0.500	0.456	0.489	0.500	0.500	0.511	0.553
$\phi_{1,11} = \ldots = \phi_{5,11}$	0.800	0.694	0.778	0.798	0.798	0.819	0.918
$\phi_{1,22} = \ldots = \phi_{5,22}$	0.800	0.715	0.780	0.800	0.800	0.820	0.901
$\phi_{1,33} = \ldots = \phi_{5,33}$	0.800	0.716	0.780	0.798	0.798	0.816	0.880
$\phi_{1,44} = \ldots = \phi_{5,44}$	0.800	0.718	0.782	0.797	0.800	0.819	0.889
$\phi_{1,55} = \ldots = \phi_{5,55}$	0.800	0.678	0.772	0.800	0.799	0.825	0.938
$\phi_{1,66} = \ldots = \phi_{5,66}$	0.800	0.709	0.781	0.800	0.801	0.820	0.889
$\phi_{1,77} = \ldots = \phi_{5,77}$	0.800	0.717	0.784	0.800	0.801	0.818	0.876
$\phi_{1,88} = \ldots = \phi_{5,88}$	0.800	0.726	0.783	0.800	0.800	0.818	0.886
$\phi_{12,22} = \phi_{23,22} = \phi_{34,22} = \phi_{45,22}$	0.200	0.110	0.184	0.199	0.200	0.215	0.279
$\phi_{13,22} = \phi_{24,22} = \phi_{35,22}$	0.050	-0.040	0.031	0.049	0.049	0.067	0.132
$\phi_{14,22} = \phi_{25,22}$	0.013	-0.101	-0.009	0.013	0.013	0.034	0.118
$\phi_{15,22}$	0.003	-0.135	-0.026	0.004	0.005	0.035	0.130

[a]The parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.2: Summary of the maximum likelihood estimates of free parameters in A_1, B_1, and Ψ obtained from the simulation study (1,000 replications, T=5, N=500, PFA(1,1)).

	Value	Min.	Q_1	Median	Mean	Q_3	Max.
$\alpha_{1,11}$	0.700	0.645	0.685	0.699	0.698	0.712	0.749
$\alpha_{1,21}$	0.200	0.134	0.186	0.201	0.201	0.216	0.271
$\alpha_{1,12}$	0.300	0.207	0.282	0.301	0.301	0.320	0.402
$\alpha_{1,22}$	0.400	0.223	0.359	0.398	0.396	0.434	0.558
$\beta_{1,11}$	0.500	0.326	0.458	0.502	0.506	0.549	0.881
$\beta_{1,21}$	0.000	-[a]	-	-	-	-	-
$\beta_{1,12}$	0.000	-	-	-	-	-	-
$\beta_{1,22}$	0.200	-0.043	0.156	0.207	0.206	0.256	0.425
ψ_{11}	1.000	0.712	0.956	1.004	0.999	1.048	1.195
ψ_{21}	0.000	-	-	-	-	-	-
ψ_{12}	0.000	-	-	-	-	-	-
ψ_{22}	1.000	0.830	0.955	0.994	0.994	1.031	1.217

[a]The parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.3: Summary of the standard error estimates associated with free parameters in Λ and Φ obtained from the simulation study (1,000 replications, T=5, N=500, PFA(1,1)).

	SD	Min.	Q_1	Median	Mean	Q_3	Max.
$\lambda_{1,11} = \ldots = \lambda_{5,11}$	-[a]	-	-	-	-	-	-
$\lambda_{1,21} = \ldots = \lambda_{5,21}$	0.011	0.010	0.011	0.011	0.011	0.011	0.013
$\lambda_{1,31} = \ldots = \lambda_{5,31}$	0.010	0.009	0.010	0.010	0.010	0.010	0.011
$\lambda_{1,41} = \ldots = \lambda_{5,41}$	0.011	0.009	0.010	0.010	0.010	0.010	0.011
$\lambda_{1,52} = \ldots = \lambda_{5,52}$	-	-	-	-	-	-	-
$\lambda_{1,62} = \ldots = \lambda_{5,62}$	0.017	0.016	0.018	0.018	0.018	0.018	0.020
$\lambda_{1,72} = \ldots = \lambda_{5,72}$	0.016	0.014	0.015	0.016	0.016	0.016	0.018
$\lambda_{1,82} = \ldots = \lambda_{5,82}$	0.016	0.014	0.015	0.016	0.016	0.016	0.018
$\phi_{1,11} = \ldots = \phi_{5,11}$	0.031	0.028	0.031	0.031	0.031	0.032	0.035
$\phi_{1,22} = \ldots = \phi_{5,22}$	0.029	0.026	0.028	0.029	0.029	0.030	0.032
$\phi_{1,33} = \ldots = \phi_{5,33}$	0.026	0.025	0.027	0.027	0.027	0.028	0.030
$\phi_{1,44} = \ldots = \phi_{5,44}$	0.028	0.025	0.027	0.027	0.027	0.028	0.030
$\phi_{1,55} = \ldots = \phi_{5,55}$	0.040	0.036	0.039	0.039	0.039	0.040	0.043
$\phi_{1,66} = \ldots = \phi_{5,66}$	0.028	0.026	0.028	0.029	0.029	0.029	0.031
$\phi_{1,77} = \ldots = \phi_{5,77}$	0.026	0.023	0.025	0.025	0.025	0.026	0.027
$\phi_{1,88} = \ldots = \phi_{5,88}$	0.026	0.023	0.025	0.025	0.025	0.026	0.028
$\phi_{12,22} = \phi_{23,22} = \phi_{34,22} = \phi_{45,22}$	0.024	0.020	0.023	0.023	0.024	0.024	0.026
$\phi_{13,22} = \phi_{24,22} = \phi_{35,22}$	0.026	0.023	0.025	0.026	0.026	0.026	0.028
$\phi_{14,22} = \phi_{25,22}$	0.032	0.028	0.030	0.031	0.031	0.032	0.034
$\phi_{15,22}$	0.044	0.040	0.043	0.044	0.044	0.045	0.048

[a]The standard deviations of parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.4: Summary of the standard error estimates associated with free parameters in A_1, B_1, and Ψ obtained from the simulation study (1,000 replications, T=5, N=500, PFA(1,1)).

