Finite Sample Performance of Standard Error Estimators for Dynamic Factor Analysis of Non-Normal Data Using the Kalman Filter Algorithm

Zijun Ke

Publication Date

19-04-2012

FINITE SAMPLE PERFORMANCE OF STANDARD ERROR ESTIMATORS FOR DYNAMIC FACTOR ANALYSIS OF NON-NORMAL DATA USING THE KALMAN FILTER ALGORITHM

A Thesis

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Master of Arts

by

Zijun Ke,

Guangjian Zhang, Director

Graduate Program in Psychology
Notre Dame, Indiana

April 2012

FINITE SAMPLE PERFORMANCE OF STANDARD ERROR ESTIMATORS FOR DYNAMIC FACTOR ANALYSIS OF NON-NORMAL DATA USING THE KALMAN FILTER ALGORITHM

Abstract

by

Zijun Ke

This master thesis is concerned with the finite sample properties of four standard error (SE) estimators for dynamic factor analysis using the Kalman filter algorithm with both normal and nonnormal data. The estimators considered are the observed information based SE estimator, Harvey's SE estimator, and the two sandwich type SE estimators. Statistical properties of these estimators are assessed using a simulation study. Results indicate that the sandwich type SE estimator proposed by Papanastassiou (2006) generally outperforms other SE estimators. However, the observed information SE estimator is still valuable in that the advantage of the sandwich type SE estimator proposed by Papanastassiou (2006) over the observed information SE estimator for non-covariance component parameters is limited.

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Guangjian Zhang for his guidance and support on the development of this Master thesis. His stimulating suggestions and encouragement have helped to make this research fun for me. Besides my advisors, I would like to thank the rest of the members of my committee, Dr. Scott Maxwell and Dr. Ke-Hai Yuan, and Dr. Zhiyong Zhang for their patience and support. I would also like to thank my colleagues Xiaoling Zhong for their sound advice and encouragement which helped me get through the difficult times.

CHAPTER 1

INTRODUCTION

1.1 Dynamic Studies in Psychology

Social and behavioral scientists are increasingly recognizing the importance of studying change. The P-technique factor analysis proposed by Cattell, Cattell, and Rhymer (1947) was an early attempt to study within individual variability. With decades of development, researchers today are afforded with better statistical techniques to model individual change over time and in many substantive areas in psychology, process-oriented research is getting increasing attention. Studies on the dynamics underlying mood structures in Parkinson's disease (Chow, Nesselroade, Shiren, & McArdle, 2004) and on the relationship between children's perceived control and school performance in terms of intraindividual changes (Musher, Nesselroade, & Schmitz, 2002) are typical examples of applying dynamic models in social and educational psychology. An illustration of applying dynamic techniques to study interindividual differences with respect to intraindividual variation in the Five Factor model is given by Hamaker, Dolan, and Molenaar (2005). Applications in the developmental area are discussed by Nesselroade and Paolo (2003). Dynamic models also provide ways to analyze various bio-signals including responses in fMRI, EEG, skin conductance, heart rate, etc., (Yang & Chow, 2010; Parkinson, Laura, & Jackson, 2010; Tarvainen, Georgiadis, Ranta-aho, & Karjalainen, 2006).

1.2 Dynamic Factor Analysis Models and Research Questions

The dynamic factor analysis (DFA) model is one promising way to study intraindividual variations (Browne & Nesselroade, 2005; Browne & Zhang, 2007; Molenaar, 1985; Nesselroade, McArdle, Aggen, & Meyers, 2002). By integrating the factor model and the dynamic model, the DFA model is able to take into account measurement errors as well as to describe the underlying dynamical mechanisms. There are two ways to fit dynamic factor models: either through the SEM approach or through the state-space approach. Despite its long history in economics and engineering, the state-space approach is still unknown to many social and behavioral scientists. This thesis will focus on the state-space approach.

When data are normally distributed, the maximum likelihood (ML) estimates of the state space model can be obtained using the Kalman filter algorithm (Harvey, 1989). Several standard error (SE) estimators for ML estimates based on the information matrix have been proposed (Harvey, 1989, Shumway & Stoffer, 2004). Using simulation studies, Papanastassiou (2006) showed that the average bias of the information based SE estimators was less than 10% of the values of the true SE when data were normally distributed.

When data are nonnormal, parameter estimates obtained by maximizing the normal likelihood function are referred to as normal theory based maximum likelihood estimates (Harvey & Shephard, 1996)[1]. Three sandwich-type standard error estimators of normal theory based ML estimates have been proposed for non-Gaussian data (Papanastassiou, 2006; White, 1984). Finite sample properties of these standard error estimators are largely unknown, however. This thesis will study the performance of standard error estimators in finite samples.

[1] They were called quasi-maximum likelihood estimates in Harvey and Shephard (1996) and many other papers in economic area (e.g., Strickland, Forbes, & Martin, 2006; Bauwens & Veredas, 2004), however. In social and behavioral sciences, these estimates are often called to normal theory based maximum likelihood estimates (e.g., Bentler & Yuan, 1999 and Yuan & Bentler, 1998), We decide to use the term “normal theory based maximum likelihood”.

The motivation for the current thesis resides in three aspects. First, studies on SE estimators are important. Some statistical inferences, such as hypothesis testing and interval estimation, involve standard error estimates. Inaccurate standard error estimates may lead to misleading statistical conclusions. For example, a significant autoregressive coefficient may turn out to be nonsignificant due to inaccurate standard error estimates (see example 3 in Chow, Ho, Hamaker, & Dolan, 2010). Therefore it is worthwhile to study the performance of standard error estimators. Second, psychological data rarely follow the Gaussian distribution (Micceri, 1989). Micceri (1989) examined 440 large empirical data sets. They were various ability tests, aptitude tests, personality inventories, measures of anger, anxiety, sociability and so on. The normality assumption failed in each of them. Consequently, the Gaussian assumption is often untenable in social and behavior research. Thus whether standard error estimators are robust to violation of the normality assumption is of interest to both methodologists and applied researchers. Third, these asymptotic standard errors are often justified by asymptotic theory. Their finite sample properties are largely unknown. In real data analysis, especially in psychological research, the sample size of time series data is generally not very large. Hence it is more informative to explore the finite sample performance of standard error estimators using simulation studies.

1.3 Goals of Current Thesis

The main purpose of the current thesis is to systematically study the finite sample performance of SE estimators using the state-space approach when fitting DFA models to non-normal data. These SE estimators include two information based SE estimators and two sandwich type SE estimators. Two practical questions to be answered are "Is it appropriate to use these SE estimators when data are not normally distributed" and "Which one is better".

The thesis is arranged as follows. Chapter 2 introduces dynamic factor analysis and compares different methods of fitting DFA models using the SEM approach. Chapter 3 introduces the state-space approach and compares different methods of fitting DFA models using the state-space approach. Chapter 4 discusses asymptotic normality and introduces standard error estimators for the state space model. Chapter 5 describes the simulation design. Chapter 6 reports simulation results and chapter 7 summarizes and discusses the findings and limitations of this thesis.

CHAPTER 2

DYNAMIC FACTOR ANALYSIS

In this chapter, dynamic factor analysis (DFA) models will be introduced. Section 2.1 gives a formal definition of the DFA model. Section 2.2 compares various methods of fitting DFA models using the SEM approach.

2.1 Introduction to DFA Models

Dynamic factor analysis is used to study intraindividual changes. There are two major types of DFA models: the shock factor analysis model (Browne & Nesselroade, 2005; Molenaar, 1985) and the process factor analysis (PFA) model (Browne & Nesselroade, 2005; Browne & Zhang, 2007). The shock factor analysis model emphasizes the latent random shocks that drive the manifest processes. These latent shocks are exogenous and are uncorrelated across time. In contrast, the process factor model emphasizes latent dynamics that represent the observed psychological processes. The latent factors in the PFA model are endogenous and they are correlated over time. In this thesis, I focus on PFA models.

The factor analysis model for the manifest variables in a typical PFA model is

$$\mathbf{y}_t = \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{u}_t, \qquad t = 1, \cdots, T, \tag{2.1}$$

where the $k \times 1$ vector $\mathbf{y}_t$ is the k observed manifest variables at time t; the $k \times m$ matrix Λ is the factor loading matrix; $\mathbf{f}_t$ is the $m \times 1$ latent factors at time t; the $k \times 1$

white noise variables $\mathbf{u}_t$ represent the measurement errors at time t. In PFA models, the factors, $\mathbf{f}_t$, follow a stationary vector ARMA model, that is,

$$\begin{aligned}\mathbf{f}_t &= \mathbf{z}_t + \mathbf{A}_1\mathbf{f}_{t-1} + \mathbf{A}_2\mathbf{f}_{t-2} + \cdots + \mathbf{A}_p\mathbf{f}_{t-p} \\ &\quad + \mathbf{B}_1\mathbf{z}_{t-1} + \mathbf{B}_2\mathbf{z}_{t-2} + \cdots + \mathbf{B}_q\mathbf{z}_{t-q},\end{aligned} \tag{2.2}$$

where the $m \times 1$ random shock variables $\mathbf{z}_t$ are the process noises at time t; and the $m \times m$ matrices $\mathbf{A}_p$ and $\mathbf{B}_q$ are the corresponding autoregressive weights and moving average weights. As shown in (2.2), in addition to incorporating the influences of the q preceding shock variables on current factors, the DFA models allow the current factors, $\mathbf{f}_t$, to depend on the p preceding factors.

The measurement errors, $\mathbf{u}_t$, and the random shocks, $\mathbf{z}_t$, have similar mathematical properties, namely,

$$\begin{aligned} E(\mathbf{u}_t) &= E(\mathbf{z}_t) = 0, \\ Cov(\mathbf{u}_t, \mathbf{u}_t') &= \Xi \text{ and } Cov(\mathbf{z}_t, \mathbf{z}_t') = \Psi \end{aligned} \tag{2.3}$$

for $t = 1, \ldots, T$. And both of them are not correlated across time. But they have different theoretical meanings. $\mathbf{u}_t$ represents measurement errors that contaminate the observations of latent factors $\mathbf{f}_t$ while $\mathbf{z}_t$ are the random shock variables that drive the latent processes. More specifically, $\mathbf{z}_t$ influences the current underlying factor and subsequent factors while $\mathbf{u}_t$ only influences the observation of the current factor. The measurement errors are assumed to be independent of each other. Thus Ξ is a diagonal matrix.

2.2 Fitting DFA Models: SEM Approach

In this section, three methods of fitting DFA models using the structural equation modeling framework are reviewed. The first two methods obtain maximum likelihood estimates and the third method obtains ordinary least square estimates.

2.2.1 Raw Data Likelihood

When analyzing time series data, the model implied covariance matrix, $\mathbf{\Sigma}(\theta)$ can be established according to the dynamic and factor structures (Du Toit & Browne, 2007). The sample covariance matrix, $\mathbf{S}$, consists of sample lag covariance matrices. The covariance matrix has a size of $N \times Tk$ where N is the number of participants; T is the time series length and k is the number of observed variables. If there is a sufficient number of subjects, say, more than 100 (and usually with a smaller T, i.e., $T \leq 10$), the sample covariance matrix, $\mathbf{S}$, can be computed in the independent data fashion, that is, averaging the squared deviations from the mean across subjects. However, if the number of subjects is not large enough, the resulting sample covariance matrix may be singular. In this situation, the raw data log-likelihood function $\mathcal{L}$ is involved to avoid using the sample covariance matrix (Hamaker, Dolan, & Molenaar, 2003). Since this procedure involves inverting a $Tk \times Tk$ matrix, it is computationally inefficient, especially when T is large.

2.2.2 Block Toeplitz-Pseudo MLE

Another way to address this problem is to use the block Toeplitz (BT) method (Molenaar, 1985; Nesselroade et al., 2002; Hamaker, Dolan, & Molenaar, 2002) to reduce the dimension of the input matrix. By treating the sample lag covariance matrix up to lag L, $\mathbf{S}(L)$ as the sample covariance matrix, parameter estimates can be obtained by minimizing the likelihood function in the standard SEM framework. At least two issues concerning this method arise. Due to the dependence in the data, the sample lag covariance matrix, $\mathbf{S}(L)$, usually does not have the Wishart distribution which is, however, assumed for maximum likelihood estimation. Consequently, this approach only produces "pseudo-maximum likelihood estimates" (Molenaar & Nesselroade, 1998). Note that it is the case even for well behaved data, i.e., Gaussian processes. This will pose a serious problem in that standard error estimates may be inaccurate and further statistical inferences such as hypothesis testing and model fit comparisons based on MLE will lose their theoretical foundation (Hamaker et al., 2002; Chow et al., 2010;Z. Zhang, Hamaker, & Nesselroade, 2008). Using a simulation study, Z. Zhang et al. (2008) showed that the BT method yielded relatively larger total absolute errors (absolute biases) compared to other methods, e.g., the state space approach. And Chow et al. (2010) showed in a simulation study that statistical inferences about some parameters can be misleading because of inaccurate SE estimates obtained from the BT method. The second issue concerning the block Toeplitz method resides in its implementation. Since there are many redundant elements in the block Toeplitz covariance matrix, one needs to place some constraints to exhaust false degrees of freedom when using the standard SEM software to fit the model. Those constraints could be complex if the model is large and complicated.

2.2.3 Block Toeplitz-OLS

A third way to analyze time series data using the SEM approach is to use ordinary least square (OLS) estimation (Browne & Zhang, 2007). In this method, the block Toeplitz matrix is still used to reduce the dimensions of the input matrix. Instead of minimizing the likelihood function, it minimizes the discrepancy function which measures the squared distance between the sample lag correlation matrix and the model implied correlation matrix. Recently, several standard error estimators for the OLS method have been developed. The sandwich-type one is analytic (G. Zhang, Chow, & Ong, 2011). Others are bootstrap based standard error estimators (G. Zhang

& Browne, 2010; G. Zhang & Chow, 2010). The OLS method has several advantages. For example, since OLS only adopts the non-duplicated elements in the block Toeplitz matrix, no complicated constraints are needed to adjust the false degree of freedom. The implementation is more straightforward. The OLS method also provides a basis for conducting the dynamic factor analysis in an exploratory way. In addition, since no distributional assumption is required, OLS can be applied to data that are not normally distributed. However, OLS is not the perfect method. Due to the use of inconsistent estimators of the limiting covariance matrix of the sample covariance matrix, OLS estimates are not the most efficient. By simulation studies, Z. Zhang et al. (2008) showed that OLS provides good, but not optimal estimates, in that it produces larger mean square standard error (standard deviation of the estimates across replications) than other estimates, e.g., the ML estimates obtained by using Kalman filter algorithm.

Another method for fitting dynamic factor analysis models is the state space approach. The state space approach provides maximum likelihood estimates. It can also provide subject specific factor scores. Methods for computing these scores within the SEM framework are unavailable. The state space model has been adapted for missing data problems and initial condition problems (Shumway & Stoffer, 2004; Harvey, 1989; De Jong, 1988; Zarchan & Musoff, 2000). The state space model and its implementation will be described in the next chapter.

CHAPTER 3

KALMAN FILTER ALGORITHM

The state-space approach is another important way of fitting DFA models (Harvey, 1989; Shumway & Stoffer, 2004). The key idea of this approach is to utilize the Gaussian Kalman filter algorithm to obtain the maximum likelihood estimates. This chapter focuses on the state space model and associated estimation issues. Section 3.1 introduces basic concepts of state-space modeling. Section 3.2 briefly reviews how to obtain the MLE for state space models using the Gaussian Kalman filter algorithm. Section 3.3 reviews different methods under the state space framework which work for non-Gaussian processes. This section also compares these methods with the Gaussian Kalman filter approach.

3.1 State Space Models

Among several key features of the state space models (SSM) in representing psychological processes are a) the capability to incorporate a whole class of special models of interest using a relatively simple form; and b) the feasibility of obtaining maximum likelihood estimates for parameters of dynamic factor analysis models.

Similar to the measurement equation of DFA models, the observation equation of SSM describes the relationship between manifest variables $\mathbf{y}_t$ and state variables $\mathbf{x}_t$ and can be expressed as

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \mathbf{v}_t, \qquad t = 1, 2, \cdots, T \tag{3.1}$$

where the $q \times 1$ vector $\mathbf{y}_t$ represents the q manifest variables at time t; the time invariant $q \times p$ matrix $\mathbf{A}$ is the observation matrix; $\mathbf{x}_t$ is the $p \times 1$ state variable at time t; the $q \times 1$ white noise variable $\mathbf{v}_t$ represents the measurement errors with covariance matrix

$$Cov\left(\mathbf{v}_t, \mathbf{v}_t^T\right) = \mathbf{R}.$$

The transition equation is specified as,

$$\mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{w}_t, \tag{3.2}$$

where the $p \times p$ matrix $\mathbf{\Phi}$ is the transition matrix and the $p \times 1$ random shock variable $\mathbf{w}_t$ is the process disturbance with covariance matrix

$$Cov\left(\mathbf{w}_t, \mathbf{w}_t^T\right) = \mathbf{Q}.$$

The state equation determines how the current state vector $\mathbf{x}_t$ updates from the past state vector $\mathbf{x}_{t-1}$. Generally, the two noise variables, the measurement errors and the process disturbance are assumed to be Gaussian white noise processes and thus carry the neat properties of Gaussian white noise sequences.

The following examples illustrate how to rewrite different models in the form of state space models. By carefully defining the manifest variables and state variables, the SSM can subsume any vector ARMA model or any DFA model. For example, the VARMA(p, q) can be rewritten as

$$\mathbf{y}_t = \left[\begin{array}{cccc} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right] \left[\begin{array}{c} \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1} \\ \epsilon_t \\ \vdots \\ \epsilon_{t-p+1} \end{array} \right] + [\mathbf{O}] \tag{3.3}$$

$$
\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+2} \\ \mathbf{y}_{t-p+1} \\ \epsilon_t \\ \epsilon_{t-1} \\ \epsilon_{t-2} \\ \vdots \\ \epsilon_{t-q+1} \end{bmatrix}
=
\begin{bmatrix}
\mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p & \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_{q-1} & \mathbf{B}_q \\
\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0}
\end{bmatrix}
\begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p+1} \\ \mathbf{y}_{t-p} \\ \epsilon_{t-1} \\ \epsilon_{t-2} \\ \epsilon_{t-3} \\ \vdots \\ \epsilon_{t-q} \end{bmatrix}
$$

$$
+ \begin{bmatrix}
\mathbf{I} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\
\mathbf{I} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}
\end{bmatrix}
\begin{bmatrix} \epsilon_t \\ \mathbf{0} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{3.4}
$$

The DFA model can be expressed using SSM in a similar way. The observation equation of a DFA model will be

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{f}_t \\ \vdots \\ \mathbf{f}_{t-p+1} \\ \mathbf{z}_t \\ \vdots \\ \mathbf{z}_{t-p+1} \end{bmatrix} + \mathbf{u}_t.$$

The state equation is same as (3.4) except that $\mathbf{y}_t$ and ϵ_t are substituted with $\mathbf{f}_t$ and $\mathbf{z}_t$ respectively.

The state space framework can incorporate not only many popular time series models but also commonly seen longitudinal models. Take a typical linear latent growth curve model for example. The model with N subjects, each of which is measured $T*$ times, can be parameterized as a state space model with time series length $T = 1$ and a null transition matrix, $\mathbf{\Phi} = \mathbf{0}$,

$$\begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T*} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & \vdots \\ 1 & T*-1 \end{bmatrix} \begin{bmatrix} intercept_i \\ slope_i \end{bmatrix} + \begin{bmatrix} v_{i,1} \\ v_{i,2} \\ \vdots \\ v_{i,T*} \end{bmatrix},$$

and

$$\begin{bmatrix} intercept_i \\ slope_i \end{bmatrix} = \begin{bmatrix} \mu_m \\ \mu_s \end{bmatrix} + \begin{bmatrix} w_{i,intercept} \\ w_{i,slope} \end{bmatrix}.$$

Actually, when $\Phi = \mathbf{0}$, the state-space model is equivalent to the structural equation model. By choosing appropriate initial prior, the two approaches can yield identical likelihoods and hence parameter estimates (Chow et al., 2010).