	SD	Min.	Q_1	Median	Mean	Q_3	Max.
$\alpha_{1,11}$	0.019	0.017	0.019	0.020	0.020	0.020	0.022
$\alpha_{1,21}$	0.021	0.017	0.019	0.020	0.020	0.021	0.024
$\alpha_{1,12}$	0.030	0.027	0.030	0.031	0.031	0.031	0.035
$\alpha_{1,22}$	0.054	0.044	0.050	0.052	0.052	0.054	0.063
$\beta_{1,11}$	0.068	0.051	0.062	0.068	0.071	0.075	0.382
$\beta_{1,21}$	-[a]	-	-	-	-	-	-
$\beta_{1,12}$	-	-	-	-	-	-	-
$\beta_{1,22}$	0.077	0.065	0.070	0.072	0.072	0.073	0.082
ψ_{11}	0.068	0.059	0.065	0.068	0.070	0.073	0.269
ψ_{21}	-	-	-	-	-	-	-
ψ_{12}	-		-	-	-	-	-
ψ_{22}	0.056	0.050	0.054	0.055	0.055	0.056	0.062

[a]The standard deviations of parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.5: Summary of the maximum likelihood estimates of free parameters in Λ and Φ obtained from the simulation study (1,000 replications, T=50, N=10, PFA(1,1)).

	Value	Min.	Q_1	Median	Mean	Q_3	Max.
λ_{11}	1.000	-[a]	-	-	-	-	-
λ_{21}	0.800	0.750	0.784	0.797	0.799	0.814	0.865
λ_{31}	0.800	0.738	0.789	0.799	0.799	0.811	0.855
λ_{41}	0.800	0.755	0.788	0.803	0.802	0.816	0.859
λ_{52}	1.000	-	-	-	-	-	-
λ_{62}	0.700	0.629	0.679	0.700	0.704	0.726	0.814
λ_{72}	0.500	0.424	0.473	0.499	0.499	0.528	0.582
λ_{82}	0.500	0.420	0.479	0.504	0.501	0.523	0.572
ϕ_{11}	0.800	0.627	0.741	0.785	0.791	0.833	0.975
ϕ_{22}	0.800	0.646	0.754	0.795	0.793	0.830	0.960
ϕ_{33}	0.800	0.657	0.763	0.796	0.800	0.834	0.985
ϕ_{44}	0.800	0.644	0.750	0.800	0.796	0.840	0.993
ϕ_{55}	0.800	0.589	0.752	0.797	0.798	0.845	1.068
ϕ_{66}	0.800	0.645	0.765	0.808	0.805	0.847	0.987
ϕ_{77}	0.800	0.671	0.756	0.803	0.804	0.847	0.969
ϕ_{88}	0.800	0.642	0.752	0.788	0.786	0.820	0.938

[a]The parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.6: Summary of the maximum likelihood estimates of free parameters in A_1, B_1, and Ψ obtained from the simulation study (1,000 replications, T=50, N=10, PFA(1,1)).

	Value	Min.	Q_1	Median	Mean	Q_3	Max.
$\alpha_{1,11}$	0.700	0.539	0.670	0.696	0.698	0.729	0.794
$\alpha_{1,21}$	0.200	0.103	0.172	0.202	0.203	0.224	0.358
$\alpha_{1,12}$	0.300	0.102	0.264	0.312	0.306	0.348	0.498
$\alpha_{1,22}$	0.400	0.018	0.331	0.396	0.391	0.467	0.590
$\beta_{1,11}$	0.500	0.220	0.416	0.484	0.473	0.543	0.812
$\beta_{1,21}$	0.000	-[a]	-	-	-	-	-
$\beta_{1,12}$	0.000	-	-	-	-	-	-
$\beta_{1,22}$	0.200	-0.064	0.124	0.215	0.216	0.296	0.677
ψ_{11}	1.000	0.780	0.959	1.005	1.021	1.086	1.318
ψ_{21}	0.000	-	-	-	-	-	-
ψ_{12}	0.000	-	-	-	-	-	-
ψ_{22}	1.000	0.733	0.904	0.981	0.990	1.083	1.278

[a]The parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.7: Summary of the standard error estimates associated with free parameters in Λ and Φ obtained from the simulation study (1,000 replications, T=50, N=10, PFA(1,1)).

	SD	Min.	Q_1	Median	Mean	Q_3	Max.
λ_{11}	-[a]	-	-	-	-	-	-
λ_{21}	0.021	0.018	0.021	0.022	0.022	0.023	0.027
λ_{31}	0.021	0.018	0.021	0.022	0.022	0.023	0.028
λ_{41}	0.022	0.018	0.021	0.022	0.022	0.023	0.028
λ_{52}	-	-	-	-	-	-	-
λ_{62}	0.035	0.032	0.038	0.040	0.040	0.042	0.048
λ_{72}	0.036	0.028	0.033	0.035	0.035	0.037	0.042
λ_{82}	0.033	0.028	0.033	0.035	0.035	0.037	0.041
ϕ_{11}	0.069	0.057	0.064	0.067	0.067	0.070	0.077
ϕ_{22}	0.059	0.050	0.058	0.060	0.060	0.062	0.070
ϕ_{33}	0.061	0.052	0.058	0.060	0.060	0.062	0.072
ϕ_{44}	0.065	0.050	0.057	0.060	0.060	0.063	0.073
ϕ_{55}	0.080	0.075	0.082	0.085	0.085	0.089	0.099
ϕ_{66}	0.065	0.053	0.061	0.064	0.064	0.066	0.075
ϕ_{77}	0.061	0.048	0.054	0.056	0.056	0.059	0.067
ϕ_{88}	0.053	0.047	0.053	0.055	0.055	0.057	0.064

[a]The standard deviations of parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.8: Summary of the standard error estimates associated with free parameters in A_1, B_1, and Ψ obtained from the simulation study (1,000 replications, T=50, N=10, PFA(1,1)).

	SD	Min.	Q_1	Median	Mean	Q_3	Max.
$\alpha_{1,11}$	0.041	0.034	0.038	0.040	0.040	0.043	0.051
$\alpha_{1,21}$	0.043	0.029	0.037	0.039	0.040	0.043	0.055
$\alpha_{1,12}$	0.066	0.050	0.059	0.063	0.063	0.067	0.077
$\alpha_{1,22}$	0.100	0.070	0.087	0.094	0.095	0.102	0.127
$\beta_{1,11}$	0.098	0.081	0.105	0.118	0.122	0.134	0.312
$\beta_{1,21}$	-[a]	-	-	-	-	-	-
$\beta_{1,12}$	-	-	-	-	-	-	-
$\beta_{1,22}$	0.133	0.106	0.124	0.129	0.130	0.135	0.157
ψ_{11}	0.103	0.109	0.128	0.135	0.137	0.144	0.292
ψ_{21}	-	-	-	-	-	-	-
ψ_{12}	-	-	-	-	-	-	-
ψ_{22}	0.121	0.097	0.112	0.119	0.118	0.124	0.137

[a]The standard deviations of parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.9: Summary of the maximum likelihood estimates of free parameters in Λ and Φ obtained from the simulation study (1,000 replications, T=100, N=1, PFA(1,1)).

	Value	Min.	Q_1	Median	Mean	Q_3	Max.
λ_{11}	1.000	-[a]	-	-	-	-	-
λ_{21}	0.800	0.686	0.767	0.795	0.800	0.836	0.908
λ_{31}	0.800	0.705	0.772	0.795	0.800	0.825	0.942
λ_{41}	0.800	0.678	0.772	0.807	0.806	0.841	0.908
λ_{52}	1.000	-	-	-	-	-	-
λ_{62}	0.700	0.476	0.631	0.705	0.702	0.770	1.048
λ_{72}	0.500	0.284	0.425	0.492	0.489	0.532	0.731
λ_{82}	0.500	0.265	0.428	0.482	0.487	0.538	0.645
ϕ_{11}	0.800	0.356	0.705	0.790	0.793	0.896	1.194
ϕ_{22}	0.800	0.466	0.662	0.778	0.771	0.854	1.071
ϕ_{33}	0.800	0.517	0.725	0.779	0.795	0.866	1.159
ϕ_{44}	0.800	0.505	0.687	0.785	0.786	0.858	1.197
ϕ_{55}	0.800	0.332	0.639	0.771	0.777	0.901	1.237
ϕ_{66}	0.800	0.402	0.656	0.768	0.761	0.852	1.022
ϕ_{77}	0.800	0.513	0.721	0.805	0.820	0.895	1.157
ϕ_{88}	0.800	0.505	0.691	0.770	0.779	0.859	1.038

[a]The parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.10: Summary of the maximum likelihood estimates of free parameters in A_1, B_1, and Ψ obtained from the simulation study (1,000 replications, T=100, N=1, PFA(1,1)).

	Value	Min.	Q_1	Median	Mean	Q_3	Max.
$\alpha_{1,11}$	0.700	0.462	0.639	0.720	0.701	0.759	0.867
$\alpha_{1,21}$	0.200	0.022	0.134	0.192	0.197	0.256	0.456
$\alpha_{1,12}$	0.300	-0.043	0.162	0.263	0.268	0.366	0.611
$\alpha_{1,22}$	0.400	-0.116	0.285	0.430	0.405	0.545	0.795
$\beta_{1,11}$	0.500	-0.104	0.280	0.400	0.375	0.491	0.776
$\beta_{1,21}$	0.000	-[a]	-	-	-	-	-
$\beta_{1,12}$	0.000	-	-	-	-	-	-
$\beta_{1,22}$	0.200	-0.463	-0.028	0.164	0.151	0.380	0.754
ψ_{11}	1.000	0.605	0.921	1.022	1.067	1.224	1.854
ψ_{21}	0.000	-	-	-	-	-	-
ψ_{12}	0.000	-	-	-	-	-	-
ψ_{22}	1.000	0.565	0.860	1.030	1.022	1.164	1.504

[a]The parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.11: Summary of the standard error estimates associated with free parameters in Λ and Φ obtained from the simulation study (1,000 replications, T=100, N=1, PFA(1,1)).

	SD	Min.	Q_1	Median	Mean	Q_3	Max.
λ_{11}	-[a]	-	-	-	-	-	-
λ_{21}	0.051	0.034	0.045	0.051	0.052	0.057	0.080
λ_{31}	0.047	0.037	0.045	0.051	0.053	0.059	0.076
λ_{41}	0.051	0.035	0.045	0.051	0.052	0.059	0.080
λ_{52}	-	-	-	-	-	-	-
λ_{62}	0.107	0.046	0.077	0.089	0.091	0.100	0.156
λ_{72}	0.084	0.049	0.072	0.080	0.081	0.089	0.132
λ_{82}	0.079	0.052	0.068	0.079	0.079	0.088	0.119
ϕ_{11}	0.153	0.102	0.138	0.149	0.151	0.166	0.206
ϕ_{22}	0.124	0.093	0.116	0.133	0.131	0.142	0.173
ϕ_{33}	0.135	0.093	0.124	0.131	0.134	0.145	0.185
ϕ_{44}	0.135	0.095	0.119	0.132	0.133	0.142	0.194
ϕ_{55}	0.188	0.122	0.176	0.188	0.191	0.205	0.273
ϕ_{66}	0.132	0.089	0.127	0.138	0.137	0.150	0.181
ϕ_{77}	0.135	0.096	0.115	0.127	0.128	0.138	0.175
ϕ_{88}	0.121	0.083	0.110	0.120	0.122	0.132	0.164

[a]The standard deviations of parameter estimates and the associated quantities are not available for fixed parameters.

Table 7.12: Summary of the standard error estimates associated with free parameters in A_1, B_1, and Ψ obtained from the simulation study (1,000 replications, T=100, N=1, PFA(1,1)).

	SD	Min.	Q_1	Median	Mean	Q_3	Max.
$\alpha_{1,11}$	0.084	0.060	0.080	0.088	0.091	0.095	0.139
$\alpha_{1,21}$	0.094	0.031	0.075	0.087	0.088	0.101	0.200
$\alpha_{1,12}$	0.133	0.095	0.119	0.132	0.135	0.149	0.201
$\alpha_{1,22}$	0.202	0.108	0.170	0.199	0.207	0.238	0.356
$\beta_{1,11}$	0.168	0.154	0.197	0.224	0.239	0.271	0.500
$\beta_{1,21}$	-[a]	-	-	-	-	-	-
$\beta_{1,12}$	-	-	-	-	-	-	-
$\beta_{1,22}$	0.273	0.178	0.256	0.282	0.284	0.311	0.489
ψ_{11}	0.234	0.208	0.250	0.285	0.286	0.312	0.553
ψ_{21}	-	-	-	-	-	-	-
ψ_{12}	-	-	-	-	-	-	-
ψ_{22}	0.208	0.155	0.243	0.268	0.269	0.292	0.378

[a]The standard deviations of parameter estimates and the associated quantities are not available for fixed parameters.