Note that in multi-subject research settings, participants can follow distinct underlying processes as long as the time series length is sufficiently long. This opens up a way to study group differences in intraindividual change. For example, some researchers believe that males and females may have different emotion regulation systems despite sharing the same measurement models. When fitting the model, one could assign a different transition matrix, $\mathbf{\Phi}$ to males and females but hold the observation matrix, $\mathbf{A}$, the same across the two groups. And with the help of appropriate fit indexes, the model with differences across groups can be compared to the model with the same parameters across groups. This will shed light on whether there is a group difference in intraindividual change.

Equation (3.1) and (3.2) provide the basic form of state space models. More general forms of state-space models includes small modifications of (3.1) and (3.2) so that the state space framework can handle more types of research questions. For example, by adding a subscript i, the state space model can be used to analyze multi-subject data, although in real data analysis, an inspection of the homogeneity of the underlying systems of different subjects is always needed before fitting the state space model to multi-subject data. Another extension is that it allows parameters to be time-variant, i.e., $\mathbf{\Phi}$ varying across different time points, becoming $\mathbf{\Phi}_t$. This would help researchers to model more complicated dynamical systems.

3.2 Kalman Filter Algorithm

Given the statistical merits of maximum likelihood (ML) estimation, this thesis centers on ML estimates. There are various algorithms that have been developed to obtain ML estimates, e.g., the Newton-Raphson algorithm. However, these algorithms involve computing values of the likelihood function given some guesses of the parameters. The value of the likelihood function for time dependent data is generally hard to obtain. The Kalman filter algorithm, however, provides one way to obtain the value of the likelihood function for the state space model given current guesses of parameter values.

The likelihood function for the state space model can be decomposed into products of conditional density functions of data:

$$L(\mathbf{y_1}, \dots, \mathbf{y_n}, \boldsymbol{\Theta}) = P(\mathbf{y}_1, \dots, \mathbf{y}_n | \boldsymbol{\Theta}) = \prod_{t=1}^{T} P_{\Theta}(\mathbf{y}_t | \mathbf{Y}_{t-1}) \tag{3.5}$$

where $\mathbf{Y}_{t-1}$ denotes all preceding observations up to time $t-1$. Usually, the Gaussian assumption is adopted to achieve theoretical simplicity and tractability. With the normal assumption, that is, assuming the two disturbance terms $\mathbf{w}_t$ and $\mathbf{v}_t$, follow independent multivariate normal distributions, the conditional density $P_{\boldsymbol{\Theta}}(\mathbf{y}_t | \mathbf{Y}_{t-1})$ is given by

$$\begin{aligned} P_{\boldsymbol{\Theta}}(\mathbf{y}_t | \mathbf{Y}_{t-1}) &= \left(\frac{1}{\sqrt{2\pi}}\right)^q cov(\mathbf{y}_t | \mathbf{Y}_{t-1})^{-1/2} \\ &\quad \times exp\left\{-\frac{1}{2}\left[\mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1})\right]' cov(\mathbf{y}_t | \mathbf{Y}_{t-1})^{-1}\left[\mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1})\right]\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^q |\boldsymbol{\Sigma}_t|^{-1/2} exp\left\{-\frac{1}{2}\mathbf{e}_t' \boldsymbol{\Sigma}_t^{-1} \mathbf{e}_t\right\} \end{aligned} \tag{3.6}$$

where $\mathbf{e}_t \equiv \mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1})$ and $\boldsymbol{\Sigma}_t \equiv cov(\mathbf{y}_t | \mathbf{Y}_{t-1}) = cov(\mathbf{e}_t)$.

The Kalman filter algorithm is introduced to compute all $\mathbf{e}_t$ and $\boldsymbol{\Sigma}_t$. However, note that

$$\begin{aligned} \mathbf{e}_t &\equiv \mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1}) \\ &= \mathbf{y}_t - \mathbf{A}E(\mathbf{x}_t | \mathbf{Y}_{t-1}) - E(\mathbf{v}_t | \mathbf{Y}_{t-1}) \\ &= \mathbf{y}_t - \mathbf{A}\mathbf{x}_t^{t-1} \end{aligned} \tag{3.7}$$

and

$$\begin{aligned}
\mathbf{\Sigma}_t &\equiv cov\left(\mathbf{y}_t|\mathbf{Y}_{t-1}\right) \\
&= \mathbf{A}cov\left(\mathbf{x}_t|\mathbf{Y}_{t-1}\right)\mathbf{A}' + cov\left(\mathbf{v}_t|\mathbf{Y}_{t-1}\right) + 2\mathbf{A}cov\left(\mathbf{x}_t,\mathbf{v}_t|\mathbf{Y}_{t-1}\right) \qquad (3.8) \\
&= \mathbf{A}\mathbf{P}_t^{t-1}\mathbf{A}' + \mathbf{R}.
\end{aligned}$$

where $\mathbf{x}_t^{t-1} = E\left(\mathbf{x}_t|\mathbf{Y}_{t-1}\right)$ and $\mathbf{P}_t^{t-1} = E\left\{\left(\mathbf{x}_t - \mathbf{x}_t^{t-1}\right)\left(\mathbf{x}_t - \mathbf{x}_t^{t-1}\right)'|\mathbf{Y}_{t-1}\right\}$. $\mathbf{A}$ and $\mathbf{R}$ are all parameters in the state space model. Thus, the Kalman filter algorithm actually obtains $\mathbf{e}_t$ and $\mathbf{\Sigma}_t$ by computing $\mathbf{x}_t^{t-1}$ and $\mathbf{P}_t^{t-1}$. Here $\mathbf{x}_t^{t-1}$ is the conditional expectation of the latent factor at time t, namely, $\mathbf{x}_t$. Generally, $\mathbf{x}_t^{t-1}$ is treated as an estimator or a guess of the values of the latent factor at time t based on the information containing in the observations up to time $t-1$. For Gaussian processes, this estimator is the best in the sense that the error covariance of this estimator is the smallest while for non-Gaussian processes, it is the best among all linear predictors (Anderson & Moore, 1979). $\mathbf{P}_t^{t-1}$ is the conditional covariance matrix associated with the estimator, $\mathbf{x}_t^{t-1}$(Anderson & Moore, 1979). For Gaussian data, $\mathbf{P}_t^{t-1}$ is actually independent of $\mathbf{Y}_{t-1}$ and hence $\mathbf{P}_t^{t-1}$ is also the unconditional error covariance of the estimator, $\mathbf{x}_t^{t-1}$ (Anderson & Moore, 1979).

To compute $\mathbf{x}_t^{t-1}$ and $\mathbf{P}_t^{t-1}$, the Kalman filter algorithm incorporates two steps: the predicting step and the updating step. This algorithm is initialized using some guesses of the initial conditions and runs through the two steps iteratively. In the predicting step, with current $\mathbf{x}_{t-1}^{t-1}$ and $\mathbf{P}_{t-1}^{t-1}$, $\mathbf{x}_t^{t-1}$ and $\mathbf{P}_t^{t-1}$ are computed using the following equations,

$$\begin{aligned}
\mathbf{x}_t^{t-1} &= \mathbf{\Phi}\mathbf{x}_{t-1}^{t-1}, \qquad (3.9) \\
\mathbf{P}_t^{t-1} &= \mathbf{\Phi}\mathbf{P}_{t-1}^{t-1}\mathbf{\Phi}' + \mathbf{Q},
\end{aligned}$$

where $\mathbf{x}_{t-1}^{t-1} = E\left(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1}\right)$ and $\mathbf{P}_{t-1}^{t-1} = E\left\{\left(\mathbf{x}_{t-1} - \mathbf{x}_{t-1}^{t-1}\right)\left(\mathbf{x}_t - \mathbf{x}_{t-1}^{t-1}\right)'|\mathbf{Y}_{t-1}\right\}$. Similar to $\mathbf{x}_t^{t-1}$, $\mathbf{x}_{t-1}^{t-1}$ is the conditional expectation of the latent factor at time $t-1$. The difference is that it also includes the information provided by observations at time $t-1$. And $\mathbf{P}_{t-1}^{t-1}$ is the associated conditional error covariance of $\mathbf{x}_{t-1}^{t-1}$.

In the updating step, the values of $\mathbf{x}_t^{t-1}$ and $\mathbf{P}_t^{t-1}$ computed in the predicting step are adopted and $\mathbf{x}_t^t$ and $\mathbf{P}_t^t$ are computed as below,

$$
\begin{aligned}
\mathbf{x}_t^t &= \mathbf{x}_t^{t-1} + \mathbf{K}_t \left(\mathbf{y}_t - \mathbf{A}\mathbf{x}_t^{t-1} \right), \\
\mathbf{P}_t^t &= \left[\mathbf{I} - \mathbf{K}_t \mathbf{A} \right] \mathbf{P}_t^{t-1},
\end{aligned}
\tag{3.10}
$$

where

$$\mathbf{K}_t = \mathbf{P}_t^{t-1} \mathbf{A}' \mathbf{\Sigma}_t^{-1}$$

The new estimates of $\mathbf{x}_t^t$ and $\mathbf{P}_t^t$ are then used in the predicting step of the next iteration.

To sum up, the Gaussian Kalman filter starts with the guesses of $\mathbf{x}_0^0$ and $\mathbf{P}_0^0$ and assigned parameter values. By (3.9) and (3.10) the filter updates to $\mathbf{x}_1^1$ and $\mathbf{P}_1^1$ from $\mathbf{x}_0^0$ and $\mathbf{P}_0^0$ and computes the one-step ahead prediction, $\mathbf{x}_1^0$ and its covariance $\mathbf{P}_1^0$ as well. The $\mathbf{e}_1$ and $\mathbf{\Sigma}_1$ are then computed using the one-step ahead prediction and its covariance. With the new estimates for the state vector and its covariance, namely, $\mathbf{x}_t^t$ and $\mathbf{P}_t^t$ where $t = 1, 2, ..., T-1$, the Gaussian Kalman filter repeats the updating procedures as from $\mathbf{x}_0^0$ to $\mathbf{x}_1^1$ till all $\mathbf{e}_t$ and $\mathbf{\Sigma}_t$ at different occasions are obtained. For details of the derivation of the algorithm, please refer to Harvey(1989, p.105-113) or Shumway and Stoffer(2000,p.330-339).

With all $\mathbf{e}_t$ and $\mathbf{\Sigma}_t$ at hand, $P_{\mathbf{\Theta}}\left(\mathbf{y}_t | \mathbf{Y}_{t-1}\right)$ can be calculated using (3.6). The likelihood function which equals the products of $P_{\mathbf{\Theta}}\left(\mathbf{y}_t | \mathbf{Y}_{t-1}\right)$ could then be obtained simply by multiplying all $P_{\mathbf{\Theta}}\left(\mathbf{y}_t | \mathbf{Y}_{t-1}\right)$. After some simple algebra, we may write the negative log-likelihood function in terms of $\mathbf{e}_t$ and $\mathbf{\Sigma}_t$,

$$-\,l\left(\theta\right) = C + \frac{1}{2} \sum_{t=1}^{T} log \left| \mathbf{\Sigma}_t \right| + \frac{1}{2} \sum_{t=1}^{T} \mathbf{e}_t^T \mathbf{\Sigma}_t^{-1} \mathbf{e}_t. \tag{3.11}$$

As shown above, the Gaussian Kalman filter serves like a device which yields the value of likelihood function based on our guesses of parameters. However, our main interest is to obtain parameter estimates rather than computing likelihood values given on parameters. To fulfill this purpose, some algorithm for finding approximations to ML estimates such as Newton-Raphson algorithm is used to update the parameter estimates. More specifically, the estimation procedure begins with initial guesses of the parameters. The Kalman filter then computes the likelihood value based on these guesses. With the likelihood value, the Newton-Raphson algorithm can update the estimates for parameters. By repeating this process, the sequence of estimates produces smaller negative log-likelihood functions of (3.11). Finally, when some criterion is met, the final estimates provide maximum likelihood estimates.

One important decision in applying the Kalman filter algorithm is to obtain the initial state vector and its covariance matrix. Note that the Gaussian Kalman filter is initialized with the mean and covariance matrix of the initial state vector, $\mathbf{x}_0$. For instance, consider an initial state vector which follows $N(\mu_0, \mathbf{P}_0)$. In order to get $\mathbf{e}_1$ and $\mathbf{\Sigma}_1$, a pair of values needs to be assigned to μ_0 and $\mathbf{P}_0$. Consequently, with finite samples, the settings of the initial state vector may influence the likelihood function and therefore the final estimates. Several ways of setting the initial conditions were developed so that the dependence on the initial conditions can be reduced. If stationarity of the state vector is assumed, the covariance matrix of the initial state vector, expressed in terms of other parameters in the model, can be defined as (Harvey, 1989, p121)

$$vec\left(\mathbf{P}_0\right) = \left[\mathbf{I} - \mathbf{\Phi} \otimes \mathbf{\Phi}\right]^{-1} vec\left(\mathbf{Q}\right). \tag{3.12}$$

The Gaussian Kalman filter algorithm which initializes with this $\mathbf{P}_0$ and a mean vector of zero provides exact MLE for stationary time series data. For non-stationary time series data, a so-called diffuse prior in which the covariance matrix $\mathbf{P}_0$ is defined as $\kappa\mathbf{I}$ is suggested (Harvey, 1989,p121-122; De Jong, 1988; Zarchan & Musoff, 2000). This is also called noninformative prior because κ is usually assigned a large positive number. Other algorithms concerning non-stationary state vectors include the method discussed by Ansley and Kohn (1985). In this thesis, only the $\mathbf{P}_0$ specified in (3.12) and diffuse prior are adopted.

3.3 Methods for Non-Gaussian Data

The Gaussian Kalman filter algorithm depends on the Gaussian assumption. The desirable properties of MLE for Gaussian processes, such as efficiency may not hold for non-Gaussian processes. To deal with non-Gaussian processes, many methods within the state space framework have been developed. For example, estimates can be obtained using numerical integration (Kitagawa, 1987), Monte Carlo techniques (Fruhwirth-Schnatter, 1994), Gaussian mixtures (Kawakatsu, 2007), particle techniques (Creal, 2008), Monte Carlo maximum likelihood (MCL) methods (Durbin & Koopman, 1997; Sandmann & Koopman, 1998), and Markov Chain Monte Carlo (MCMC) techniques from Bayesian perspective [1](Carlin, Polson, & Stoffer, 1992; Shumway & Stoffer, 2004; Jacquier, 1994).

Another method is the normal theory based ML estimation method which estimates the parameters by minimizing the Gaussian likelihood function regardless of whether or not the data are normally distributed (Harvey & Shephard, 1996). This method still uses the Gaussian Kalman filter algorithm. The only difference is that data are not normally distributed. In this thesis, the term "normal theory based ML approach" and "Gaussian Kalman filter approach" are used interchangeably. Sim-

[1]A review on MCMC method addressing non-normal state-space models can be found in Shumway and Stoffer (2004). The general MCMC method assumes the densities of the two noise variables are scale mixtures of normal distributions. This method first works on the complete data likelihood which assumes the latent state vectors are observed and then numerically integrates out the state vectors to obtain the marginal posterior distribution. Parameter estimates are obtained by computing certain statistics of the marginal posterior distribution, e.g., the expectation. If the non-normality comes from linearizing a nonlinear Gaussian model, some specific MCMC methods have been proposed (e.g., Jacquier, 1994). These methods generally directly work on the nonlinear Gaussian model.

ulation studies in economics showed that the performance of normal theory based ML depended on the signal-to-noise ratio and time series length. The signal-to-noise ratio is the ratio of the variation of state vectors to the variation of the measurement errors. According to the simulation results, with short time series length and low signal-to-noise ratio, the normal theory based ML estimates were less accurate and less efficient compared to MCMC methods (Sandmann & Koopman, 1998; Strickland et al., 2006).

However, normal theory based ML estimation is still an important and useful alternative to other existing methods due to the following three reasons. First, normal theory based ML yields comparable estimates when the variation in latent states dominates the variation in measurement errors. For example, great improvements of normal theory based ML have been observed when larger variance of the process disturbance is observed (Sandmann & Koopman, 1998; Strickland et al., 2006). The poor performance of normal theory based ML when variation in the state vector was small may be at least partially due to finite sample likelihood associated problems. As illustrated in table 2 in Sandmann and Koopman (1998), the MCL also suffered when the signal-to-noise ratio was low. Note that MCL is claimed to maximize the true likelihood function, which is approximated by sampling techniques. Thus this suggests that even if the true likelihood was maximized, the resulting estimates were not likely to perform well in the low signal-to-noise ratio situation (with a finite sample). The results found by Z. Zhang et al. (2008) where the performance of MLE for Gaussian process was studied paralleled the results above. And in other conditions, the improvement of MCL over normal theory based ML was not impressive (Sandmann & Koopman, 1998).

Second, those methods which involve numerical integration and sampling techniques have, in addition to their inefficiency in computational effort, nontrivial problems during implementation, i.e, convergence problems and sensitivities to model specification (here the specification refers to, for example, specification of prior distribution in MCMC methods, and specification of marginal distribution in particle filter methods). For example, the computational bottleneck of methods involving numerical integration when evaluating the prediction and updating densities makes it difficult to apply to dynamical systems with high dimensions (Fruhwirth-Schnatter, 1994; Kawakatsu, 2007). Small misspecification in marginal distributions can makes particle filter methods suffer (Creal, 2008), even when the observations are measured at high frequencies. Designs with intensive measurement are the situations when particle filter may provide moderate improvements over normal theory based ML. Convergence problems and sensitivities to prior distribution have long existed in Bayesian methods, such as MCMC approaches. Many factors may influence the severity of the problems. In real data analysis, sophisticated modifications of the prior distributions or of the sampler are needed to ensure convergence, although theoretically the sequence will always converge. Normal theory based ML is thus still valuable because the implementation is straightforward and the risks of mal-implementation are relatively low.

Third, other methods that are easy to implement such as GMM and EMM are at best as good as normal theory based ML. One issue associated with the results reported by Jacquier (1994) arises in the literature. Several researchers re-examined the performance of normal theory based ML reported in Jacquier (1994) and they found that the performance of normal theory based ML is "nowhere as near as bad as reported by JPR (Jacquier (1994))", (Sandmann & Koopman, 1998, p.284; Breidt & Carriquiry, 1996). Compared to other researchers results (Harvey & Shephard, 1996; Kirby, 2006), normal theory based ML seems to perform significantly poorer in Jacquier's paper. These researchers argued that Jacquier's results may suffer from problems of poor starting values, different convergence criteria or inefficient implementation of the algorithm and thus advocated paying special caution to the results concerning normal theory based ML method in Jacquier (1994). Subsequent work, such as that in Andersen, Chung, and Sorensen (1999), where the author's conclusions were rested on Jacquier's results, also need special attention (this paper simply quoted the results from Jacpuier's paper). If using Sandman and Koopman's results, normal theory based ML provides better estimates than EMM with a time series length of 500 and comparable estimates with EMM when the time series length increases to 2000 (Sandmann & Koopman, 1998; Andersen et al., 1999). GMM seems to be relatively inefficient and inaccurate compared to EMM and normal theory based ML.

In short, the normal theory based ML method or the Gaussian Kalman filter approach is a valuable alternative to other existing methods that cater for non-Gaussian processes. This thesis only focuses on the normal theory based ML approach. In the next chapter the asymptotic properties of normal theory based ML estimators and associated standard error estimators will be discussed.

CHAPTER 4

STANDARD ERROR ESTIMATORS FOR SSM

This chapter will introduce two types of SE estimators for state space models. The first type of SE estimators are based on the information matrix and they assume data are normally distributed. The second type of SE estimators are based on the sandwich asymptotic covariance matrix. They are proposed to adjust for violations of the normality assumption. Section 4.1 will introduce the asymptotic theory for normal data and SE estimators which are based on the information matrix. Section 4.2 will introduce the asymptotic theory for non-normal data and the sandwich type SE estimators.