Chapter 8

Discussion

8.1 Objectives accomplished

Objectives of this dissertation were as follows: 1) to develop and implement the *EM* algorithm to obtain maximum likelihood estimates for Process Factor Analysis models; 2) to derive closed-form expressions of the gradient vector and the Hessian matrix for implementing the M-step of the *EM* algorithm; 3) to test the viability of the proposed algorithm by applying to simulated data; 4) to develop and implement a method of standard error estimation in the context of the EM algorithm.

To accomplish these objectives, the target function *i.e.* the Q-function in the *EM* algorithm was constructed in such a way that the algorithm can be implemented in structural equation modeling framework. The analytic expressions of the gradient vector and the Hessian matrix of the Q-function were derived for the M-step. A simulation study was conducted to demonstrate the viability of the proposed algorithm in terms of the recovery of population parameter values and the results verified that the algorithm is viable. The results of the simulation study also showed the proposed method of standard error estimation fairly accurately quantifies the actual variability of the parameter estimates over repeated sampling. Therefore, the objectives of the project have been completed.

8.2 A methodological implication of the objectives accomplished

The results of the simulation study clearly demonstrated that the covariance structure in (4.10), the proposed *EM* algorithm and standard error estimation method can be employed not only for repeated time series data analysis, but also for single-subject time series data analysis.

Typically, the analysis of single-subject time series data requires a large number of measurements, e.g. $T = 100$, which will play the role of an effective sample size in parameter estimation, causing a sharp increase of the dimension of the covariance structure in (4.10). Although, with the rapid increase in computing power, today's personal computers can invert and multiply matrices of order of hundreds by hundreds without difficulty, differentiating a matrix of the same size is another matter, especially when the elements of the matrix are complex nonlinear functions of unknown variables, as in the covariance structure for a VARMA(p, q) process in (4.10) with $T = 100$.

It is worth noting that the analytic expressions of the gradient vector and the Hessian matrix derived in Chapter 5 (see Section 5.5) replace numerical differentiations required in the *M-step* with simple matrix multiplications and inversions, which can be handled without difficulty in personal computers. This simplification of the *M-step* and the consequent capability of the proposed *EM* algorithm to deal with a large covariance structure matrix constitute a major benefit of the objectives accomplished regarding the derivations of the closed-form expressions of the gradient vector and the Hessian matrix of the target function. The results of the simulation study also showed that standard error estimates associated with the parameter estimates can be obtained by simple matrix multiplications and inversions as the closed-form expressions are derived in Chapter 6 (see Section 6.3 and Appendix B).

8.3 A substantive implication of the objectives accomplished

In psychology, panel data with relatively large sample size, N, and a moderate number of measurement occasions, usually called *longitudinal data set*, are very common. The proposed *EM* algorithm is well suited for analyzing such data, providing a useful tool for fitting process-oriented time-series methods to longitudinal data. By applying time-series methods, researchers can empirically test the effect of a latent psychological construct X on another construct Y, taking account of the autoregressive effects of Y on itself, which is a widely known method of examining causality of X on Y in time series literature (Granger, 1969, p.431). The maximum likelihood estimates of such effects and the associated standard error estimates are provided by the proposed *EM* algorithm.

8.4 On the analysis of non-stationary data under the proposed algorithm

The proposed algorithm can be applied both to single-subject time series data analysis and to repeated time series data analysis. The single-subject time series data analysis assumes that the data be generated by a stationary process for the purpose of parameter estimation, and consequently the proposed algorithm can be directly applied. The proposed algorithm can also be employed for parameter estimation of repeated time-series data, under the assumption that each subject is an independent replication following the identical stationary time series.

However, either theories or data could reveal that stationarity is not a plausible constraint. A nice property of the proposed *EM* algorithm is that it has the capability to deal with non-stationary process of certain kinds. Such capability of the algorithm can be achieved by flexible adaptations of the initial status covariance matrix Θ specified in (4.11).

For example, when day to day variability of emotional processes of ordinary adults is of interest, it may be sensible to assume that the same stationary underlying process as the VARMA(p, q) process postulated in the model has started in the distant past and continued to the first measurement, thereby to adopt the covariance matrix of the initial status specified as in (4.11). However, when either substantive theory or the data suggest a change of emotional process as the first measurements were made, for example, as a result of promotion, marriage, etc., one can incorporate such hypothesis by specifying the initial status covariance matrix (denoted by Θ-free) as any positive definite matrix without any constraints imposed on elements of Θ. In a completely different context such as a learning experiment where subjects are required to conduct completely unfamiliar tasks, it may be a plausible hypothesis that there is no influence of prior experiences on the initial performance so that the covariance matrix of the initial status vector, Θ is set to be a null matrix i.e. $\Theta = 0$. In this model, the time series data are assumed to start with the first observation.

8.5 On relationships to the quasi-simplex model

The current model specification for PFA models given in Equations (4.1) and (4.2) provides a great flexibility as a general model containing a wide range of specific models as its special cases. Therefore, by allowing the parameters to be fixed, free, or constrained, the developed methods for the estimation of PFA models can be directly employed to obtain the maximum likelihood estimates and the associated standard error estimates for such special cases. A very important class of such special cases in the context of longitudinal data analysis include the quasi-simplex model.

Since the seminal paper of Guttman (1954) regarding simplex correlation structures to be constructed by tests ordered by their complexity, many investigators have studied so-called the simplex or the quasi-simplex model in terms of model formulation (Anderson, 1959), model identification and estimation (Jöreskog, 1970; Jöreskog & Sörbom, 1977), and applications to sociological panel data analysis (Heise, 1969; Wiley & Wiley, 1970). In particular, Heise (1969) and Wiley and Wiley (1970) proposed employing quasi-simplex models for separating reliability and stability in test-retest correlation, thereby obtaining reliability estimates adjusted for the attenunation caused by changes in true-scores across time. Most recently, Biemer, Christ, and Wiesen (2009) proposed a generalized simplex model (GSM) for estimating scale score reliability for panel survey data where true scores change over time.