Regularity conditions need to be assumed to ensure appropriateness of asymptotic theory for estimators. In this thesis, all state space models are assumed to be time invariant, stationary, controllable and observable so that the asymptotic theory introduced in the following sections is valid. Although the state space model can be time-varying, this is beyond the scope of this thesis. Stationarity, controllability and observability are typical requirements for state space models to use the asymptotic theory (Shumway & Stoffer, 2004, p.345-347). For definitions of stationarity, controllability and observability, please refer to Appendix B.

4.1 Information Based SE Estimators

4.1.1 Asymptotic Covariance Matrix and Standard Error

Under the Gaussian assumption, the asymptotic properties of ML estimates for the standard state space model have been well-studied. For Gaussian processes, under regularity conditions, the ML estimates are consistent and asymptotically normal (Shumway & Stoffer, 2004, p.345-347; Caines, 1988, p.426-480),

$$\sqrt{T}\left(\hat{\theta}-\theta_0\right) \to N\left(\mathbf{0}, \mathcal{I}^{-1}\left(\theta_0\right)\right)$$

where θ_0 is the true parameter and the information matrix, $\mathcal{I}\left(\theta\right)$, is

$$\mathcal{I}\left(\theta\right) \quad = \quad -\lim T^{-1}\sum_{t=1}^{T} E\left[\ddot{l}_t\left(\theta\right)\right]$$

where $\ddot{l}_t\left(\theta\right)$ is the second derivative of the log-likelihood function. The definition of the information matrix is slightly different from that of the information matrix for independent data. The limit of average $E\left[\ddot{l}_t\left(\theta\right)\right]$ is used here because, although $E\left[\ddot{l}_t\left(\theta\right)\right]$ becomes stable as $t \to \infty$, it can vary substantially with small ts.

The asymptotic covariance matrix is computed by dividing the inverse of the information matrix by T. And the standard error for the ith component in θ is the square root of the ith diagonal element in the covariance $\frac{\left[\mathcal{I}^{-1}(\theta_0)\right]}{T}$. Algebraically, $SE_{info,i} = \sqrt{\frac{\left[\mathcal{I}^{-1}(\theta_0)\right]_{ii}}{T}}$.

4.1.2 Standard Error Estimators

Three SE estimators have been proposed for $SE_{info,i}$. They are the observed information based SE estimator, $\widehat{SE}_O$ (Shumway & Stoffer, 2004, p.347), the expected information based SE estimator, $\widehat{SE}_E$ (Cavanaugh & Shumway, 1996) and Harvey's SE estimator, $\widehat{SE}_H$ (Harvey, 1989, p.140-142).

The observed information based SE estimator uses the negative Hessian matrix as an estimator of $\mathcal{I}\left(\theta_0\right)$. The negative Hessian matrix, also called the observed information matrix, is given by,

$$\widehat{\mathcal{I}}_O = -\frac{1}{T}\sum_{t=1}^{T}\ddot{l}_t\left(\hat{\theta}\right). \tag{4.1}$$

As it is shown in the above equation, the Hessian matrix is the second derivative of the log-likelihood function. The SE estimates are then obtained by the equation, $\widehat{SE}_{O,i} = \sqrt{\frac{\left(\widehat{\mathcal{I}}_O^{-1}\right)_{ii}}{T}}$.

The expected information based SE estimator takes the expectation of the Hessian matrix, and uses it to compute the SE estimates. Algebraically, $\widehat{SE}_E$ is given by,

$$\widehat{SE}_{E,i} = \sqrt{\frac{\left(\widehat{\mathcal{I}}_E^{-1}\right)_{ii}}{T}}$$

where

$$\widehat{\mathcal{I}}_E = -\frac{1}{T}\sum_{t=1}^{T}E\left[\ddot{l}_t\left(\hat{\theta}\right)|\mathcal{H}_0\right]. \tag{4.2}$$

The expectation operation is taken under the assumption ($\mathcal{H}_0$) that the data, $\{\mathbf{y}_t\}$, follow normal distributions and that the model is correctly specified.

The third estimator is Harvey's SE estimator. When Harvey derived the estimator of the information matrix, he first took the expectation of the Hessian matrix so that the formula was greatly simplified. After taking expectation, he got the expected information matrix as follows (Harvey, 1989, p.142),

$$\mathcal{I}_{E,ij}(\theta) = -\frac{1}{T}E\left(\frac{\partial^2 lnL}{\partial\theta_i\partial\theta_j}\right) = \frac{1}{T}\sum_{t=1}^{T}\left\{E\left(\left(\frac{\partial\mathbf{e}_t}{\partial\theta_i}\right)'\mathbf{\Sigma}_t^{-1}\frac{\partial\mathbf{e}_t}{\partial\theta_j}\right) - \frac{1}{2}tr\left(\frac{\partial\mathbf{\Sigma}_t^{-1}}{\partial\theta_i}\frac{\partial\mathbf{\Sigma}_t}{\partial\theta_i}\right)\right\}$$

where $\frac{\partial^2 lnL}{\partial\theta_i\partial\theta_j}$ is the second derivative of the log-likelihood function. The computation of the first term, however, is very complicated (Cavanaugh & Shumway, 1996). Harvey then suggested replacing expectation of the first term by current estimates. The resulting matrix, namely, $\mathcal{I}_H$, can be used as an approximation of $\mathcal{I}(\theta_0)$. Replacing parameters in $\mathcal{I}_H$ by parameter estimates will yield $\widehat{\mathcal{I}}_H$. $\widehat{\mathcal{I}}_H$ can be used to compute SE estimator. The resulting SE estimator is defined as $\widehat{SE}_H = \sqrt{\frac{\left(\widehat{\mathcal{I}}_H^{-1}\right)_{ii}}{T}}$.

Harvey's SE estimator only involves first derivatives and thus is the most computationally efficient. The expected information based SE involves complicated recursions to take the expectation and thus it is the most computationally intensive. In this thesis only the observed information based SE and Harvey's SE estimator are studied.

4.2 Sandwich Type SE Estimators

4.2.1 Asymptotic Covariance Matrix and Standard Error

When the distribution assumption is violated, while the point estimates are still consistent (Dunsmuir, 1979; Caines, 1988), the standard error estimates may be biased. Several lines of research have studied the asymptotic theory for various time series models when data are not normally distributed (Dunsmuir, 1979; Dunsmuir & Hannan, 1976, 1978; Caines, 1988; White, 1984). They arrived at similar conclusions of asymptotic normality, although they used different sets of conditions and they addressed the problem from different perspectives, i.e., from frequency domain or time domain. Only Caines' work is presented here to illustrate the asymptotic normality of (Q)ML estimates because all formulas in Caines' work are expressed in the form of time domain representation.

Caines (1988) studied the asymptotic properties of a general class of estimators for non-Gaussian dependent data. This class of estimators includes (Q)ML estimators, least square estimators, and other estimators that minimize a discrepancy function of prediction errors. Let $\theta_T^\star$ be the vector that maximizes $\bar{L}_T = T^{-1}\sum E\left[l_t(\theta)\right]$ where $l_t(\theta)$ is the log-likelihood function at time t. Under mild regularity conditions (Caines, 1988; p.488-489, p.494-496, p.498-500, p.518-522),

$$\sqrt{T}\mathbf{B}_T^{-1/2}\mathbf{A}_T\left(\widehat{\theta}_T - \theta_T^\star\right) \to N\left(\mathbf{0}, \mathbf{I}\right), \tag{4.3}$$

where

$$\mathbf{A}_T = -T^{-1}\sum_{t=1}^{T} E\left[\ddot{l}_t\left(\theta_T^\star\right)\right]$$

and

$$\mathbf{B}_T = var\left(T^{-1/2}\sum_{t=1}^{T}\dot{l}_t\left(\theta_T^{\star}\right)\right) = T^{-1}E\left[\sum_{t=1}^{T}\dot{l}_t\left(\theta_T^{\star}\right)\sum_{t=1}^{T}\dot{l}_t'\left(\theta_T^{\star}\right)\right].$$

So the estimator $\hat{\theta}_T$ roughly follows a normal distribution with a mean of $\theta_T^{\star}$ and a covariance matrix of $\mathbf{A}_T^{-1}\mathbf{B}_T\mathbf{A}_T^{-1}/T$.

If further assumes that $\mathbf{B} = \lim \mathbf{B}_T$ exist then,

$$\sqrt{T}\left(\widehat{\theta}_T - \theta^{\star}\right) \to N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right), \tag{4.4}$$

where $\mathbf{A} = \lim \mathbf{A}_T$ and $\theta^{\star}$ becomes the maximum of $\bar{L} = \lim \bar{L}_T$.

Caines' conditions are complicated and further work on whether a standard state space model satisfies those conditions is needed. Results from Dunsmuir's (1976) work, however, can be used to show that the sampling distribution of (Q)ML estimates for a standard state space model does follow a normal distribution. The asymptotic covariance matrix that Dunsmuir obtained (Dunsmuir, 1979) also had the form of sandwich, that is, $\mathbf{\Omega}^{-1}\mathbf{\Pi}\mathbf{\Omega}^{-1}$. And $\mathbf{\Omega}$ (or$\mathbf{\Pi}$) was also the limiting value of the second (the variance of the first) derivative of the log-likelihood function. But the likelihood function used in Dunsmuir's paper was defined from the frequency perspective. Further work is needed to examine whether the asymptotic covariance matrix derived from frequency domain is the same as Caines' covariance matrix. I also include some discussion on the regularity conditions in Appendix C.

4.2.2 Standard Error Estimators

Three sandwich type SE estimators have been proposed based on asymptotic normality (4.3). These SE estimators rest on different estimators of the covariance matrix, namely, $\mathbf{A}_T^{-1}\mathbf{B}_T\mathbf{A}_T^{-1}/T$. The two sandwich type SE estimators proposed by Papanastassiou (2006) use $\widehat{\mathcal{I}}_E$ as the estimator of $\mathbf{A}_T$. For estimation of $\mathbf{B}_T$, Papanastassiou (2006) showed numerically that the distances between $\widehat{\mathbf{B}}_2 = T^{-1}\sum_{t=1}^{T} E\left[\dot{l}_t\left(\widehat{\theta}\right)\dot{l}_t'\left(\widehat{\theta}\right)\middle|\mathcal{H}_0\right]$ and $\mathbf{B}_T$ and distances between $\widehat{\mathbf{B}}_1 = T^{-1}\sum_{t=1}^{T}\dot{l}_t\left(\widehat{\theta}\right)\dot{l}_t'\left(\widehat{\theta}\right)$

and $\mathbf{B}_T$ were negligible. The distances were measured in terms of the relative MSE of the ordered eigenvalues of the matrices and the distances are smaller than 1×10^{-3}. And then Papanastassiou (2006) simply used the two combinations of $\widehat{\mathcal{I}}_E$, $\widehat{\mathbf{B}}_1$ and $\widehat{\mathbf{B}}_2$ to estimate the covariance matrix. More specifically, the first SE estimator is given by,

$$\widehat{SE}_{SW_{P1},i} = \sqrt{\left(\widehat{\mathcal{I}}_E^{-1}\widehat{\mathbf{B}}_1\widehat{\mathcal{I}}_E^{-1}\right)_{ii}/T}$$

and the second SE estimator is given by,

$$\widehat{SE}_{SW_{P2},i} = \sqrt{\left(\widehat{\mathcal{I}}_E^{-1}\widehat{\mathbf{B}}_2\widehat{\mathcal{I}}_E^{-1}\right)_{ii}/T}.$$

The expectation operation in $\mathbf{A}_T$ is generally complicated to compute. $\widehat{\mathcal{I}}_O$ is an alternative estimator other than $\widehat{\mathcal{I}}_E$, because if the asymptotic normality is validate $\widehat{\mathcal{I}}_O$ is also a consistent estimator of $\mathbf{A}_T$ in that $\widehat{\mathcal{I}}_O - \mathbf{A}_T \to 0$ as $T \to \infty$ (Caines, 1988, p.523-524). Thus in this thesis, I only include $\widehat{SE}_{SW_{P1},i}$ and I drop the expectation operation in $\widehat{SE}_{SW_{P1},i}$.

White (1984) also proposed a sandwich type SE estimator based on asymptotic normality discussed here. Instead of using the expected information based SE estimator to estimate $\mathbf{A}_T$, the observed information based SE estimator, $\widehat{\mathcal{I}}_O$, was used. For $\mathbf{B}_T$, White (1984) proposed a new estimator, namely,

$$\begin{aligned}\widehat{\mathbf{B}}_3 \quad = \quad & T^{-1}\left\{\sum_{t=1}^{T} \dot{l}_t\left(\widehat{\theta}\right)\dot{l}_t^T\left(\widehat{\theta}\right) + \sum_{\tau=1}^{s}\sum_{t=\tau+1}^{T}\left[\dot{l}_t\left(\widehat{\theta}\right)\dot{l}_{t-\tau}^T\left(\widehat{\theta}\right)\right.\right. \\ & \left.\left. +\dot{l}_{t-\tau}\left(\widehat{\theta}\right)\dot{l}_t^T\left(\widehat{\theta}\right)\right]\right\},\end{aligned}$$

where $s = o\left(T^{1/3}\right)$ and $s \to \infty$ as $T \to \infty$. Thus the formula for SE estimator is,

$$\widehat{SE}_{SW_W,i} = \sqrt{\left(\widehat{\mathcal{I}}_O^{-1}\widehat{\mathbf{B}}_3\widehat{\mathcal{I}}_O^{-1}\right)_{ii}/T}.$$

In short, this thesis will explore the finite sample properties of four SE estimators, $\widehat{SE}_O$, $\widehat{SE}_H$, $\widehat{SE}_{SW_{P1}}$ and $\widehat{SE}_{SW_W}$ under different conditions. And in the next chapter the specific design for simulation studies will be described.

CHAPTER 5

SIMULATION DESIGN

The main objective of this thesis is to compare SE estimators for dynamic factor analysis when data are non-normally distributed. Five factors that may influence the performance of SE estimators are studied: distributions of the data, the time series length, the number of subjects, initial settings and the signal-to-noise ratio. Section 5.1 discusses how these five factors may influence the performance of SE estimators. Section 5.2 describes the simulation design.

5.1 Possible Influences on the Performance of SE Estimator

Distribution of the data is one key factor that would influence the performance of SE estimators. Since the information based SE estimators are derived under the assumption that data are normally distributed, it is expected that they would be, to some extent, influenced by violations of the normal assumption. Results from independent data showed that the information based SE estimators fail to give consistent estimation of standard errors because they ignore the influences of kurtosis (Yuan & Hayashi, 2006). Consequently, sandwich type SE estimators which are derived from the asymptotic normality that takes into account the deviation from the Gaussian distribution are then recommended (Yuan & Hayashi, 2006). However, the inferences of kurtosis on some parameter estimates of dependent data may be limited. For example, Dunsmuir (1979) proved that the asymptotic covariance matrix for AR and MA weights in a vector ARMA model does not depend on the fourth moment of random shocks. The asymptotic covariance matrix for these parameters is actually in the same form as when data are normally distributed. But the asymptotic covariance matrix for covariance components, such as the variances of random shocks, is still influenced by kurtosis. Note that Dunsmuir's results are legitimate asymptotically. Finite sample performance may be quite different. The vector ARMA model is a highly related model. As shown in section 3.1, it is a special case of the state space model.

The second factor concerning the finite performance of SE estimators is the time series length. The performance of SE in long time series could be very different from the performance in short time series. This is because large sample performances are governed by theorems of convergence. These theorems are applicable only when time series length becomes fairly large. Finite sample properties of consistent SE estimators need to be carefully studied.

The third factor that influences the performance of SE estimators is initial settings. Two types of initial settings that can be commonly seen in substantive research are the noninformative initial setting and the stationary initial setting. When the noninformative initial setting is used, a diagonal matrix, $\kappa\mathbf{I}$, where κ is usually a large positive number, is utilized as the initial factor covariance matrix. This type of initial setting is suitable for data with long time series length. When time series is short, point estimates and likely SE estimates would be influenced and more biased. The stationary initial setting expresses the unknown initial factor covariance matrix in terms of other model parameters. By equation, the initial factor covariance matrix is given by $vec\,(\mathbf{P}_0) = [\mathbf{I} - \mathbf{\Phi} \otimes \mathbf{\Phi}]^{-1}\, vec\,(\mathbf{Q})$. This type of initial setting is suitable for stationary data.

Another factor that influences SE estimation is the number of participants. Extra participants may contribute extra information and thus with more participants, the performance of (quasi-)maximum likelihood ((Q)ML) estimators and SE estimators can be expected to be better. In other words, there are two ways to increase the number of observations, one is to increase the time series length and the other is to collect more participants. In this thesis the combination effects of time series length and the number of participants will be examined. This combined effect is of interest because in real data analysis, collecting observations from extra participants is less expensive then increasing the time series length. Suppose a Researcher A collects data from 50 participants and each has 10 observations while another Researcher B surveys 10 participants but each has 50 observations. Both data sets have 500 observations, but the latter design is likely to be more expensive. It is thus worthwhile to see whether the performance of quasi-maximum likelihood estimators and SE estimators for the first data set is comparable to performance of estimators for the second data set. This may help for substantive researchers choose less expensive designs while maintaining the performance of estimators.

The signal-to-noise ratio is the last factor. It is the ratio of variance that can be explained by latent processes to variance of measurement errors. In a standard state space model, the signal-to-noise ratio for the ith component of a stationary process, $\{\mathbf{y}_t\}$, is defined as, $\frac{(\mathbf{A}'\mathbf{P}\mathbf{A})_{ii}}{(\mathbf{R})_{ii}}$, where $\mathbf{A}$ is the observation matrix; $\mathbf{P}$ is the steady covariance matrix of the latent state; and $\mathbf{R}$ is the covariance matrix of the measurement errors. The influences of the signal-to-noise ratio on (Q)ML estimates have been studied indirectly. Previous research found that the finite sample performance of (Q)ML estimator was influenced by the variability of the latent factors (Jacquier, 1994; Harvey & Shephard, 1996). With larger variation in the latent factors, the (Q)ML estimates got closer to the true value and yielded smaller root mean square errors (Jacquier, 1994; Harvey & Shephard, 1996). However, in those models, the variation of measurement errors and the factor loadings were known. Thus the key factor that influenced the performance of (Q)ML estimator was the signal-to-noise ratio, which is the ratio of the variation of latent factors to the variation of measurement errors. SE estimation may also be vulnerable to a low signal-to-noise ratio. To ensure that it is the signal-to-noise ratio that influences that the performance of SE estimates, the signal-to-noise ratio will be manipulated in different ways and results of conditions varying in how signal-to-noise ratio is manipulated will be compared.

In short, the thesis investigates the influences of these five factors using one simulation study.

5.2 Simulation Design

The finite sample performance of two types of standard error estimators will be studied under various distribution, length, number of participants, initial setting, and signal-to-noise ratio conditions using a Monte Carlo simulation.

5.2.1 The True Model

A bivariate AR(1) latent state process measured by six indicators in total constitute the true model. The measurement model is:

$$\begin{array}{ccccccc} \mathbf{y}_t & = & \mathbf{A} & \mathbf{x}_t & + & \mathbf{v}_t \\ \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \\ y_{5t} \\ y_{6t} \end{bmatrix} & = & \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{12} & 0 \\ \lambda_{13} & 0 \\ 0 & \lambda_{21} \\ 0 & \lambda_{22} \\ 0 & \lambda_{23} \end{bmatrix} & \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} & + & \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \\ v_{6t} \end{bmatrix} \end{array} \tag{5.1}$$

in which $\lambda_{11} = \lambda_{21} = 1$ and $\lambda_{12} = \lambda_{13} = \lambda_{22} = \lambda_{23} = .8$. The observation noise variables, $\mathbf{v}_t$, are identically and independently distributed with a mean vector of zero and a diagonal covariance matrix of $\mathbf{R}$. The state vector $\mathbf{x}_t$ follows a VAR(1) model, that is,

$$\mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{w}_t \tag{5.2}$$

where $\mathbf{\Phi} = \begin{bmatrix} .6 & .3 \\ .3 & .6 \end{bmatrix}$ and $\mathbf{w}_t \sim MD\left(0, \mathbf{Q}\right)$. $MD\left(\mu, \mathbf{\Sigma}\right)$ represents a distribution with a mean of μ and a covariance matrix of $\mathbf{\Sigma}$. The covariance matrix of the initial latent factor $\mathbf{x}_0$, $\mathbf{P}_0$, satisfies the equation (3.12). This will ensure that the time series is generated from a stationary system. Figure 1 gives the visual representation of this model.

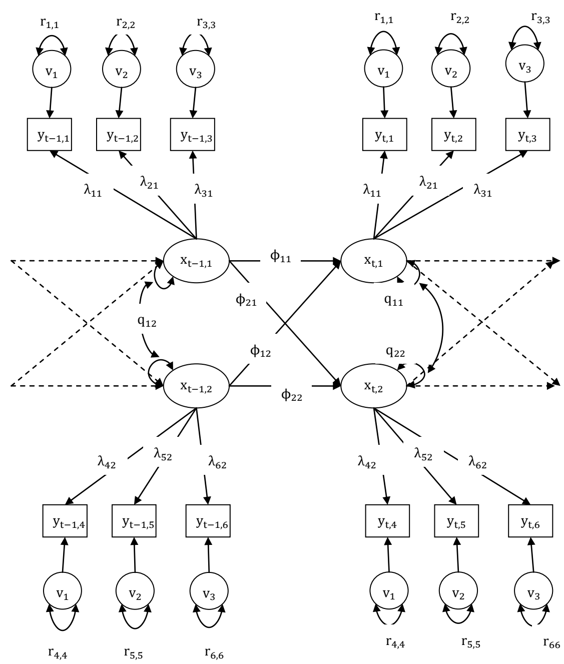

Figure 5.1: The path diagram of the simulated model.

5.2.2 Factors Manipulated

Five variables are manipulated in this study: the distribution of data (of noise variables), time series length, number of participants, initial settings and the signal-to-noise ratio. For distributions of data (of noise variables), three different distributions will be used: Normal, log-normal (with mean shifted to $\mathbf{0}$), and $\epsilon-$contaminated normal ($\epsilon = .05$, c=15). Log-normal is included to study the impact of skewness and kurtosis while $\epsilon-$contaminated normal are used to study the influence of kurtosis. All of the distributions will share the same first two moments. Specifically, the means of the two noise variables are fixed at zero. The covariance matrices, $\mathbf{R}$ and $\mathbf{Q}$, under different signal-to-noise conditions are summarized in table 5.2.1. The third and fourth moments of the two noise variables vary across different distribution conditions. For example, in the contaminated normal condition, the noise variables still have zero mean and the same second and third moments as that in normal condition, but the kurtosis is different. The kurtosis in the contaminated normal condition is higher than that in the normal condition. Table 5.2.2 reports the measures of skewness and kurtosis of the two noise variables under different distributional conditions.

TABLE 5.1

PARAMETER VALUES UNDER DIFFERENT SIGNAL-TO-NOISE RATIO CONDITIONS

$\mathbf{P}_{ii}$	2.5	1	0.5
$r_{s/n} = 5$ or 3.2	Model 1-1 $\mathbf{Q}_3\mathbf{R}_3$	Model 1-2 $\mathbf{Q}_2\mathbf{R}_2$	Model 1-3 $\mathbf{Q}_1\mathbf{R}_1$
$r_{s/n} = 1$ or 0.64	Model 2-1 $\mathbf{Q}_3\mathbf{R}_6$	Model 2-2 $\mathbf{Q}_2\mathbf{R}_4$	Model 2-3 $\mathbf{Q}_1\mathbf{R}_3$
$r_{s/n} = .25$ or .16	Model 3-1 $\mathbf{Q}_3\mathbf{R}_8$	Model 3-2 $\mathbf{Q}_2\mathbf{R}_7$	Model 3-3 $\mathbf{Q}_1\mathbf{R}_5$
Resulting $\mathbf{P}$	$\begin{bmatrix} 2.5 & 1 \\ 1 & 2.5 \end{bmatrix}$	$\begin{bmatrix} 1 & .4 \\ .4 & 1 \end{bmatrix}$	$\begin{bmatrix} .5 & .2 \\ .2 & .5 \end{bmatrix}$

TABLE 5.2

MEASURES OF SKEWNESS AND KURTOSIS FOR THE TWO NOISE VARIABLES

		Normal				LogNormal				Contaminated Normal			
		Mardia		Marginal		Mardia		Marginal		Mardia		Marginal	
	Covariance	Skew	Kurt	Skew	Kurt	Skew	Kurt	Skew	Kurt	Skew	Kurt	Skew	Kurt
$\mathbf{w}_t : \mathbf{Q}$	$\begin{bmatrix} .203 & -.07 \\ -.07 & .203 \end{bmatrix}$	0	8	0	3	3.750	14.582	1.320	6.25	0	33.772	0	12.664
	$\begin{bmatrix} .406 & -.14 \\ -.14 & .406 \end{bmatrix}$	0	8	0	3	7.405	21.813	1.843	9.588	0	33.772	0	12.664
	$\begin{bmatrix} 1.015 & -.35 \\ -.35 & 1.015 \end{bmatrix}$	0	8	0	3	18.496	46.875	2.865	20.465	0	33.772	0	12.664
$\mathbf{v}_t : \mathbf{R}$	$diag\,(.1, 6)$	0	48	0	3	5.254	57.578	.936	4.596	0	202.630	0	12.664
	$diag\,(.2, 6)$	0	48	0	3	10.305	67.209	1.311	6.202	0	202.630	0	12.664
	$diag\,(.5, 6)$	0	48	0	3	24.883	97.069	2.036	11.178	0	202.630	0	12.664
	$diag\,(1, 6)$	0	48	0	3	48.541	151.082	2.844	20.180	0	202.630	0	12.664
	$diag\,(2, 6)$	0	48	0	3	96.000	276.000	4.000	41.000	0	202.630	0	12.664
	$diag\,(2.5, 6)$	0	48	0	3	120.174	346.698	4.475	52.783	0	202.630	0	12.664
	$diag\,(4, 6)$	0	48	0	3	194.954	590.125	5.700	93.354	0	202.630	0	12.664
	$diag\,(10, 6)$	0	48	0	3	526.931	1993.631	9.371	327.272	0	202.630	0	12.664

For time series length and number of participants, eight conditions are included. To evaluate the influence of time series length, four length conditions are used, $T = 50$, 100, 200 and 500. 50 and 100 are time series length that can be commonly seen in psychological dynamic studies. These are chosen to study performance in small to moderate samples. The two longer time series length are included for comparison purpose. To evaluate the combined effects of time series length and number of participants, four conditions that share the same total number of observations but differ in time series length and the number of participants are considered. Specifically, the following four conditions are adopted: $N = 10$ & $T = 50$, $N = 20$ & $T = 25$, $N = 25$ & $T = 20$, and $N = 50$ & $T = 10$.

Two types of initial settings are used, the noninformative initial setting and the stationary initial setting. The noninformative initial setting uses a diagonal matrix $\kappa\mathbf{I}$ as the initial factor covariance matrix. In this thesis, κ is set to be 100. The stationary initial setting expresses the unknown initial factor covariance matrix in terms of other model parameters. By equation, the initial factor covariance matrix is given by $vec\,(\mathbf{P}_0) = [\mathbf{I} - \mathbf{\Phi} \otimes \mathbf{\Phi}]^{-1}\, vec\,(\mathbf{Q})$.

For signal-to-noise ratio conditions, three levels are chosen: high, moderate and low. Signal-to-noise ratio is defined as $(\mathbf{APA})_{ii}\,/\mathbf{R}_{ii}$ where $vec\,(\mathbf{P}) = [\mathbf{I} - \mathbf{\Phi} \otimes \mathbf{\Phi}]^{-1}$ $vec\,(\mathbf{Q})$. When time series approaches steady state, the signal-to-noise ratio can be manipulated by altering values of $\mathbf{A}$, $\mathbf{R}$, $\mathbf{\Phi}$ and $\mathbf{Q}$. To simplify the problem, this thesis uses one set of $\mathbf{A}$ and $\mathbf{\Phi}$. With a given signal-to-noise ratio, three sets of $\mathbf{R}$ and $\mathbf{Q}$ are chosen to obtain the target signal-to-noise ratio. Specifically, I first choose three $\mathbf{Q}$s so that the variance of latent factors, $\mathbf{P}_{ii}$, equal to .5, 1 and 2.5. After choosing $\mathbf{Q}$s, I pick up the variance of the measurement errors which, in combination of the predefined $\mathbf{A}$, $\mathbf{\Phi}$ and $\mathbf{Q}$, will achieve the target signal-to-noise ratio. The specific values of the model parameters in this study are listed as follows: (a) the covariance matrix of $\mathbf{w}_t$ has three conditions $\mathbf{Q}_1 = \begin{bmatrix} .203 & -.07 \\ -.07 & .203 \end{bmatrix}$, $\mathbf{Q}_2 = \begin{bmatrix} .406 & -.14 \\ -.14 & .406 \end{bmatrix}$, and $\mathbf{Q}_3 = \begin{bmatrix} 1.015 & -.35 \\ -.35 & 1.015 \end{bmatrix}$; (b) the covariance matrix of measurement errors will be assigned to seven different sets of values, $\mathbf{R}_1 = diag\,(.1, 6)$, $\mathbf{R}_2 = diag\,(.2, 6)$, $\mathbf{R}_3 = diag\,(.5, 6)$, $\mathbf{R}_4 = diag\,(1, 6)$, $\mathbf{R}_5 = diag\,(2, 6)$, $\mathbf{R}_6 = diag\,(2.5, 6)$ and $\mathbf{R}_7 = diag\,(4, 6)$, $\mathbf{R}_8 = diag(10, 6)$. Table 4.1 lists models with different signal-to-noise ratios $r_{s/n}$ under each distributional condition. In short, three levels of signal-to-noise ratio are included and for each signal-to-noise ratio condition, the signal-to-noise ratio is manipulated via three ways.

To summarize, the simulation study is carried out using a $3 \times 8 \times 2 \times 3 \times 3$ (Distribution×T_N×Initial Settings×Ratio×Ways of Manipulating Ratio) design.

5.2.3 Data Generation

With predefined model parameters, the following procedures are employed to generate samples accordingly:

S1. Generate an initial state vector, $\mathbf{x}_0$ from $MD\,(\mathbf{0}, \mathbf{P}_0)$. This thesis uses the $\mathbf{P}_0$ as defined in (3.12).

S2. Generate a process noise vector $\mathbf{w}_t$ from $MD\,(\mathbf{0}, \mathbf{Q})$ where $t = 1, 2, \ldots, T$.

S3. Compute a new state vector $\mathbf{x}_t$ using the equation (3.2) and previous latent factor, $\mathbf{x}_{t-1}$.

S4. Generate an observation noise vector $\mathbf{v}_t$ from $MD\,(\mathbf{0}, \mathbf{R})$.

S5. Compute a new observation, $\mathbf{y}_t$, using the equation (3.1) and current latent factor, $\mathbf{x}_t$.

S6. Update the time indicator subscript t and repeat Step 2-5 T times.

S7. Save the T data points.

Note that MD can be any of the four aforementioned distributions. Appendix A describes how to generate samples from the four aforementioned distributions. And in the above procedure, no burn-in periods are used because the initial conditions have taken into account the possible influences of the settings of initial conditions.

5.2.4 The Fitted Model

Data are fitted by the true model, that is, a latent vector AR(1) model with three indicators for each factor. Algebraically, the fitted model is defined by (5.1) and (5.2). λ_{11} and λ_{21} are fixed to one to help identification. For each simulated sample, the (Q)ML estimates and the four SE estimates $\widehat{SE}_O$, $\widehat{SE}_H$, $\widehat{SE}_{SW_{P1}}$ and $\widehat{SE}_{SW_W}$ will be computed. In the computation of $\widehat{SE}_{SW_W}$, s will be set to $s = \left[T^{1/3.1}\right]$ where $[x]$ is the largest integer that is no larger than x. $\left[T^{1/3.1}\right]$ is used so that more cross-product terms are included and thus the difference between $\widehat{SE}_{SW_W}$ and $\widehat{SE}_{SW_{P1}}$ is adequately large.

Results of interest are convergence rates, the performance of the normal theory based ML estimator, the performance of standard error estimators and interval estimates. The accuracy of the (Q)ML estimator is evaluated using bias. The bias is given by $bias_i = M^{-1}\sum_{j=1}^{M}\left(\hat{\theta}_{i,j} - \theta_i\right)$ where $\hat{\theta}_{i,j}$ is the estimate of the ith parameter obtained when analyzing the jth replication; θ_i is the true value of the ith parameter; and M is the number of converged replications in the simulation study. To evaluate the performance of SE estimators, the relative bias of the ith parameter, $rbias_i = M^{-1}\sum_{j}^{M}\left(\widehat{SE}_{i,j} - SE_{emp,i}\right)/SE_{emp,i}$, is computed. $SE_{emp,i}$ is the empirical SE estimate of the ith parameter.

Generally speaking, SE estimators provide information on the accuracy of point estimates. They can be used to test a hypothesis about parameters and to construct confidence intervals. A confidence interval gives a range of plausible values for parameters. In addition to the single value provided by point estimates, a confidence interval can tell how likely the true value will fall into the interval. If the SE estimator of interest gives good estimates of the true standard error and the sampling distribution of the target consistent statistic is symmetric, the probability that the confidence interval computed using this SE estimator fails to cover the true value should be close to the nominal level. This probability is called the mis-coverage rate. Thus, by examining how close to the nominal level the empirical mis-coverage rate is, we can examine the performance of SE estimator indirectly. This is important because in real data analysis, researchers are more concerned about the accuracy of the inferences they make based on SE estimators. In thesis, the mis-coverage rates of confidence intervals based on different SE estimators will be studied.

The parameter estimation is conducted using the nlm optimizer in R and when a Heywood case is encountered, the sample is analyzed again using the optim optimizer. All SE estimates are computed using a customized program written in R.

CHAPTER 6

RESULTS

Simulation results are reported in this chapter. Section 6.1 discusses convergence rates for all conditions. Section 6.2 and section 6.3 summarize results concerning point estimates and standard error estimates respectively. Since different ways of manipulating signal-to-noise ratios do not influence the bias of point estimates and the relative bias of standard error estimates in most conditions, I decide to focus on one way of manipulating signal-to-noise ratios (models in the second column of table 5.1). Conditions in which the influences of methods of manipulating signal-to-noise ratios are observed are summarized and discussed in Appendix D. Detailed results of all conditions are included in the supplemental file and are available on request to the author.

6.1 Convergence Rates

The number of samples to be included in the analysis is summarized in table 6.1 and table 6.2. In each condition, 1000 samples are simulated. Samples that do not converge or samples with estimated covariance matrices that are not positive-semidefinite are excluded from further analysis in this thesis. Since results are based on clean samples, one should not overgeneralize the findings of this thesis to irregular samples.

From the following table we can see that convergence rates are influenced by the time series length. As the time series length increases, convergence rates increase.

Adding participants also helps improving convergence rates. And as a product of the time series length and the number of participants, the total number of observations also influences convergence rates. Convergence rates increase as the total number of observations amplifies. The effect of the data distribution depends on the total number of observations. When the total number of observations is small, the convergence rates in the normal data condition are substantially higher than in the non-normal data conditions. However, when the total number of observations is large, the differences in convergence rates across distribution conditions are negligible. Initial settings also show influences on convergence rates. Convergence rates are lower when the stationary initial setting is used than when the noninformative initial setting is used. This is perhaps due to the fact that when the stationary initial setting is used, initial factor covariance matrix is based on the estimates of model parameters. The likelihood function is then more complicated and thus it is more difficult to find the solution. Therefore convergence rates are lower when the stationary initial setting is used.

The effect of the signal-to-noise ratio is not uniform. As shown in table 6.1 and table 6.2, when the total number of observations is large, the convergence rates are lower in high ratio conditions. One possible explanation of this seemingly counterintuitive phenomenon is that convergence rates are influenced by two factors which bind to each other in the simulation study but affect convergence rates differently. These two factors are the signal-to-noise ratio and the distance between the true parameter values and the boundary of parameter space. In this thesis, signal-to-noise ratios are manipulated by adjusting the variances of measurement errors. As a result, the variances of measurement errors are closer to 0, the boundary of parameter space, in the high ratio conditions than in the low ratio conditions. Hence in this thesis, a higher signal-to-noise ratio is always associated with a shorter distance and a lower ratio with a longer distance. However, these two factors, ratio and distance, affect convergence rates differently. When the total number of observations is small, the influence of the signal-to-noise ratio dominates. And since a higher ratio indicates more information, convergence rates thus increase as the signal-to-noise ratio increases. This is confirmed by the fact that the empirical standard error estimates in the low ratio conditions are larger than in the high ratio conditions. However, the influences of signal-to-noise ratios reduce as the total number of observations increases. This is because when the total number of observations is large, even the data sets in the low ratio condition contain enough information to ensure convergence. In other words, when the total number of observations is sufficiently large, the influence of signal-to-noise ratios is negligible and the effect of distances becomes more obvious. When the true parameter value is closer to the boundary of parameter space, the chances of suffering from convergence problems are higher and thus the convergence rates are lower. Since the distance between the true parameter value and the boundary of parameter space is smaller in the high ratio condition, the convergence rates are lower in the high ratio condition when the total number of observations is sufficiently large.

Convergence rates appear to be somewhat lower than some previous papers, i.e., Z. Zhang et al. (2008) and Song and Ferrer (2009). This is perhaps due to three reasons. First, the signal-to-noise ratio in this study is lower. For example, Z. Zhang et al. (2008) used models with signal-to-noise ratios ranging from 1 to 10 and Song and Ferrer (2009) used signal-to-noise ratios ranging from around 2.5 to 16. In this thesis, however, the signal-to-noise ratio ranges from .25 to 5. When the total number of observations is small, models with lower ratios tend to have lower convergence rates. Second, different initial settings are used in this thesis. In the two previous papers, the initial factor covariance matrix was treated as known and it was computed based on true parameter values and the stationarity assumption. In other words, these two papers used extra information that is usually unknown in real data analysis during the estimation process. In contrast, the two initial settings used in this thesis do not rely on the information that is usually unknown in the real data analysis. The stationary initial setting treats the initial factor covariance matrix as unknown and estimates the initial factor covariance matrix based on estimates of other parameters. Therefore the estimation using this initial setting is more difficult. The noninformative initial setting uses $\kappa\mathbf{I}$ as the initial factor covariance matrix where κ is an arbitrary large positive number. In this thesis $\kappa = 100$. This matrix is quite far away from the true initial covariance matrix. Thus it is not surprise that lower convergence rates are observed. Third, Song and Ferrer (2009) used the expectation-maximization algorithm (EM) to obtain point estimates. It is generally accepted that the EM algorithm has higher convergence rates than the Newton-type algorithm. Since the Quasi-Newton algorithm was used in Z. Zhang et al. (2008) and the Newton-Raphason algorithm is used in this thesis, it is not surprise that the convergence rates in Song and Ferrer (2009) were higher than the convergence rates in Z. Zhang et al. (2008) and in the current thesis. The shortcoming of the EM algorithm is that SE estimates cannot be obtained directly. As noted in Song and Ferrer (2009, p360), even though there exist some approaches to overcome this limitation, e.g., a modified EM algorithm (Meng & Rubin, 1991) and a bootstrapping method specific to DFA models (G. Zhang & Browne, 2010), examination on whether those methods can be applied to the state space model is needed.

Finally, since the convergence rates in the conditions with short times series lengths and low signal-to-noise ratios are strikingly low, results in these conditions are not as faithful as results in other conditions. A 50%-convergence-rate standard is used to judge whether or not results of this condition are trustworthy. And thus the condition $T50N1$ with low to medium signal-to-noise ratios and condition $T100N1$ with medium signal-to-noise ratios are excluded from further analysis. A complete report of results is included in the supplemental file and is available on request to the author.