It is worth noticing that the GSM proposed in Biemer et al. (2009, p.406) can be specified as a PFA(1,0) model with $T = 3$ where a single latent variable, measured by two manifest indicators ($m = 1$, $k = 2$), follows a non-stationary AR(1) process with time-varying autoregressive coefficients and random-shock variances with a null initial status covariance matrix. As special cases of the generalized simplex model, the quasi-simplex model and the simplex model can also be specified as special cases of PFA models. As such, the generality of PFA models, encompassing the aforementioned class of simplex models as special cases, suggests that the developed estimation methods for PFA models can be employed to achieve the same goal that the previous investigators have pursued, under different or more general conditions.

More specifically, by employing PFA models, further generalizations beyond the currently most general QSM approaches (Biemer et al., 2009) may be possible for the appropriate estimation of reliability in the context of longitudinal data analysis. In other words, PFA models can provide a way of obtaining reliability estimates under the situation where the true scores are multidimensional, measured by multiple indicators, and changing over time following a more complex process, e.g. a non-stationary VARMA(p, q) process. In particular, when the changes in true scores can justifably be assumed to be a stationary VARMA(p, q) process, the parameter estimation method proposed in this dissertion can provide an effective way of incorporating such processes for the reliability estimation. Furthermore, using the standard error estimates obtained by the proposed method, the possibility of quantifying the sampling variaibility of the reliability estimates in the context of longitudinal data analysis can be explored.

8.6 Future studies

Although the success of the proposed algorithm for maximum likelihood estimation of PFA models has been shown in previous chapters, there are a few important issues that remain to be addressed for the proposed algorithm to be adopted fully in stand-alone PFA model-fitting software and applied psychological research. Those issues include obtaining good starting values, incorporating comprehensive measurement models, and identifying the order of latent time series, among others. Here, possible ways open to addressing each issue are discussed.

8.6.1 Starting values

The proposed *EM* algorithm includes an iterative process of updating the 'current' parameter estimates in the next step until convergence is achieved. Thus, at the outset, some values need to be assigned as initial estimates of the parameters for the algorithm to operate on. It is generally advantageous to provide good initial estimates to facilitate the iterative procedures. In the simulation study of the proposed algorithm, the starting values were arbitrarily chosen close enough to the population parameter values because the purpose of the study was not in developing a procedure to find good starting values but in testing the viability of the proposed algorithm assuming reasonable starting values that will lead to the convergence of the algorithm. In this instance convergence was reached without difficulty. In empirical studies, however, population parameter values are not known. Thus, it is recommended to employ the best available starting values because convergence may not be attained or may be a very slow process with poor starting values. Extension of Yule-Walker equations to latent vector time series models will be explored as a method of determining reasonable starting values for the proposed algorithm[1]. The two stage least squares (2SLS) estimators (Bollen, 1996) or ordinary least squares estimator (Browne & Zhang, 2007a) can be considered as other candidates

[1]By personal communication with Dr. Stephen du Toit.

for non-iterative algorithms for providing starting values. Further research on this issue is necessary.

8.6.2 Comprehensive measurement models

In psychological research, observed data are not always suitable to be modeled with normal random vectors. Dichotomous responses, for example, produced by yes or no response choices, are very common. It is also not uncommon for each item in a Likert Scale, which is commonly used in questionnaires, to offer three to four alternatives reflecting level of agreement to a certain statement. In such cases, modeling the item responses according to continuous normal distributions is likely to distort the statistical inferences, especially when the normality assumption is not tenable (Browne, 1984). For example, for modeling a random variable taking values 0 and 1, it is more reasonable to suppose a Bernoulli distribution rather than continuous normal distributions (Moustaki & Knott, 2000; Moustaki, 1996). Similarly, a Poisson distribution may be a better choice for modeling a random variable consisting of counts (Moustaki & Knott, 2000).

The proposed algorithm has been developed based on the assumption that all relevant random vectors are multivariate normally distributed. However, this assumption can be easily relaxed so as to encompass more general response data generating mechanisms. For example, the logistic distribution can be used for modeling dichotomous manifest variables. In this sense, it would be fair to say that the proposed algorithm laid out the foundations for including more comprehensive measurement models in maximum likelihood estimation for PFA(p, q) models. Equation (5.1) clearly shows that various types of distributions for measurement models can be handled in the first term separately without affecting the distribution and the derived covariance structure of the latent process variables. This flexibility constitutes a major advantage of the proposed *EM* algorithm.

It may be difficult to analytically compute the conditional expectation in equations (5.2), (5.9), (5.10), and (5.11), especially when the predictive density of the latent variable, $p(\eta|Y, \xi^{(\ell)})$ is of mathematically intractable form. A numerical integration method such as Gauss-Hermite Adaptive Quadrature (Liu & Pierce, 1994) may be adopted here, but the demand for computational resources, due to the high dimensionality of numerical integrations in PFA models even under the moderate size of T, can easily exceed the computational capacity of today's high-speed desktop computers. A possible solution to such computational difficulties, known as the curse of dimensionality, can be found in Monte Carlo methods (Wei & Tanner, 1990), where independent samples drawn from $p(\eta|Y, \xi^{(\ell)})$ approximate the expectations or integrals. When direct draws of independent samples from $p(\eta|Y, \xi^{(\ell)})$ are not plausible, Importance Sampling (Rubin, 1988) or the Markov Chain Monte Carlo method (Gilks & Spiegelhalter, 1996; Robert & Casella, 2004) can be explored as viable choices.

8.6.3 Identification of order of latent time series

Applications of time series data analysis require a decision of the appropriate order of the underlying (vector) autoregressive moving average process. In most areas of application of time series, methods used to select a model have been developed for a time series for manifest variables. Exploratory procedures based on autocorrelation functions and partial autocorrelation functions are not directly applicable when the time series of interest is latent (Browne & Nesselroade, 2005). A potential advantage of the proposed *EM* algorithm is the availability of the estimated latent scores by equation (5.9). Thus, the order of the latent time series may be determined based on these estimated factor scores by employing various model selection techniques such as AIC, BIC, and MDL, etc. Another important advantage of the SEM methodology is that it provides model fit indices such as RMSEA (Steiger & Lind, 1980; Browne & Cudeck, 1993). Such indices will provide applied researchers with useful tools of model evaluation and the selection of proper order of the latent process variables. Further research is warranted on this issue.

8.7 Conclusion

The proposed EM algorithm for obtaining maximum likelihood estimates for PFA(p, q) models can be effectively used in applied psychological research by providing statistically optimal solutions, i.e. MLEs. Most of all, the proposed algorithm provides a method for estimating temporal relationships among (stable) latent process factors with their measurement errors being accounted for. Once the aforementioned issues are well addressed, the proposed algorithm will have the potential of full functioning as an established statistical tool to aid applied researchers who need to use latent variable time series models.