TABLE 6.1

NUMBER OF CONVERGED SAMPLES WHEN THE STATIONARY INITIAL SETTING IS USED

	T50N1			T100N1			T200N1			T500N1		
	N	LN	CN	N	LN	CN	N	LN	CN	N	LN	CN
M11	719	651	694	903	838	836	920	925	919	939	940	944
M12	657	622	549	811	759	781	864	838	823	839	830	861
M13	569	530	558	741	729	688	783	799	771	795	789	785
M21	665	590	547	891	861	854	952	961	967	961	960	978
M22	672	615	516	895	824	812	923	895	911	927	952	928
M23	560	547	439	831	806	787	892	888	880	937	862	867
M31	188	105	126	556	380	398	897	750	759	998	889	948
M32	156	176	160	561	501	433	884	850	709	971	966	934
M33	179	174	162	550	525	443	890	854	738	981	918	918

	T10N50			T20N25			T25N20			T50N10		
	N	LN	CN	N	LN	CN	N	LN	CN	N	LN	CN
M11	931	941	955	940	940	959	945	951	944	936	942	938
M12	832	841	842	790	818	865	829	817	861	846	828	870
M13	757	788	826	789	788	832	778	800	813	798	785	802
M21	974	974	980	969	978	981	971	967	980	974	978	977
M22	930	962	958	939	947	956	935	951	952	937	929	931
M23	931	855	878	925	829	875	935	847	856	938	847	827
M31	988	890	943	994	931	957	998	935	934	999	940	963
M32	949	959	929	970	973	934	978	970	942	972	984	948
M33	967	931	910	973	936	907	972	927	925	972	940	926

Note: Symbols in the first column stand for models in table 5.1. M11, M12, and M13 belong to the high signal-to-noise ratio condition and M33, M32, and M33 belong to the low signal-to-noise ratio condition. The three models in the middle have medium signal-to-noise ratios. The length of the series is given by the number following the letter T in the first row and the number of participants is given by the number following the letter N in the first row. N, LN and CN in the second row stand for the normal, the log-normal and the contaminated normal distribution respectively.

TABLE 6.2

NUMBER OF CONVERGED SAMPLES WHEN THE NONINFORMATIVE INITIAL SETTING IS USED

	T50N1			T100N1			T200N1			T500N1		
	N	LN	CN	N	LN	CN	N	LN	CN	N	LN	CN
M11	803	766	789	970	927	941	998	985	987	1000	1000	1000
M12	763	748	723	954	942	880	991	988	971	994	999	992
M13	776	707	747	912	898	904	934	921	942	961	959	950
M21	700	656	607	938	912	905	998	986	996	1000	1000	1000
M22	711	677	610	957	934	917	997	989	997	1000	1000	1000
M23	678	714	602	946	941	919	994	989	988	999	998	997
M31	222	239	200	609	616	491	925	885	812	1000	966	982
M32	252	245	215	663	600	490	809	789	668	999	987	982
M33	260	253	197	640	611	492	943	921	793	999	999	985

	T10N50			T20N25			T25N20			T50N10		
	N	LN	CN	N	LN	CN	N	LN	CN	N	LN	CN
M11	923	928	934	934	931	934	938	869	935	940	932	947
M12	919	920	944	896	917	939	839	849	885	940	955	947
M13	862	850	884	870	869	885	898	855	879	894	906	919
M21	986	972	983	973	957	974	976	965	976	1000	997	1000
M22	951	947	952	954	940	956	916	888	912	952	928	947
M23	954	926	958	967	951	972	975	962	958	968	942	928
M31	1000	879	941	1000	935	962	1000	935	976	1000	915	976
M32	991	932	889	980	950	933	988	960	942	998	984	961
M33	984	862	826	973	915	892	980	924	919	983	934	935

Note: Symbols in the first column stand for models in table 5.1. M11, M12, and M13 belong to the high signal-to-noise ratio condition and M33, M32, and M33 belong to the low signal-to-noise ratio condition. The three models in the middle have medium signal-to-noise ratios. The length of the series is given by the number following the letter T in the first row and the number of participants is given by the number following the letter N in the first row. N, LN and CN in the second row stand for the normal, the log-normal and the contaminated normal distribution respectively.

6.2 Point Estimates

To evaluate the performance of point estimates, bias and total absolute bias under various conditions are summarized in the following tables. Results across various length conditions are included in table 6.3, 6.4,6.5, 6.6, 6.7, and 6.8. Results of Multi-subject conditions are included in table 6.11, 6.12, 6.9, and 6.10. Since convergence rates are low in conditions $T50N1$ and $T100N1$ with low signal-to-noise ratios and condition $T50N1$ with medium ratios, results of those conditions are not reported in the following four tables. Thus our attention centers on conditions with reasonable convergence rates.

From table 6.3, 6.4,6.5, 6.6, 6.7, and 6.8, we can see that as expected, the accuracy of point estimates generally improves as the times series length increases. Similarly, as the signal-to-noise ratio increases, the accuracy of point estimates improves. Moreover, the bias in the non-normal data conditions is generally larger than that in the normal data condition.

Concerning the comparisons between the two initial settings, we can see by comparing table 6.3, 6.4, 6.5 and table 6.6, 6.7, and 6.8 that the effect of initial settings depends on the time series length and the signal-to-noise ratio. As the time series length increases, the influence of initial settings on the accuracy of point estimates decreases. Similarly, as the signal-to-noise ratio increases, the effect of initial settings diminishes gradually. In general, as opposed to using the noninformative initial setting, using the stationary initial setting results in more accurate estimates of the transition matrix $\mathbf{\Phi}$. The differences between the two initial conditions for other parameters are not uniform. In some conditions, the stationary initial setting outperforms the noninformative initial setting whereas in other conditions, the noninformative initial setting gives more accurate parameter estimates.

With regard to the trade-off between increasing the time series length and the number of participants, the bias of point estimates, among the conditions that share the same total number of observations, is smaller in conditions with longer time series lengths but fewer participants (see table 6.11, 6.12, 6.9, and 6.10). For example, estimates obtained in condition $T500N1$ are more accurate than estimates in condition $T10N50$. The differences between conditions with longer time series lengths and fewer participants and conditions with shorter time series lengths and more participants depend on initial settings and signal-to-noise ratios. By comparing table 6.9 and table 6.10, we can see that when the stationary initial setting is used, the differences in bias among conditions that share the same total number of observations are within an acceptable range. In contrast, when the noninformative initial setting is used, bias is remarkably larger in conditions with shorter time series lengths than in conditions with longer time series lengths. This suggests that for designs with multiple participants and short time series lengths, it is better not to use the noninformative initial setting. Moreover, the signal-to-noise ratio also influences the differences among conditions that share the same total number of observations. By comparing table 6.9, 6.9 and table 6.11, 6.12, we can find that as the ratio increases, the differences gradually decrease. In short, even though the estimates in conditions with longer time series lengths are still more accurate, designs with shorter time series lengths but more participants have the advantage of lower experimental costs. Thus when the stationary initial setting is used, designs with shorter time series lengths are still useful alternatives that deserve considerations. This is especially true when the time series length is not extremely short, i.e., the condition $T50N10$ and when the signal-to-noise ratio is high, that is, when most variation in the observed variables is caused by the latent process rather than the variation in measurement errors.

The signal-to-noise ratio also influences the accuracy of point estimates. The bias of point estimates in the high ratio conditions is smaller than that in the low ratio conditions.

TABLE 6.3

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OBTAINED USING THE STATIONARY INITIAL SETTINGS: HIGH RATIO CONDITION (MODEL M12)

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	.20	.20	.20	.20	.20	.20	TAB
	T50N1	-4.4	0.3	-0.8	-3.3	0.3	-0.2	1.3	0.9	-1.3	0.9	-1.4	-0.6	0.0	-0.3	0.4	-0.4	-0.6	17.3
N	T100N1	-2.4	0.8	0.2	-1.8	0.0	0.1	0.6	0.3	-0.4	0.3	-0.6	-0.2	-0.1	0.2	-0.3	-0.2	-0.2	8.5
	T200N1	-0.9	0.1	0.3	-1.0	0.0	-0.1	0.0	0.0	-0.1	0.1	-0.5	-0.1	-0.1	-0.1	0.0	-0.1	0.0	3.5
	T500N1	-0.3	0.1	-0.1	-0.2	0.0	-0.1	0.0	0.1	0.0	0.1	-0.2	0.0	0.0	0.0	0.0	-0.2	0.0	1.4
	T50N1	-3.3	-0.3	-1.1	-3.2	0.0	0.5	1.1	1.3	-1.4	1.1	-1.3	-0.5	0.0	-0.2	0.6	-0.1	-0.4	16.5
LN	T100N1	-1.9	0.8	-0.3	-1.8	0.1	0.2	0.4	0.0	-0.4	0.6	-0.6	-0.5	0.2	-0.1	-0.3	-0.3	-0.1	8.5
	T200N1	-0.7	-0.3	0.3	-0.9	0.1	0.1	0.1	0.0	0.0	0.1	-0.2	-0.2	-0.2	0.0	-0.2	-0.1	0.2	3.7
	T500N1	-0.3	0.1	0.1	-0.5	-0.1	-0.1	0.0	0.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	1.8
	T50N1	-4.6	-1.0	-0.7	-3.6	0.3	1.7	2.0	0.3	2.5	0.3	1.2	0.4	1.8	0.5	1.2	0.3	1.0	23.3
CN	T100N1	-2.6	1.1	-0.1	-3.1	0.4	0.4	-0.2	0.3	1.5	0.3	0.9	0.2	0.0	0.0	0.0	-0.2	-0.5	11.7
	T200N1	-1.2	0.0	0.3	-1.4	0.1	0.0	0.2	0.2	1.0	0.2	0.3	0.0	0.1	-0.3	-0.1	0.0	0.3	5.7
	T500N1	-0.4	0.2	0.1	-0.5	0.0	0.2	0.1	0.2	0.3	0.0	0.2	0.1	0.1	-0.1	0.1	0.0	0.0	2.5

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$. The asterisks in the table indentify the conditions whose convergence rates are lower than the 50

TABLE 6.4

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OBTAINED USING THE STATIONARY INITIAL SETTINGS: MEDIUM RATIO CONDITION (M22)

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	1.00	1.00	1.00	1.00	1.00	1.00	TAB
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
N	T100N1	-2.0	1.5	-0.3	-2.9	1.2	1.6	1.0	1.6	-0.6	1.4	0.1	-1.3	-1.1	-0.8	-2.1	-1.5	-2.3	23.2
	T200N1	-1.3	0.0	0.8	-1.4	0.3	0.4	1.4	1.5	0.4	0.3	-0.6	-0.7	0.0	-1.1	-0.3	-0.9	-0.5	11.7
	T500N1	-0.6	0.4	0.2	-0.6	0.2	-0.1	0.3	0.0	0.4	0.1	0.2	-0.2	-0.3	-0.4	-0.6	-0.7	-0.5	5.7
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
LN	T100N1	-2.5	1.4	-0.1	-2.9	0.7	0.9	2.1	1.5	0.8	1.4	-0.1	-3.1	-3.0	-1.4	-1.1	-1.8	-2.7	27.2
	T200N1	-1.6	0.1	0.7	-1.4	0.5	1.6	0.9	0.6	-0.1	0.6	-0.2	-1.5	0.3	-0.5	-2.7	-2.4	-0.1	15.9
	T500N1	-0.6	0.1	0.1	-0.6	0.3	-0.1	0.5	0.2	0.1	0.2	0.2	0.0	-0.7	0.5	-0.6	-0.5	-0.6	5.8
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CN	T100N1	-3.9	2.2	1.3	-6.1	3.3	1.2	2.3	4.3	6.9	-0.2	7.8	-3.9	-2.7	-2.0	-5.5	-4.3	-5.9	63.7
	T200N1	-2.0	0.8	1.0	-2.5	0.1	1.0	1.2	0.6	3.3	-0.1	2.3	-1.6	-1.0	-3.2	-2.2	-1.2	-1.1	25.1
	T500N1	-0.9	0.4	0.4	-1.1	-0.1	0.1	0.7	0.6	1.2	-0.1	1.0	-0.3	-0.3	-1.4	-0.1	-0.3	-0.2	9.3

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$. The asterisks in the table indentify the conditions whose convergence rates are lower than the 50

TABLE 6.5

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OBTAINED USING THE STATIONARY INITIAL SETTINGS: LOW RATIO CONDITION (M32)

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	4.00	4.00	4.00	4.00	4.00	4.00	TAB
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
N	T100N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	T200N1	-1.9	3.9	1.4	-4.5	5.1	4.4	3.6	3.8	5.8	2.0	7.9	-8.2	-6.9	-6.2	-10.2	-7.7	-6.6	90.0
	T500N1	-1.0	0.8	0.9	-1.4	0.9	1.0	2.0	2.4	1.5	1.2	1.0	-1.5	-1.5	-2.9	-2.8	-4.2	-2.7	29.6
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
LN	T100N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	T200N1	-3.8	2.5	2.4	-2.8	3.8	4.8	3.7	2.4	6.6	1.8	7.0	-23.7	-7.7	-15.3	-30.6	-27.6	-12.3	158.8
	T500N1	-1.6	0.5	1.6	-1.1	0.9	-0.2	3.6	2.5	2.8	0.9	0.8	-4.1	-9.5	0.9	-4.1	-7.2	-6.1	48.2
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CN	T100N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	T200N1	-6.9	4.8	4.7	-7.3	6.7	7.8	8.5	7.0	27.4	-4.3	27.0	-29.2	-23.9	-30.6	-39.3	-21.6	-15.0	272.2
	T500N1	-2.4	2.8	1.7	-3.0	2.2	1.6	2.7	3.0	9.4	-0.5	8.3	-7.0	-4.8	-7.8	-3.9	-4.8	-4.7	70.4

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$. The asterisks in the table indentify the conditions whose convergence rates are lower than the 50

TABLE 6.6

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OBTAINED USING THE NONINFORMATIVE INITIAL SETTINGS: HIGH RATIO CONDITIONS (M12)

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	.20	.20	.20	.20	.20	.20	TAB
	T50N1	-8.2	3.3	1.8	-7.0	0.0	-0.3	0.9	0.5	-1.0	0.5	-0.7	-0.8	-0.2	-0.5	0.1	-0.5	-0.6	26.7
N	T100N1	-4.6	2.2	1.7	-4.1	-0.1	-0.1	0.5	0.1	-0.2	0.1	-0.4	-0.3	-0.3	0.0	-0.5	-0.3	-0.3	15.6
	T200N1	-2.1	0.8	1.2	-2.3	-0.1	-0.2	-0.1	-0.2	0.0	-0.1	-0.4	-0.3	-0.1	-0.1	-0.2	-0.2	-0.1	8.3
	T500N1	-0.8	0.3	0.2	-0.7	0.0	-0.1	0.0	0.0	0.0	0.0	-0.2	0.0	-0.1	-0.1	-0.1	-0.2	0.0	2.9
	T50N1	-8.4	3.5	2.4	-7.5	-0.6	-0.5	0.3	0.7	-0.1	0.7	-0.5	-0.8	-0.3	-0.5	0.0	-0.2	-0.4	27.1
LN	T100N1	-4.4	2.4	1.4	-4.0	0.0	-0.1	0.0	-0.2	-0.2	0.4	-0.5	-0.6	-0.1	-0.2	-0.6	-0.3	-0.3	15.7
	T200N1	-2.1	0.8	1.4	-2.2	-0.1	0.0	0.0	0.0	-0.2	0.0	-0.3	-0.3	-0.1	-0.2	-0.4	-0.2	0.0	8.3
	T500N1	-0.8	0.3	0.3	-0.8	-0.1	-0.1	0.0	0.1	0.1	0.0	-0.3	-0.1	-0.1	0.0	-0.1	-0.2	-0.1	3.3
	T50N1	-8.9	2.8	2.4	-9.0	-0.3	1.5	0.5	0.1	2.5	-0.3	2.5	-0.3	0.5	-0.1	-0.2	-0.3	0.5	32.7
CN	T100N1	-5.1	2.8	1.9	-5.4	0.2	0.2	-0.2	-0.1	1.7	0.1	1.2	0.1	-0.1	-0.1	-0.3	-0.4	-0.7	20.7
	T200N1	-2.4	0.9	1.2	-2.7	-0.1	-0.2	0.1	0.1	1.1	0.0	0.2	-0.3	0.1	-0.4	-0.4	0.0	0.0	10.2
	T500N1	-0.9	0.4	0.4	-1.0	0.0	0.1	0.1	0.1	0.3	0.0	0.2	-0.1	-0.1	-0.2	0.0	-0.1	-0.1	4.0

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$. The asterisks in the table indentify the conditions whose convergence rates are lower than the 50

TABLE 6.7

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OBTAINED USING THE NONINFORMATIVE INITIAL SETTINGS: MEDIUM RATIO CONDITION (M22)

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	1.00	1.00	1.00	1.00	1.00	1.00	TAB
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
N	T100N1	-5.6	4.1	2.5	-6.5	-0.2	0.6	0.0	0.4	1.1	0.7	1.8	-2.7	-1.4	-1.1	-3.4	-1.9	-2.4	36.3
	T200N1	-3.3	1.8	2.5	-3.5	-0.1	-0.1	1.0	0.8	1.3	-0.2	0.4	-1.2	-0.4	-1.1	-1.1	-1.1	-0.6	20.4
	T500N1	-1.5	1.1	0.9	-1.4	-0.1	-0.2	0.1	-0.2	0.8	-0.2	0.5	-0.4	-0.4	-0.5	-0.9	-0.8	-0.4	10.4
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
LN	T100N1	-5.9	3.8	2.8	-6.5	0.1	0.2	1.1	0.6	1.7	0.8	1.5	-4.5	-3.4	-1.7	-2.1	-1.5	-2.9	40.9
	T200N1	-3.7	1.9	2.6	-3.5	0.2	1.2	0.2	0.0	0.9	0.1	0.6	-1.8	0.0	-1.4	-3.3	-2.3	-0.6	24.1
	T500N1	-1.4	0.8	0.8	-1.4	0.2	-0.3	0.3	0.1	0.5	0.0	0.5	-0.2	-1.0	0.1	-1.0	-0.6	-0.9	10.2
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CN	T100N1	-7.6	5.6	3.8	-9.2	2.0	0.6	1.4	3.0	8.0	-0.3	8.3	-6.2	-4.0	-3.5	-7.6	-5.2	-7.1	83.2
	T200N1	-4.0	2.5	2.8	-4.7	-0.8	0.1	0.4	0.1	4.2	-0.7	3.3	-3.0	-1.6	-3.2	-3.6	-1.8	-1.8	38.6
	T500N1	-1.8	1.2	1.2	-1.9	-0.3	-0.2	0.4	0.2	1.7	-0.3	1.4	-0.8	-0.5	-1.5	-0.7	-0.5	-0.4	14.8

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$. The asterisks in the table indentify the conditions whose convergence rates are lower than the 50

TABLE 6.8

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OBTAINED USING THE NONINFORMATIVE INITIAL SETTINGS: LOW RATIO CONDITION (M32)

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	4.00	4.00	4.00	4.00	4.00	4.00	TAB
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
N	T100N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	T200N1	-7.3	7.0	4.9	-9.4	1.5	1.8	1.7	0.7	10.3	0.5	11.0	-11.3	-6.1	-7.8	-12.3	-7.8	-6.6	108.1
	T500N1	-3.5	3.0	3.0	-4.1	-0.1	0.0	0.9	1.4	3.5	0.1	3.1	-3.2	-2.2	-3.5	-4.5	-4.5	-3.2	43.8
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
LN	T100N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	T200N1	-7.5	4.6	5.2	-6.9	2.5	2.0	1.0	-0.5	7.6	1.2	9.2	-35.4	-18.9	-24.2	-45.6	-27.5	-23.3	223.0
	T500N1	-4.2	2.3	3.9	-3.4	0.3	-1.2	2.3	1.3	4.5	0.0	2.6	-5.7	-8.4	-2.0	-6.6	-7.3	-5.6	61.4
	T50N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CN	T100N1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	T200N1	-11.3	6.2	6.4	-12.2	0.9	3.0	4.6	3.8	32.0	-2.0	27.8	-31.9	-19.0	-30.8	-41.6	-23.2	-19.7	276.2
	T500N1	-5.2	4.4	3.8	-5.4	1.4	0.8	2.1	1.9	11.5	-1.4	10.0	-9.4	-6.0	-8.4	-5.9	-6.1	-5.1	88.5