Appendix A: Proof of equations (5.9) and (5.10)

During the derivation, the following well-known facts about the multivariate normal distribution (Mardia et al, 1976) will be used: Suppose

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right] \tag{1}$$

Here $\mathcal{N}$ denotes the multivariate normal distribution and X_1 and X_2 are vectors of arbitrary dimension. Then, the conditional distribution of X_1 given $X_2 = x_2$ is

$$\mathcal{N}\left[\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right] \tag{2}$$

Conversely, if $X_2 \sim \mathcal{N}(\mu_2, \Sigma_{22})$ and (2) holds, then so does (1). Note that

$$Y|\eta \sim \mathcal{N}(\Lambda\eta, \Phi) \tag{3}$$

Identifying $X_1 = Y$ and $X_2 = \eta$, and applying (1) and (2), it can be seen that

$$\mu_2 = 0 \tag{4}$$

$$\Sigma_{22} = \Sigma(\tau) \tag{5}$$

$$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2) = \Lambda\eta \tag{6}$$

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Phi \tag{7}$$

Rearranging the terms yields,

$$\Sigma_{12}\Sigma_{22}^{-1} = \Lambda \tag{8}$$

$$\Sigma_{12} = \Lambda\Sigma(\tau) \tag{9}$$

$$\mu_1 = \Sigma_{12}\Sigma_{22}^{-1}\mu_2 = 0 \tag{10}$$

Finally,

$$\Sigma_{11} = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + V_t = \Sigma_{12}\Sigma_{22}^{-1}(\Sigma_{22})\Sigma_{22}^{-1}\Sigma_{21} + V_t = \Lambda\Sigma(\tau)\Lambda' + \Phi \tag{11}$$

Therefore,

$$\begin{pmatrix} Y \\ \\ \eta \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, \begin{pmatrix} \Lambda\Sigma\Lambda' + \Phi & \Lambda\Sigma(\tau) \\ \\ \Sigma(\tau)\Lambda' & \Sigma(\tau) \end{pmatrix}\right] \tag{12}$$

Now applying (1) an (2) again, reversing the role of X_1 and X_2, the conditional distribution of η given Y is

$$\eta|Y \sim \mathcal{N}\left(\mathrm{E}\left[\eta|Y\right], \mathrm{cov}\left[\eta, \eta'|Y\right]\right) \tag{13}$$

where

$$\begin{aligned}
\mathrm{E}\left[\eta|Y\right] &= \Sigma(\tau)\Lambda'\left(\Lambda\Sigma\Lambda' + \Phi\right)^{-1} Y \\
&= \left(\Lambda'\Phi^{-1}\Lambda + \Sigma(\tau)^{-1}\right)^{-1} \Lambda'\Phi^{-1}Y && (14) \\
\mathrm{cov}\left[\eta, \eta'|Y\right] &= \Sigma(\tau) - \Sigma(\tau)\Lambda'\left(\Lambda\Sigma\Lambda' + \Phi\right)^{-1}\Lambda\Sigma(\tau) \\
&= \left(\Lambda'\Phi^{-1}\Lambda + \Sigma(\tau)^{-1}\right)^{-1} && (15)
\end{aligned}$$

The last step uses the matrix identity, known as *Sherman-Morrison-Woodbury Matrix Identity*, (Searle, 1982),

$$(A + URV')^{-1} = A^{-1} - A^{-1}U\left(R^{-1} + V'A^{-1}U\right)^{-1} V'A^{-1} \tag{16}$$

where A, U, R and V are matrices of appropriate order with A and R being nonsingular.

Appendix B: The closed form expressions for typical elements of missing information matrix

$$
\begin{aligned}
\frac{\partial^2 H(\xi|\hat{\xi})}{\partial\lambda_r\partial\lambda_s} = &-\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left[\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_s}\right]\Omega(\xi)\left[\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right]\right\} \\
&-\mathrm{tr}\Bigg\{\Omega^{-1}(\xi)\left[-\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_s}\right)\Omega(\xi)\Lambda'\Phi^{-1}Y+\Omega(\xi)\left(\frac{\partial\Lambda}{\partial\lambda_s}\right)\Phi^{-1}Y\right] \\
&\qquad\left[-\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\Omega(\xi)\Lambda'\Phi^{-1}Y+\Omega(\xi)\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)\Phi^{-1}Y\right]'\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
= &-\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_s}\right)\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\right\} \\
&-\mathrm{tr}\left\{\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_s}\right)\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\Omega(\xi)\right\} \\
&+\mathrm{tr}\left\{\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_s}\right)\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)\Omega(\xi)\right\} \\
&+\mathrm{tr}\left\{\left(\frac{\partial\Lambda}{\partial\lambda_s}\right)\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\Omega(\xi)\right\} \\
&-\mathrm{tr}\left\{\left(\frac{\partial\Lambda}{\partial\lambda_s}\right)\Phi^{-1}YY'\Phi^{-1}\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)\Omega(\xi)\right\}
\end{aligned}
$$

$$
\begin{aligned}
= &-\frac{1}{2}\mathrm{vec}'\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_s}\right)\left(\Omega(\xi)\otimes\Omega(\xi)\right)\mathrm{vec}\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right) \\
&-\mathrm{vec}'\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_s}\right)\big(\Omega(\xi)\otimes\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\big)\mathrm{vec}\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right) \\
&+\mathrm{vec}'\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_s}\right)\big(\Omega(\xi)\otimes\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\big)\mathrm{vec}\left(\frac{\partial\Lambda}{\partial\lambda_r}\right) \\
&+\mathrm{vec}'\left(\frac{\partial\Lambda}{\partial\lambda_s}\right)\big(\Omega(\xi)\otimes\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\big)\mathrm{vec}\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right) \\
&+\mathrm{vec}'\left(\frac{\partial\Lambda}{\partial\lambda_s}\right)\big(\Omega(\xi)\otimes\Phi^{-1}YY'\Phi^{-1}\big)\mathrm{vec}\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)
\end{aligned} \tag{17}
$$