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$. The asterisks in the table indentify the conditions whose convergence rates are lower than the 50

TABLE 6.9

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OF MODEL M32: STATIONARY INITIAL SETTING

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	1.00	1.00	1.00	1.00	1.00	1.00	TAB
	T500N1	-1.0	0.8	0.9	-1.4	0.9	1.0	2.0	2.4	1.5	1.2	1.0	-1.5	-1.5	-2.9	-2.8	-4.2	-2.7	29.6
	T10N50	-0.7	2.2	-1.8	-2.2	1.5	2.0	-1.7	2.3	1.5	1.6	4.1	1.2	-4.0	-10.1	-9.1	0.0	-1.0	46.8
N	T20N25	-4.5	-1.8	4.2	1.0	2.7	1.9	3.2	2.9	4.6	1.5	-2.0	-4.7	-10.0	-1.9	3.3	0.2	-2.0	52.2
	T25N20	-2.9	0.8	1.7	-0.4	1.4	1.2	1.6	3.4	4.0	-0.7	-0.4	-6.0	-5.6	1.0	-0.3	-1.4	-0.9	33.8
	T50N10	-1.0	2.1	0.4	-1.7	2.0	1.2	1.8	0.5	0.8	0.8	1.4	-4.8	0.1	0.0	-2.4	1.1	-0.3	22.4
	T500N1	-1.6	0.5	1.6	-1.1	0.9	-0.2	3.6	2.5	2.8	0.9	0.8	-4.1	-9.5	0.9	-4.1	-7.2	-6.1	48.2
	T10N50	-0.5	4.3	-3.2	-3.2	3.0	4.6	0.0	1.9	3.1	1.4	2.6	31.1	7.4	-40.5	-2.5	5.4	8.6	123.4
LN	T20N25	-4.9	-2.1	4.6	1.1	1.7	2.9	2.3	2.7	5.1	0.9	-1.6	15.0	-16.5	8.8	-14.4	-23.9	-21.9	130.3
	T25N20	-2.1	-0.1	1.3	0.0	1.6	0.8	1.9	2.5	2.7	0.1	0.4	5.5	-3.8	2.9	-21.5	-11.1	7.8	66.1
	T50N10	-1.4	0.5	0.7	-0.1	1.6	3.1	1.9	1.3	1.2	0.9	-0.1	3.6	-7.0	1.5	-15.9	-12.8	-2.3	56.1
	T500N1	-2.4	2.8	1.7	-3.0	2.2	1.6	2.7	3.0	9.4	-0.5	8.3	-7.0	-4.8	-7.8	-3.9	-4.8	-4.7	70.4
	T10N50	-5.0	3.4	1.0	-4.1	3.8	4.3	1.9	3.9	12.2	0.5	15.1	-8.8	11.7	0.9	-2.1	3.6	8.6	90.8
CN	T20N25	-3.9	1.3	4.0	-3.7	1.6	3.1	2.4	4.6	8.7	-0.7	10.7	-6.0	13.0	13.0	5.4	13.6	8.7	104.4
	T25N20	-4.3	2.6	3.2	-2.7	3.5	1.7	3.2	4.3	9.0	-1.5	8.7	-7.7	6.6	14.9	-5.1	5.5	-4.7	89.1
	T50N10	-1.2	2.6	1.4	-3.6	1.8	4.1	4.4	3.9	4.3	0.6	9.3	-1.7	5.0	-2.2	-5.3	-2.9	0.5	54.4

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$.

TABLE 6.10

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OF MODEL M32: NONINFORMATIVE INITIAL SETTING

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	1.00	1.00	1.00	1.00	1.00	1.00	TAB
	T500N1	-3.5	3.0	3.0	-4.1	-0.1	0.0	0.9	1.4	3.5	0.1	3.1	-3.2	-2.2	-3.5	-4.5	-4.5	-3.2	43.8
	T10N50	-42.5	-16.2	-16.9	-41.6	-17.5	-16.8	-20.4	-15.4	93.2	39.8	95.4	-57.1	1.8	-5.0	-67.2	7.7	4.0	558.5
N	T20N25	-22.5	2.2	3.0	-20.1	-9.8	-10.2	-11.9	-12.4	34.3	9.1	29.2	-24.5	-8.8	-0.8	-20.0	1.8	-0.5	220.9
	T25N20	-18.6	5.3	5.8	-16.2	-10.2	-10.2	-9.5	-8.5	26.4	2.1	20.6	-22.6	-4.2	2.1	-15.8	-0.9	0.2	179.2
	T50N10	-12.8	8.8	7.8	-12.9	-5.1	-5.7	-4.5	-6.0	11.8	-1.1	11.4	-13.7	0.4	0.1	-10.0	1.3	-0.1	113.4
	T500N1	-4.2	2.3	3.9	-3.4	0.3	-1.2	2.3	1.3	4.5	0.0	2.6	-5.7	-8.4	-2.0	-6.6	-7.3	-5.6	61.4
	T10N50	-41.6	-16.1	-16.4	-42.5	0.1	-3.8	11.1	2.2	91.6	32.2	78.2	-28.8	-9.3	-38.6	-42.5	-14.1	-0.8	469.9
LN	T20N25	-24.4	1.4	3.1	-21.7	-8.9	0.4	-12.0	-9.9	35.7	8.6	35.0	-7.4	-13.8	1.4	-44.5	-18.5	-19.0	265.5
	T25N20	-19.6	4.5	5.7	-17.3	-8.6	-7.7	-6.4	-7.8	26.8	3.4	21.7	-11.9	-0.7	4.7	-36.7	-8.9	6.8	199.1
	T50N10	-13.2	7.5	8.1	-12.0	-3.5	-3.1	-0.3	-4.3	11.4	-0.5	9.3	-5.3	-7.5	1.1	-22.6	-18.3	-0.1	128.1
	T500N1	-5.2	4.4	3.8	-5.4	1.4	0.8	2.1	1.9	11.5	-1.4	10.0	-9.4	-6.0	-8.4	-5.9	-6.1	-5.1	88.5
	T10N50	-44.8	-17.4	-18.7	-43.5	-6.8	-1.0	8.7	10.2	125.7	35.0	97.6	-94.0	14.6	-2.8	-54.1	-5.6	-1.7	581.9
CN	T20N25	-26.2	-2.3	-2.1	-25.9	-10.6	-10.0	-5.1	-4.1	61.6	12.7	54.5	-46.1	13.6	13.6	-29.4	9.0	5.7	332.6
	T25N20	-20.7	3.2	3.0	-19.4	-8.3	-9.7	-7.3	-5.2	37.7	3.7	37.4	-30.1	8.2	15.3	-28.6	5.8	-4.2	247.9
	T50N10	-12.6	7.7	6.6	-14.9	-4.1	-4.2	-1.7	-3.1	14.5	-0.3	19.9	-11.5	2.7	-2.1	-14.7	-3.4	0.0	123.9

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$.

TABLE 6.11

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OF MODEL M12: STATIONARY INITIAL SETTING

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	1.00	1.00	1.00	1.00	1.00	1.00	TAB
	T500N1	-0.3	0.1	-0.1	-0.2	0.0	-0.1	0.0	0.1	0.0	0.1	-0.2	0.0	0.0	0.0	0.0	-0.2	0.0	1.4
	T10N50	-1.0	0.8	0.0	-0.8	-0.3	-0.1	-0.1	-0.1	0.6	-0.1	0.0	0.0	0.0	-0.5	-0.6	0.2	0.1	5.5
N	T20N25	-0.9	-0.3	0.4	-0.5	0.0	0.0	0.1	0.2	0.0	0.7	-0.6	-0.1	-0.5	0.1	0.2	0.1	-0.1	4.7
	T25N20	-0.6	-0.4	0.0	-0.1	-0.1	-0.1	0.2	0.3	0.0	0.2	-0.6	-0.2	-0.3	0.1	0.2	-0.1	0.0	3.6
	T50N10	-0.5	0.4	-0.5	-0.3	0.1	0.2	0.2	0.1	-0.5	0.2	0.2	-0.2	0.1	0.1	0.2	0.1	0.0	3.5
	T500N1	-0.3	0.1	0.1	-0.5	-0.1	-0.1	0.0	0.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	1.8
	T10N50	-1.1	1.2	-0.2	-1.5	0.1	0.2	-0.2	-0.1	2.2	-0.1	0.1	0.5	0.3	-0.9	-0.4	0.4	0.0	9.5
LN	T20N25	-0.7	-1.0	0.2	-0.3	0.0	0.2	0.2	0.2	0.8	0.7	-0.9	0.1	-0.7	0.2	0.2	0.0	-0.4	6.5
	T25N20	-0.6	-0.4	-0.2	-0.1	-0.2	-0.1	0.0	0.2	0.5	0.1	-0.4	-0.1	-0.1	0.2	0.1	-0.2	0.1	3.5
	T50N10	-0.3	0.2	-0.6	-0.2	-0.2	0.2	0.1	-0.1	-0.1	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.1	2.4
	T500N1	-0.4	0.2	0.1	-0.5	0.0	0.2	0.1	0.2	0.3	0.0	0.2	0.1	0.1	-0.1	0.1	0.0	0.0	2.5
	T10N50	-1.3	0.0	-0.3	-0.9	-0.2	0.0	0.1	0.2	0.5	-0.1	2.5	0.1	1.2	0.6	0.4	0.6	0.8	9.9
CN	T20N25	-0.6	-0.6	0.3	-0.7	-0.4	-0.2	-0.1	0.1	0.5	-0.5	2.1	0.1	1.3	1.1	0.8	1.2	0.7	11.1
	T25N20	-0.7	-0.4	-0.2	-0.5	0.1	-0.2	0.3	0.5	0.6	-0.1	1.2	0.0	0.6	1.1	0.7	0.6	-0.1	7.8
	T50N10	-0.4	0.0	-0.3	-0.3	0.2	0.2	0.4	0.0	0.0	-0.5	0.7	0.6	0.6	0.1	0.4	0.1	0.4	5.1

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$.

TABLE 6.12

BIAS×100 OF NORMAL THEORY BASED ML ESTIMATES OF MODEL M12: NONINFORMATIVE INITIAL SETTING

		$\mathbf{\Phi}_{11}$	$\mathbf{\Phi}_{12}$	$\mathbf{\Phi}_{21}$	$\mathbf{\Phi}_{22}$	$\mathbf{A}_{21}$	$\mathbf{A}_{31}$	$\mathbf{A}_{52}$	$\mathbf{A}_{62}$	$\mathbf{Q}_{11}$	$\mathbf{Q}_{12}$	$\mathbf{Q}_{22}$	$\mathbf{R}_{11}$	$\mathbf{R}_{22}$	$\mathbf{R}_{33}$	$\mathbf{R}_{44}$	$\mathbf{R}_{55}$	$\mathbf{R}_{66}$	
True Values		.60	.30	.30	.60	.80	.80	.80	.80	.41	-.14	.41	1.00	1.00	1.00	1.00	1.00	1.00	TAB
	T500N1	-0.8	0.3	0.2	-0.7	0.0	-0.1	0.0	0.0	0.0	0.0	-0.2	0.0	-0.1	-0.1	-0.1	-0.2	0.0	2.9
	T10N50	-14.6	10.4	9.7	-14.1	-1.4	-1.1	-1.2	-1.2	3.0	-1.0	2.1	-0.4	0.0	-0.4	-1.0	0.2	0.1	62.1
N	T20N25	-12.1	9.0	9.7	-11.7	-0.5	-0.6	-0.6	-0.5	1.6	-0.3	0.9	-0.3	-0.5	0.1	-0.1	0.2	0.0	48.6
	T25N20	-10.5	8.0	8.3	-9.9	-0.7	-0.7	-0.3	-0.4	1.4	-0.8	0.8	-0.5	-0.2	0.2	-0.1	-0.1	0.0	42.5
	T50N10	-5.2	4.4	3.4	-5.0	-0.2	-0.2	-0.2	-0.3	0.2	-0.4	0.8	-0.3	0.1	0.1	0.0	0.0	0.0	20.8
	T500N1	-0.8	0.3	0.3	-0.8	-0.1	-0.1	0.0	0.1	0.1	0.0	-0.3	-0.1	-0.1	0.0	-0.1	-0.2	-0.1	3.3
	T10N50	-14.5	10.6	9.9	-14.8	-1.0	-0.9	-1.2	-1.1	4.3	-0.9	2.2	0.1	0.3	-0.9	-0.9	0.4	0.1	64.0
LN	T20N25	-12.1	8.7	10.1	-11.9	-0.7	-0.5	-0.5	-0.4	2.6	-0.5	0.5	-0.2	-0.6	0.2	-0.1	-0.1	-0.3	49.9
	T25N20	-10.8	8.3	8.6	-10.1	-0.7	-0.7	-0.6	-0.4	1.8	-0.9	0.9	-0.4	-0.1	0.2	-0.3	-0.2	0.2	45.0
	T50N10	-5.4	4.5	3.6	-5.1	-0.5	-0.2	-0.2	-0.4	0.5	-0.4	0.6	-0.2	0.0	0.0	0.0	-0.1	0.1	21.6
	T500N1	-0.9	0.4	0.4	-1.0	0.0	0.1	0.1	0.1	0.3	0.0	0.2	-0.1	-0.1	-0.2	0.0	-0.1	-0.1	4.0
	T10N50	-14.9	10.3	9.6	-14.3	-1.3	-1.1	-1.1	-0.9	3.0	-0.9	4.7	-0.4	1.3	0.6	-0.1	0.6	0.8	65.9
CN	T20N25	-11.5	9.2	9.7	-12.1	-1.1	-0.9	-0.7	-0.6	2.1	-1.5	3.7	-0.3	1.3	1.1	0.5	1.2	0.6	58.2
	T25N20	-10.7	8.5	8.5	-10.4	-0.5	-0.8	-0.4	-0.1	1.9	-1.1	2.5	-0.1	0.6	1.1	0.4	0.6	0.0	48.1
	T50N10	-5.3	4.3	3.9	-5.3	-0.2	-0.2	0.0	-0.3	0.8	-1.0	1.4	0.3	0.4	0.0	0.1	0.0	0.2	23.8

Note: The entries in the table equal bias×100. TAB stands for total absolute bias, that is, $TAB = \sum_{i=1}^{17} |bias_i|$.

6.3 Standard Error Estimates

In this section, simulation results of standard error estimators are summarized according to the influences of related factors. These factors include the distribution of data, the time series length, the initial settings, the trade-off between increasing the time series length and increasing the number of participants and the signal-to-noise ratio.

6.3.1 Data Distribution

To illustrate the effect of the data distribution on the performance of standard error estimators, figure 6.1 which compares results of different distribution conditions is included. Figure 6.1 graphs the results of SE estimators obtained when fitting model M22 using the stationary initial setting under different distribution conditions. The relative bias (the ratio of average bias to the empirical SE estimates) of SE estimators is visualized in the first row of figure 6.1 and mis-coverage rates using different SE estimators visualized in the second row. Figure 6.1 shows that the relative bias of SE estimators in the normal data condition is generally smaller compared to the two non-normal conditions and the mis-coverage rates in the normal data condition are closer to the nominal level.

With regard to the comparisons among SE estimators, we can see that differences in the performance of SE estimators vary across distribution conditions. In the normal data condition, the relative bias of sandwich-type SE estimators is comparable to that of information based SE estimators whereas in the non-normal data conditions, sandwich-type SE estimators yield smaller relative bias compared with information based SE estimators. More specifically, when data do not follow the normal distribution, information based SE estimators tend to underestimate the empirical standard error estimates of parameters. When evaluating the mis-coverage rates, results are slightly different. Only Papanastasious' sandwich SE estimator, $\widehat{SE}_{SW_{P1}}$, outperforms information based SE estimators in the non-normal data conditions. The mis-coverage rates using White's sandwich SE estimator, $\widehat{SE}_{SW_W}$, are more inflated than using information based SE estimators in the normal and log-normal conditions. In short, $\widehat{SE}_{SW_{p1}}$ is more preferable for non-normal data compared with other SE estimators.

The effect of data distributions interacts with the effect of the time series length. Generally, the differences across distribution conditions decrease as the time series length increases. Detailed discussions on the interaction effects are included in subsection 6.3.2.

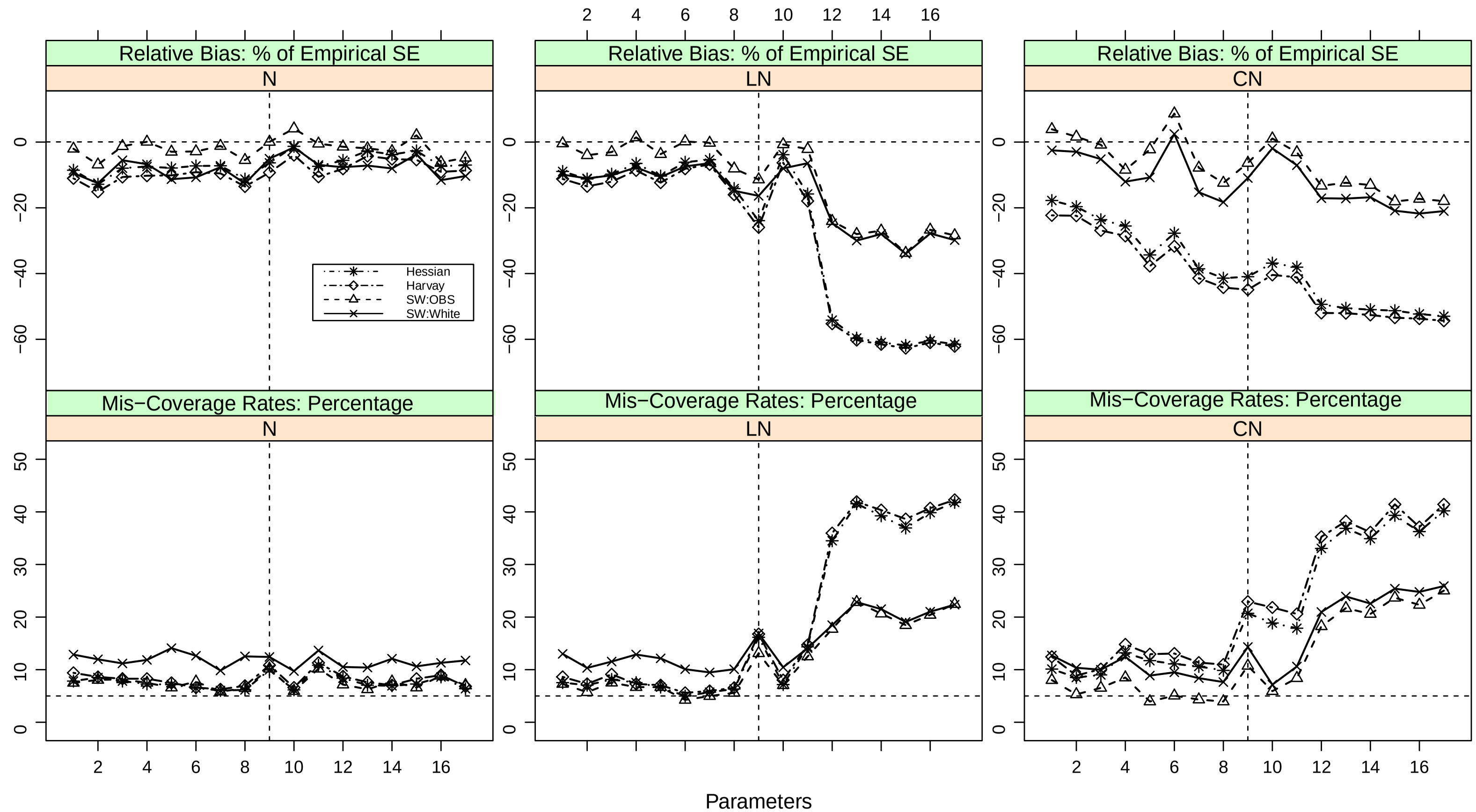

Figure 6.1: Relative bias of SE estimators and mis-coverage rates obtained when fitting model M22 under different distribution conditions. $T = 100$. The stationary initial setting is used. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

6.3.2 Time Series Length

Figure 6.2 and figure 6.3 are included to illustrate the effect of the time series length on the performance of SE estimators. Since for information based SE estimators, the effect of time series length varies across distribution conditions, figure 6.2 and figure 6.3 compare results of SE estimators across different length and distribution conditions. The relative bias of SE estimators obtained when fitting model M22 using the stationary initial setting is visualized in figure 6.2 and the corresponding mis-coverage rates are visualized in figure 6.3.