$$
\begin{aligned}
\frac{\partial^2 H(\xi|\hat{\xi})}{\partial\lambda_r\partial\phi_s} &= -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left[-\Lambda'\Phi^{-1}\left(\frac{\partial\Phi^{-1}}{\partial\phi_s}\right)\Phi^{-1}\Lambda\right]\Omega(\xi)\left[\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right]\right\} \\
&\quad - \mathrm{tr}\left\{\Omega^{-1}(\xi)\left[\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}Y - \Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\lambda_s}\right)\Phi^{-1}Y\right]\right. \\
&\qquad \left.\left[-\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\Omega(\xi)\Lambda'\Phi^{-1}Y + \Omega(\xi)\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)\Phi^{-1}Y\right]'\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\right\} \\
&\quad + \mathrm{tr}\left\{\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\Omega(\xi)\right\} \\
&\quad - \mathrm{tr}\left\{\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)\Omega(\xi)\right\} \\
&\quad - \mathrm{tr}\left\{\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\Omega(\xi)\right\} \\
&\quad + \mathrm{tr}\left\{\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}YY'\Phi^{-1}\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)\Omega(\xi)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2}\mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\otimes\Phi^{-1}\Lambda\Omega(\xi)\right)\mathrm{vec}\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right) \\
&\quad + \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\otimes\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\right)\mathrm{vec}\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right) \\
&\quad - \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\otimes\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Lambda(\xi)}{\partial\lambda_r}\right) \\
&\quad - \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\otimes\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\right)\mathrm{vec}\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right) \\
&\quad + \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\otimes\Phi^{-1}YY'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Lambda(\xi)}{\partial\lambda_r}\right) \qquad (18)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 H(\xi|\hat{\xi})}{\partial\lambda_r\partial\gamma_s} = -\frac{1}{2}\mathrm{tr}\Bigg\{&\Omega(\xi)\left[-\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\right]\Omega(\xi)\left[\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right]\Bigg\}\\
-\mathrm{tr}\Bigg\{&\Omega^{-1}(\xi)\left[\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}Y\right]\\
&\left[-\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\Omega(\xi)\Lambda'\Phi^{-1}Y+\Omega(\xi)\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)\Phi^{-1}Y\right]'\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
= \quad &\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\right\}\\
&+\mathrm{tr}\left\{\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\Omega(\xi)\right\}\\
&-\mathrm{tr}\left\{\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\left(\frac{\partial\Lambda}{\partial\lambda_r}\right)\Omega(\xi)\right\}
\end{aligned}
$$

$$
\begin{aligned}
= \quad &\frac{1}{2}\mathrm{vec}'\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\left(\Sigma_\tau^{-1}\Omega(\xi)\otimes\Sigma_\tau^{-1}\Omega(\xi)\right)\mathrm{vec}\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\\
&+\mathrm{vec}'\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\left(\Sigma_\tau^{-1}\Omega(\xi)\otimes\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\right)\mathrm{vec}\left(\frac{\partial\Omega^{-1}(\xi)}{\partial\lambda_r}\right)\\
&-\mathrm{vec}'\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\left(\Sigma_\tau^{-1}\Omega(\xi)\otimes\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Lambda}{\partial\lambda_r}\right) \qquad (19)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 H(\xi|\hat{\xi})}{\partial\phi_r\partial\phi_s} &= -\frac{1}{2}\mathrm{tr}\Bigg\{\Omega(\xi)\left[-\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\right]\Omega(\xi)\left[-\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\right]\Bigg\} \\
&\quad - \mathrm{tr}\Bigg\{\Omega^{-1}(\xi)\left[\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}Y - \Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}Y\right] \\
&\qquad\qquad \left[\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}Y - \Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}Y\right]'\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\right\} \\
&\quad - \mathrm{tr}\left\{\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi^{-1}}{\partial\phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\right\} \\
&\quad + \mathrm{tr}\left\{\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\right\} \\
&\quad + \mathrm{tr}\left\{\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\right\} \\
&\quad - \mathrm{tr}\left\{\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_s}\right)\Phi^{-1}YY'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{2}\mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\otimes\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\phi_r}\right) \\
&\quad - \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\big(\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1} \\
&\qquad\qquad \otimes\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\big)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\phi_r}\right) \\
&\quad + \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\otimes\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\phi_r}\right) \\
&\quad + \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\otimes\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\phi_r}\right) \\
&\quad - \mathrm{vec}'\left(\frac{\partial\Phi}{\partial\phi_s}\right)\left(\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\otimes\Phi^{-1}YY'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\phi_r}\right) \qquad (20)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 H(\xi|\hat{\xi})}{\partial\phi_r\partial\gamma_s} &= -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left[-\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\right]\Omega(\xi)\left[-\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\right]\right\} \\
&- \mathrm{tr}\Bigg\{\Omega^{-1}(\xi)\left[\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}Y\right] \\
&\quad \left[\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi^{-1}}{\partial\phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}Y - \Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}Y\right]'\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\right\} \\
&\quad - \mathrm{tr}\left\{\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\right\} \\
&\quad + \mathrm{tr}\left\{\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\left(\frac{\partial\Phi}{\partial\phi_r}\right)\Phi^{-1}\Lambda\Omega(\xi)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{2}\mathrm{vec}'\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\left(\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}\otimes\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\phi_r}\right) \\
&\quad - \mathrm{vec}'\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\left(\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}\otimes\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\Lambda'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\phi_r}\right) \\
&\quad + \mathrm{vec}'\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\left(\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}\otimes\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\right)\mathrm{vec}\left(\frac{\partial\Phi}{\partial\phi_r}\right) \qquad (21)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 H(\xi|\hat{\xi})}{\partial\gamma_r\partial\gamma_s} = & -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\left[-\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\right]\Omega(\xi)\left[-\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_r}\right)\Sigma_\tau^{-1}\right]\right\} \\
& - \mathrm{tr}\left\{\Omega^{-1}(\xi)\left[\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}Y\right]\right. \\
& \left.\left[\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_r}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}Y\right]'\right\} \\
= & -\frac{1}{2}\mathrm{tr}\left\{\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_r}\right)\Sigma_\tau^{-1}\right\} \\
& - \mathrm{tr}\left\{\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\Sigma_\tau^{-1}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_r}\right)\Sigma_\tau^{-1}\Omega(\xi)\right\} \\
= & -\frac{1}{2}\mathrm{vec}'\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\left(\Sigma_\tau^{-1}\Omega(\xi)\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\Omega(\xi)\Sigma_\tau^{-1}\right)\mathrm{vec}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_r}\right) \\
& - \mathrm{vec}'\left(\frac{\partial\Sigma_\tau}{\partial\gamma_s}\right)\left(\Sigma_\tau^{-1}\Omega(\xi)\Sigma_\tau^{-1}\otimes\Sigma_\tau^{-1}\Omega(\xi)\Lambda'\Phi^{-1}YY'\Phi^{-1}\Lambda\Omega(\xi)\Sigma_\tau^{-1}\right)\mathrm{vec}\left(\frac{\partial\Sigma_\tau}{\partial\gamma_r}\right)
\end{aligned}
\tag{22}
$$

Bibliography

Anderson, T. W. (1959). Some stochastic process models for intelligence test scores. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences.* Stanford, CA: Stanford University Press.