From figure 6.2 and figure 6.3, we can see that the effect of the time series length differs across different SE estimators. For sandwich-type SE estimators, the relative bias decreases and the corresponding mis-coverage rates approach the nominal level as the time series length increases. For information based SE estimators, however, the effect of time series length displays different patterns across distribution conditions. In the normal data condition, the accuracy of SE estimates and the control on mis-coverage rates improve with the time series length. In the non-normal data conditions, however, as the time series length increases, no substantial improvements are observed for covariance component parameters (the variance-covariances of state noise $\mathbf{Q}$ and the variance-covariances of measurement errors $\mathbf{R}$). But for non-covariance component parameters (the measurement matrix $\mathbf{A}$ and the transition matrix $\mathbf{\Phi}$), small improvements in the accuracy of information based SE estimates and the control on corresponding mis-coverage rates are observed as the time series length increases.

The effect of the time series length interacts with the effect of the signal-to-noise ratio. Figure 6.4 and figure 6.5 evaluate the effect of the time series length under different ratio conditions. We can see from these two figures that among conditions that have the same time series length, the relative bias of SE estimators is larger in conditions with lower ratios. This indicates that for data with low signal-to-noise ratios, a longer time series length is needed to obtain desired control on mis-coverage rates and desired accuracy of SE estimates.

Concerning the comparisons among SE estimators, Papanastassiou's sandwich SE estimator, $\widehat{SE}_{SW_{p1}}$, shows advantages in estimating the empirical standard error estimates for non-normal data in terms of smaller relative bias and better controlled mis-coverage rates. Even though the relative bias of information based SE estimators for $\mathbf{A}$ and $\mathbf{\Phi}$ is negligible in conditions with sufficiently long series lengths and adequately high signal-to-noise ratios, it is not otherwise. The same pattern applies to the mis-coverage rates based on information based SE estimators. When the signal-to-noise ratio is low, the mis-coverage rates of $\mathbf{\Phi}$ and $\mathbf{A}$ using information based SE estimator are inflated. Thus information based SE estimators are less preferable to Pananastassiou's sandwich SE estimator. White's sandwich SE estimator, $\widehat{SE}_{SW_W}$, is not recommended because when the time series length is short and when the signal-to-noise ratio is high, the mis-coverage rates using $\widehat{SE}_{SW_W}$ are more inflated compared with using other SE estimators (see figure 6.3). This is especially apparent in the normal data and the log-normal data conditions.

In short, the simulation study suggests that for normally distributed data, all SE estimators are suitable for the standard errors of parameters. For non-normal data, however, while sandwich-type SE estimators are still appropriate for all parameters, information based SE estimators are only suitable for non-covariance component parameters and only when the time series length is sufficiently long and the signal-to-noise ratio is adequately high.

Results in most other conditions agree with the findings discussed above (Please contact the author for specific results of other conditions). However, there are some exceptions. In the low ratio and the log-normal data conditions, the relative bias of SE estimators may increase as the time series length increases. This reveals some inconsistency in the estimation. Nevertheless, even in these conditions, mis-coverage rates are still better controlled in long length conditions compared with short length conditions.

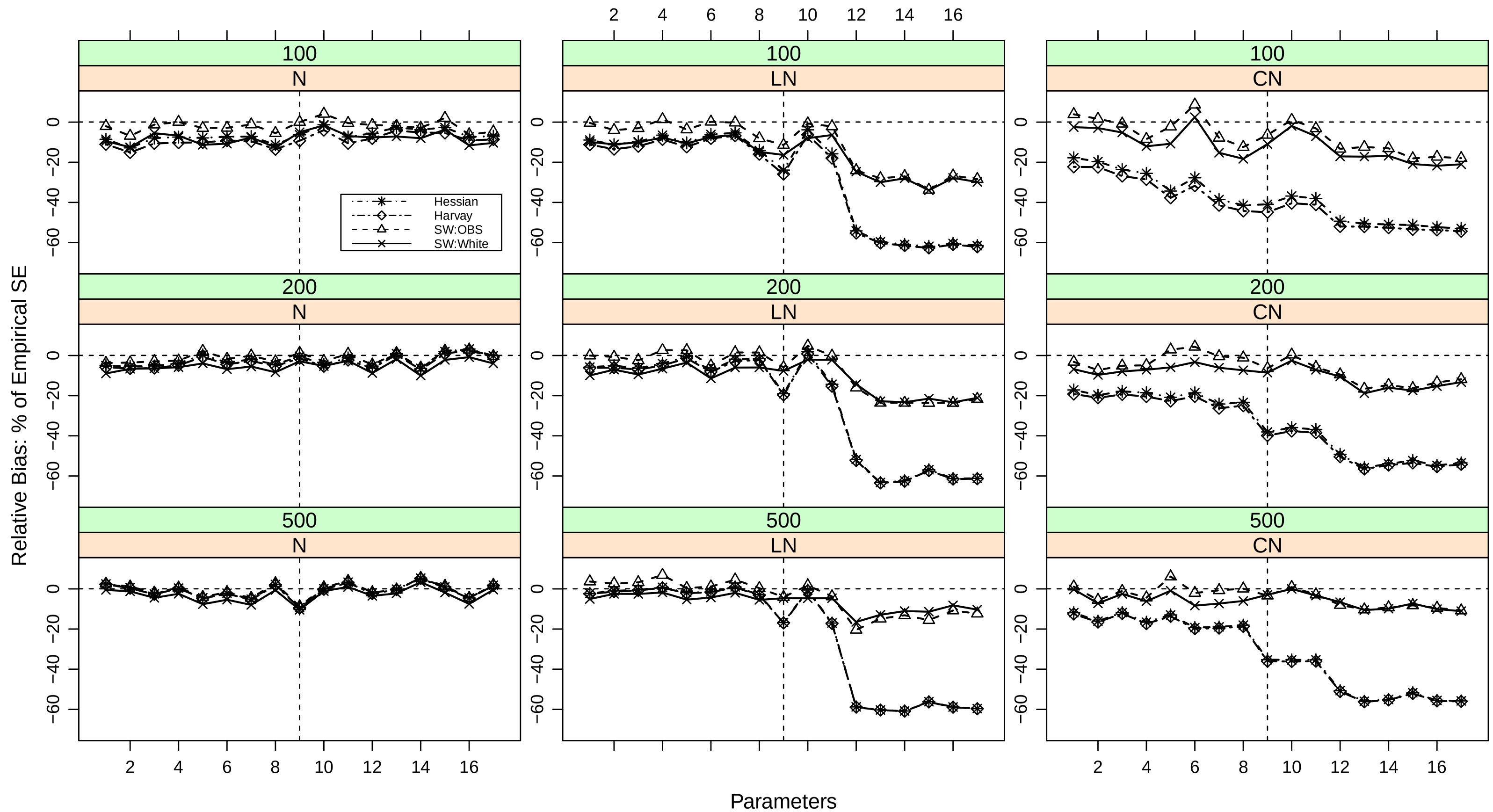

Figure 6.2: Relative bias of SE estimators obtained when fitting model M22 under different length and distribution conditions. The stationary initial setting is used. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator proposed (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator proposed (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

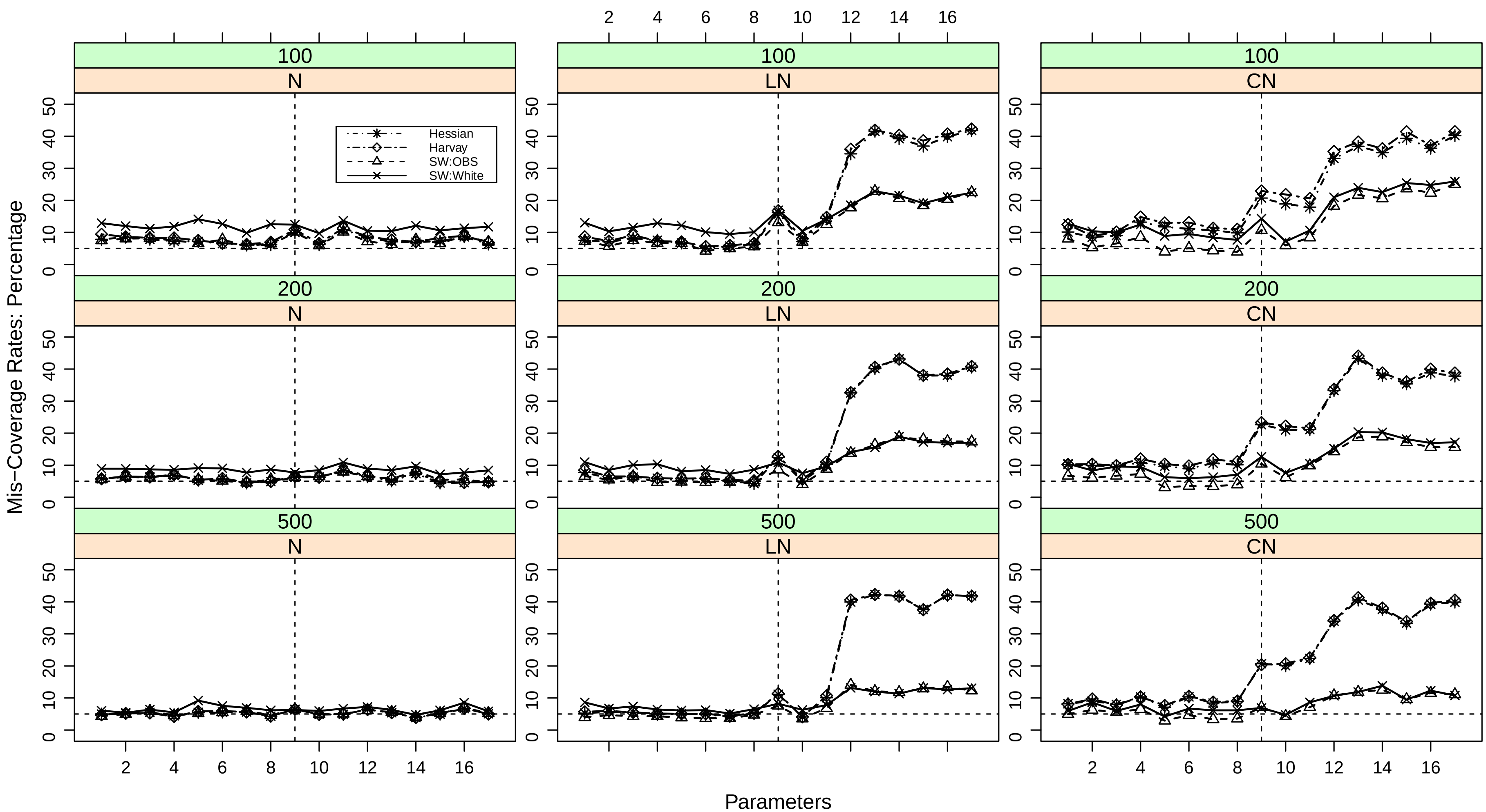

Figure 6.3: Mis-coverage rates obtained when fitting model M22 under different length and distribution conditions. The stationary initial settings is used. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

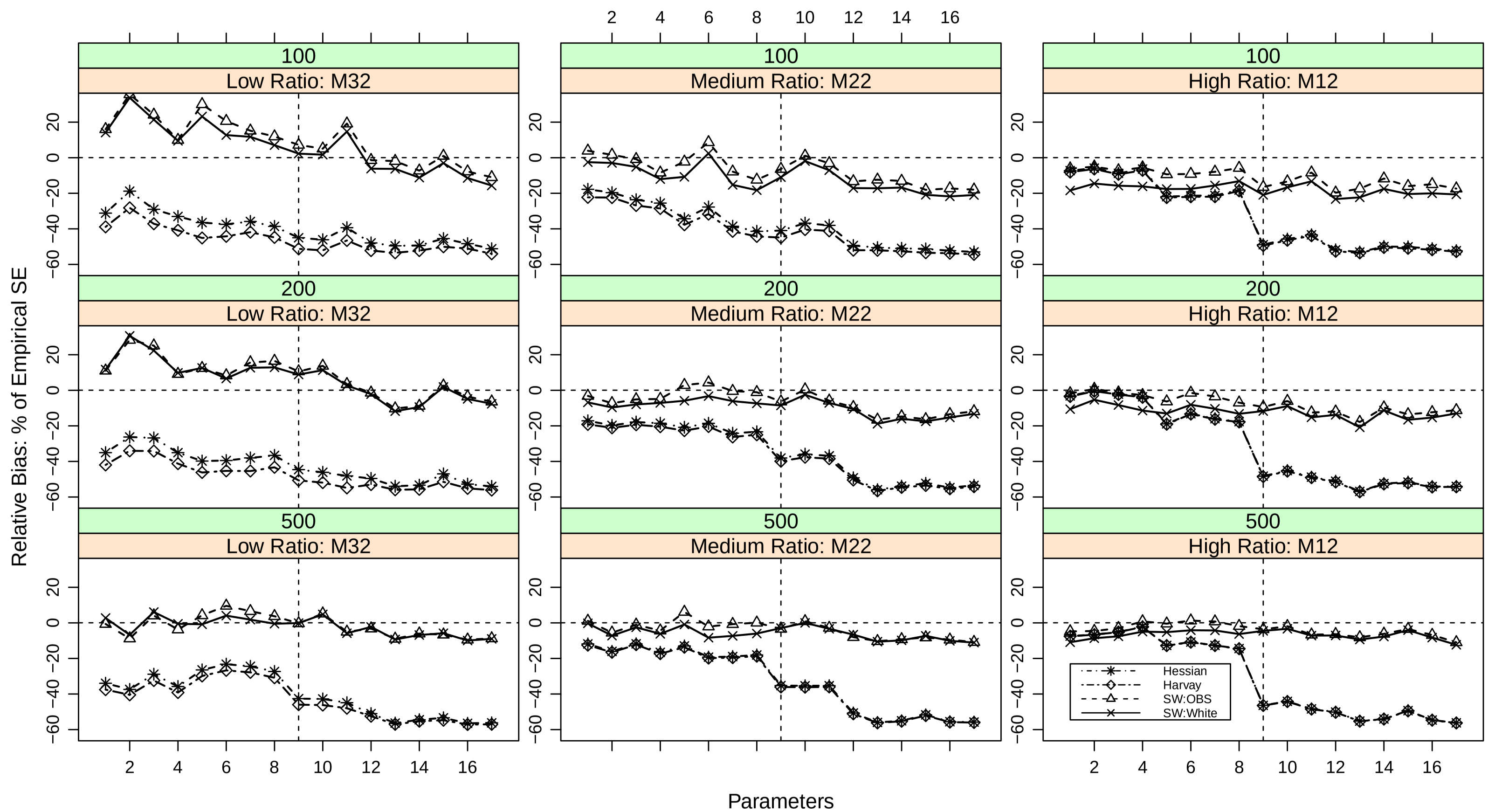

Figure 6.4: Relative bias of SE estimators obtained using the stationary initial settings under different length and ratio conditions. Measurement errors and state noise variables follow the multivariate contaminated normal distribution respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

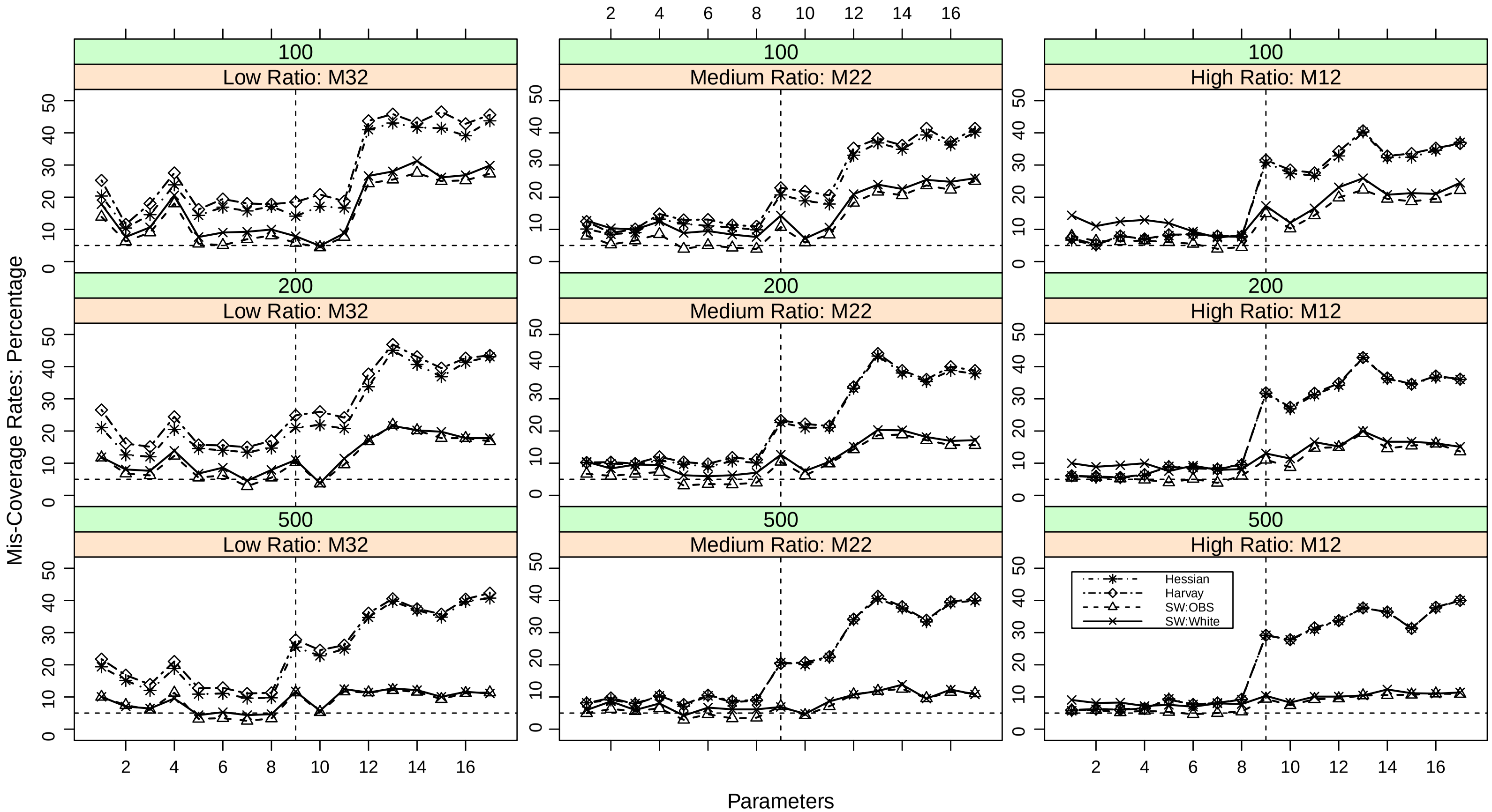

Figure 6.5: Mis-coverage rates obtained using the stationary initial settings under different length and ratio conditions. Measurement errors and state noise variables follow the multivariate contaminated normal distribution respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

6.3.3 Initial Settings

Figure 6.6 and figure 6.7 are included to illustrate the effect of initial settings. To visualize related interaction effects, figure 6.6 and figure 6.7 compare the relative bias and mis-coverage rates obtained when fitting model M12 across different initial setting, length and number-of-participant conditions. Results in the first row of these two figures are obtained using the stationary initial setting and results in the second row using the noninformative initial setting. Columns in these two figures differ in the time series length and the number of participants.