Bickel, P. J., & Doksum, K. A. (2001). *Mathematical statistics: basic ideas and selected topics*. Upper Saddle River, N. J.: Prentice Hall.

Biemer, P. P., Christ, S. L., & Wiesen, C. A. (2009). A general approach for estimating scale score reliability for panel survey data. *Psychological methods*, *14*(4), 400–412.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the *em* algorithm. *Psychometrika*, *46*(4), 443–449.

Bock, R. D., & Bargmann, R. (1966). Analysis of covariance structure. *Psychometrika*, *31*(4), 507–534.

Bollen, K. A. (1996). An alternative two stage least squares (2sls) estimator for latent variable equations. *Psychometrika*, *61*(1), 109–121.

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting*. New York: Springer.

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

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & S. Long (Eds.), *Testing structural equation models* (pp. 131–161). Newbury Park, CA: Sage.

Browne, M. W., & du Toit, S. H. C. (1991). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: recent advances, unanswered questions, future directions* (pp. 47–68). Washington, DC: APA.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: some promising uses and extensions of autoregressive moving average time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: a festschrift for roderick p. mcdonald* (pp. 415–452). Mahwah, NJ: Lawrence Erlabum Associates.

Browne, M. W., & Zhang, G. (2007a). Developments in the factor analysis of individual time series. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: historical developements and future directions* (pp. 265–291). Mahwah, NJ: Lawrence Erlbaum Associates.

Browne, M. W., & Zhang, G. (2007b). Repeated time series models for learning data. In M. J. Boker S. M. an Wenger (Ed.), *Data analytic techniques for dynamica systems* (pp. 25–45). Mahwah, NJ: Lawrence Erlbaum Associates.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society. Series B*, *39*, 1–38.

du Toit, S. H. C., & Browne, M. W. (2007). Structural equation modeling of multivariate time series. *Multivariate Behavioral Research*, *42*(1), 67–101.

du Toit, S. H. C., & du Toit, M. (2008). Multilevel structural equation modeling. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 435–478). New York: Springer.

Engle, R., & Watson, M. (1981). A one-factor multivariate time series model of metroppolitan wage rates. *Journal of the American Statistical Association*, *76*, 774–781.

Geweke, J., & Singleton, K. (1981). Maximum likelihood confirmatory factor analysis of economic time series. *International economic review*, *22*(1), 37–54.

Gilks, W. R., & Spiegelhalter, D. J. (1996). *Markov chain monte carlo in practice*. CRC: Champman & Hall.

Granger, C. W. J. (1969). Investigating causal relationship by econometric models and cross spectral methods. *Econometrica*, *37*, 424–438.

Guttman, L. (1954). A new approach to factor analysis: The Radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348). New York: Columbia University Press.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2002). On the nature of sem estimates of ARMA parameters. *Structural Equation Modeling*, *9*(3), 347–368.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2003). ARMA-based SEM when the number of time points t exceeds the number of cases: raw data maximum likelihood. *Structural Equation Modeling*, *10*(3), 352–379.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: rationale and application of multivariate stationary time series analysis. *Structural Equation Modeling*, *40*(2), 207–233.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American sociological review*, *34*, 93–101.

Jöreskog, K. G. (1966). Testing a simple structure hypothesis in factor analysis. *Psychometrika*, *31*(2), 165–178.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*(4), 443–482.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183–202.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, *23*, 121–145.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426.

Jöreskog, K. G., & Sörbom, D. (1977). Statistical models and methods for analysis of longitudinal data. In D. Aigner & A. Goldberger (Eds.), *Latent variables in socioeconomic models* (pp. 285–325). New York: Wiley.

Jöreskog, K. G., & Sörbom, D. (2006). Lisrel 8.8 for windows [Computer software manual]. Lincolnwood, IL.

Lee, S.-Y., & Poon, W.-Y. (1998). Analysis of two-level structural equation models via em type algorithms. *Statistica Sinica*, *8*, 749–766.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation.* New York: Springer.

Liang, J., & Bentler, P. M. (2004). An em algorithm for fitting two-level structural equation models. *Psychometrika*, *69*(1), 101–122.

Liu, Q., & Pierce, D. A. (1994). A note on gauss-hermite quadrature. *Biometrika*, *81*(3), 624–629.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, *44*(2), 226–233.

Meng, X.-L., & Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the Americal Statistical Association*, *86*(416), 899–909.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202.

Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology*, *49*, 313–334.

Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, *65*(3), 391–411.

Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. (2002). Dynamic factor analysis models for multivariate tim series analysis. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling individual variability with repeated measures data: advances and techniques* (pp. 235–265). New Jersey: Lawrence Erlbaum Associates.

Oakes, D. (1999). Direct calculation of the information matrix via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *61*, 479–482.

Orchard, T., & Woodbury, M. A. (1972). A missing information principle: Theory and applications. *Proceedings of the 6th Berkeley Symposium on Mathematial Statitics*, *1*, 697–715.

R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org` (ISBN 3-900051-07-0)

Robert, C. P., & Casella, G. (2004). *Monte carlo statitical methods.* New York: Springer-Verlag.

Rubin, D. B. (1988). Using the SIR algorithms to simulate posterior distributions. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics* (pp. 395–402). New York: Oxford University Press.

Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, *47*(1), 69–76.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Champman & Hall.

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society. Iowa City, Iowa.

van Buuren, S. (1997). Fitting ARMA time series by structural equation models. *Psychometrika*, *62*(2), 215–236.

van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge: Cambridge University Press.

Wei, G. C. G., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's Data Augmentation algorithms. *Journal of the American Statistical Association*, *85*, 699–704.

Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American sociological review*, *35*, 112–117.

Zhang, G. (2006). *Bootstrap procedures for dynamic factor analysis.* Unpublished doctoral dissertation, The Ohio State University, Ohio.