We can see from the following figures that the effect of initial setting is different across different SE estimators. Harvey's SE estimator is most sensitive to initial settings. The relative bias of Harvey's SE estimator for $\mathbf{\Phi}$ and $\mathbf{A}$ is smaller and the corresponding mis-coverage rates are better controlled when the stationary initial setting is used compared with when the noninformative initial setting is used. The other three SE estimators display smaller differences between the two initial conditions compared with Harvey's SE estimator, indicating that these SE estimators are less influenced by initial settings.

Generally, the effect of initial settings is attenuated by the time series length. With a time series length as long as $T = 500$, the differences between the two initial conditions are negligible. By comparing the first column and the third column of the following figures, we can see that the effect of adding participants when the time series length is short differs across the two initial conditions. If the noninformative initial setting is used, adding participants when the time series length is short does not improve the accuracy of SE estimators and the control of mis-coverage rates. In contrast, if the stationary initial setting is used, the accuracy of SE estimators and the control of mis-coverage rates are substantially improved in the conditions with more participants. In other words, the differences between the two initial conditions are larger when data involve multiple participants.

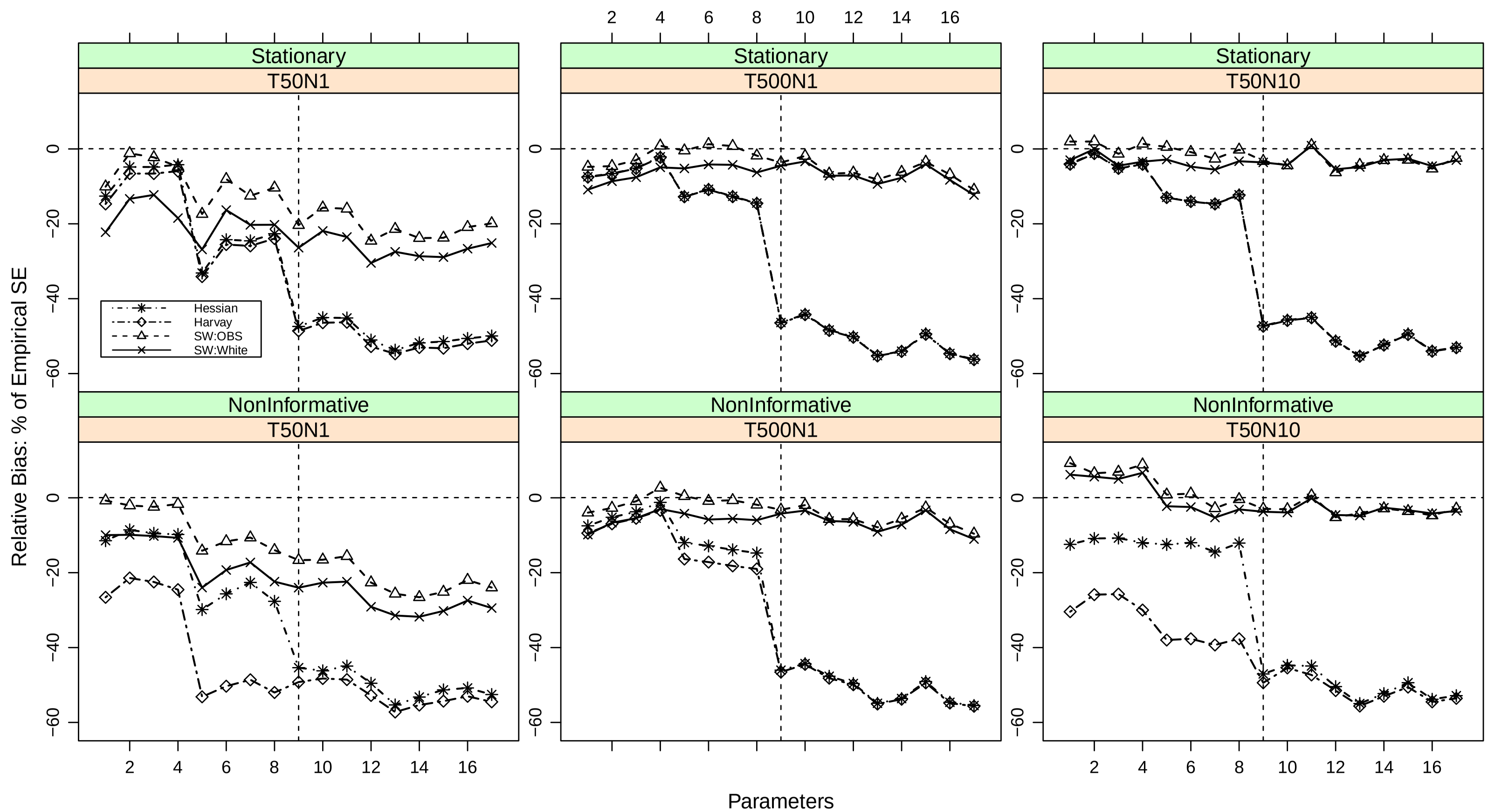

Figure 6.6: Relative bias of SE estimators obtained when fitting model M12 under different initial setting, length and size conditions. Measurement errors and state noise variables in these conditions follow the multivariate contaminated normal distribution respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

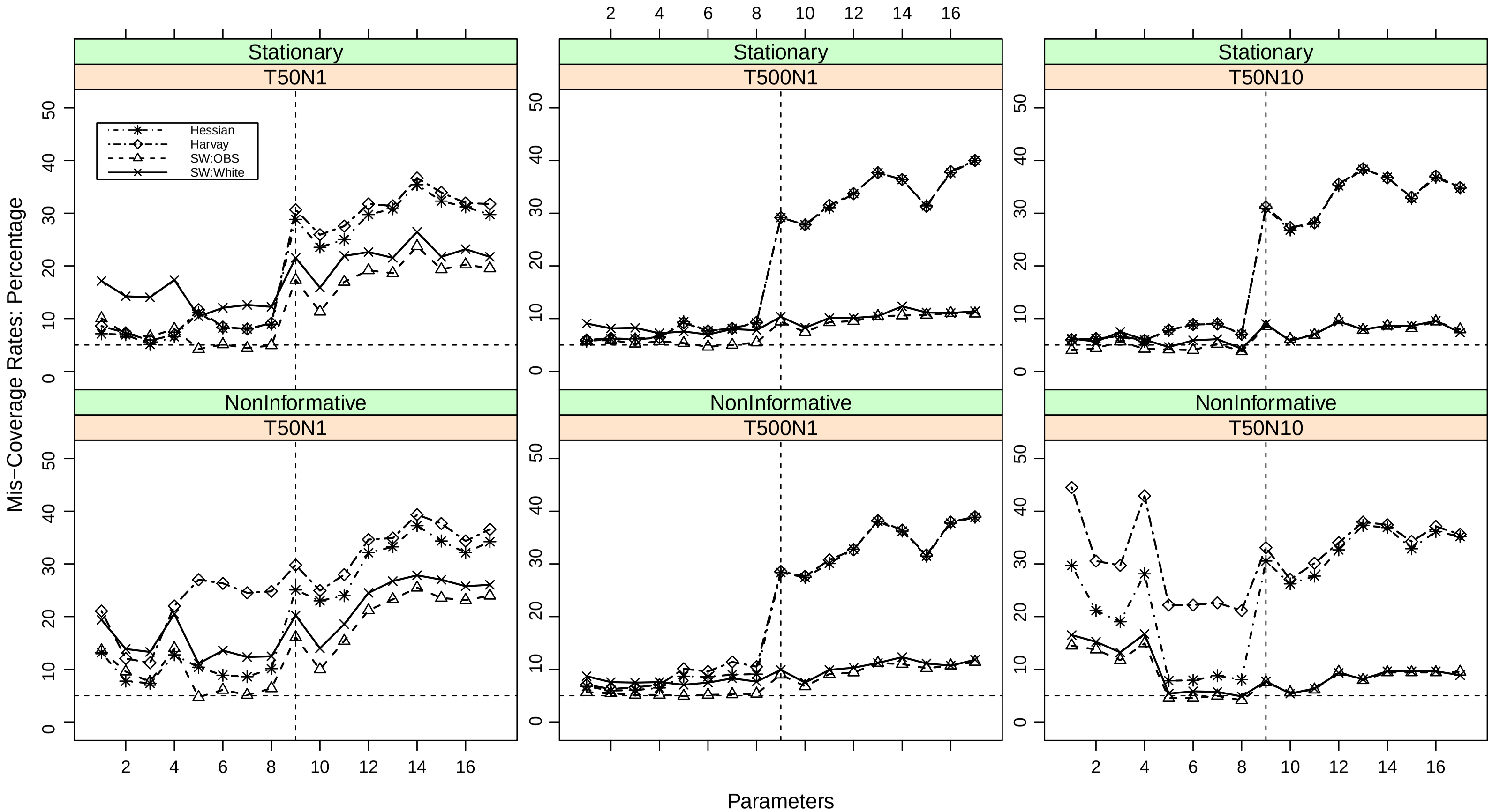

Figure 6.7: Mis-coverage rates obtained when fitting model M12 under different initial setting, length and number of participants conditions. Measurement errors and state noise variables in these conditions follow the multivariate contaminated normal distribution respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

6.3.4 Trade-Off Between Increasing T and N

Either increasing the time series length or increasing the number of participants improves the accuracy of estimation. However, increasing the time series length is more expensive. Thus between among designs that share the same total number of observations, designs with shorter time series lengths are less expensive. And it would be encouraging if the results of designs with shorter time series lengths are comparable to the results of designs with longer time series lengths.

To illustrate the differences among conditions that share the same total number of observations, figure 6.8 and figure 6.9 are included. Given the limits of space, only the results of conditions $T10N50$, $T50N10$ and $T500N1$ are included. Since the simulation study shows that the differences among conditions with the same total number of observations are influenced by initial settings. Figure 6.8 and figure 6.9 graph the results of conditions that share the same total number of observations across different initial conditions.

From figure 6.8 and figure 6.9 we can see that when the stationary initial setting is used, the differences among conditions that share the same total number of observations are negligible whereas when the noninformative initial setting is used, results are substantially better in conditions with longer time series lengths. The SE estimates obtained using the noninformative initial setting in conditions with short time series lengths, i.e., $T10N50$, are unacceptably biased and the mis-coverage rates in these conditions are severely inflated.

Concerning the comparisons among SE estimators, sandwich-type SE estimators generally outperform information based SE estimators. However, the advantage of using sandwich-type SE estimators decreases as the signal-to-noise ratio increases. Detailed discussions on the effect of the signal-to-noise ratio are included in the subsection 6.3.5. However, given the much shorter time series lengths of multi-subject conditions considered in this subsection, figure 6.10 and figure 6.11 are still included to illustrate the effect of the signal-to-noise ratio. By examining figure 6.10 and figure 6.11, we can see that the differences in the relative bias of $\mathbf{\Phi}$ and $\mathbf{A}$ among SE estimators are substantially smaller in the high ratio conditions compared with the low ratio conditions. And the mis-coverage rates using different SE estimators are comparable to each other.

In short, the simulation study suggests that when the stationary initial setting is used, conditions with shorter time series lengths but more participants yield similar results as conditions with longer time series lengths but fewer participants. In addition, sandwich-type SE estimators generally outperform information based SE estimators even though information based SE estimators work adequately well when the signal-to-noise ratio is high.

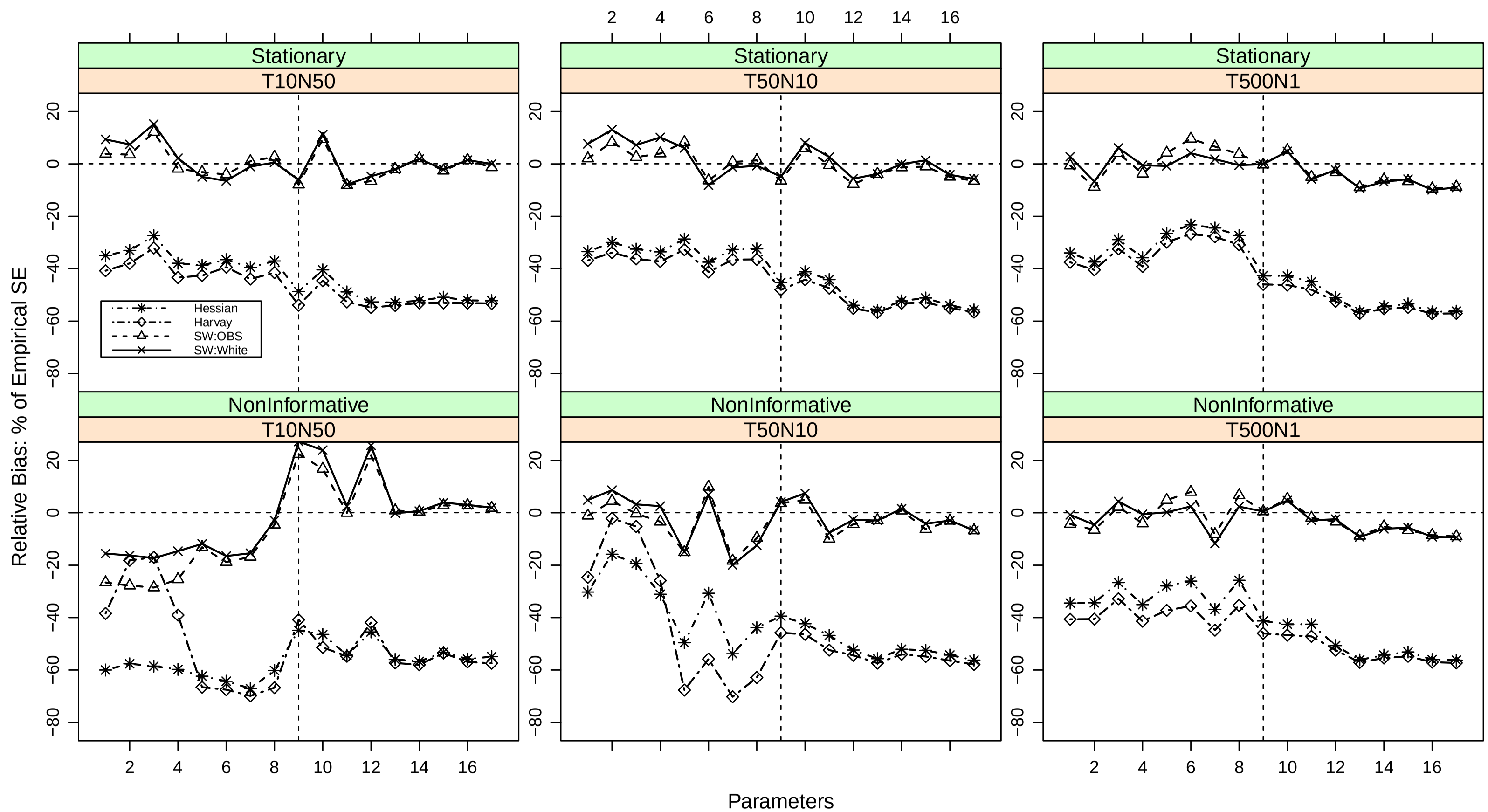

Figure 6.8: Relative bias of SE estimates obtained when fitting model M32 under different length and number-of-participants conditions. Measurement errors and state noise variables in these conditions follow the multivariate contaminated normal distribution respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: $1\mathbf{\Phi}_{11}$, $2\mathbf{\Phi}_{12}$,$3\mathbf{\Phi}_{21}$,$4\mathbf{\Phi}_{22}$, $5\mathbf{A}_{21}$, $6\mathbf{A}_{31}$, $7\mathbf{A}_{52}$, $8\mathbf{A}_{62}$, $9\mathbf{Q}_{11}$, $10\mathbf{Q}_{21}$, $11\mathbf{Q}_{22}$, $12\mathbf{R}_{11}$, $13\mathbf{R}_{22}$, $14\mathbf{R}_{33}$, $15\mathbf{R}_{44}$, $16\mathbf{R}_{55}$, $17\mathbf{R}_{66}$.

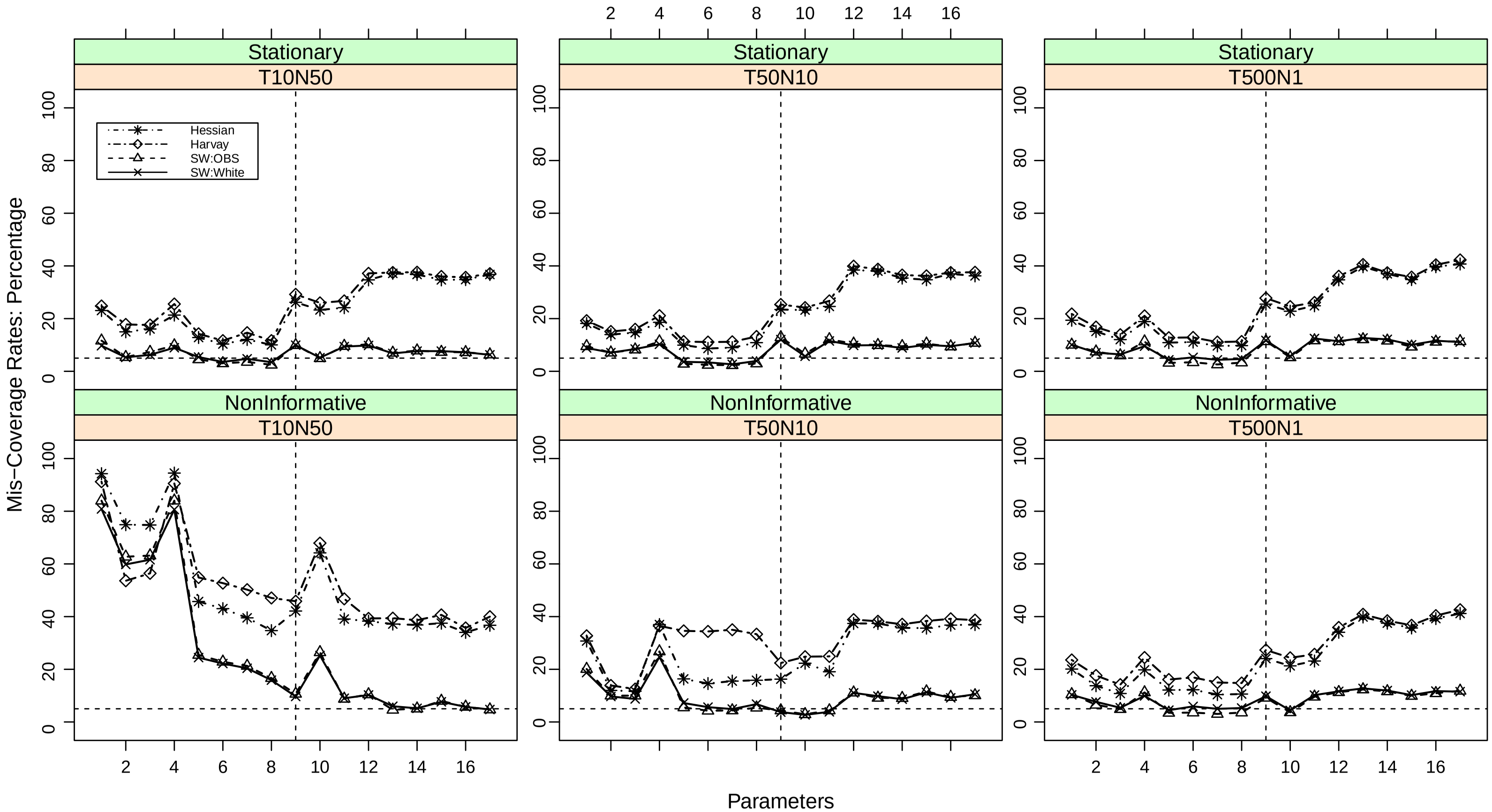

Figure 6.9: Mis-coverage rates under obtained when fitting model M32 different length and number-of-participant conditions. Measurement errors and state noise variables in these conditions follow the multivariate contaminated normal distribution respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

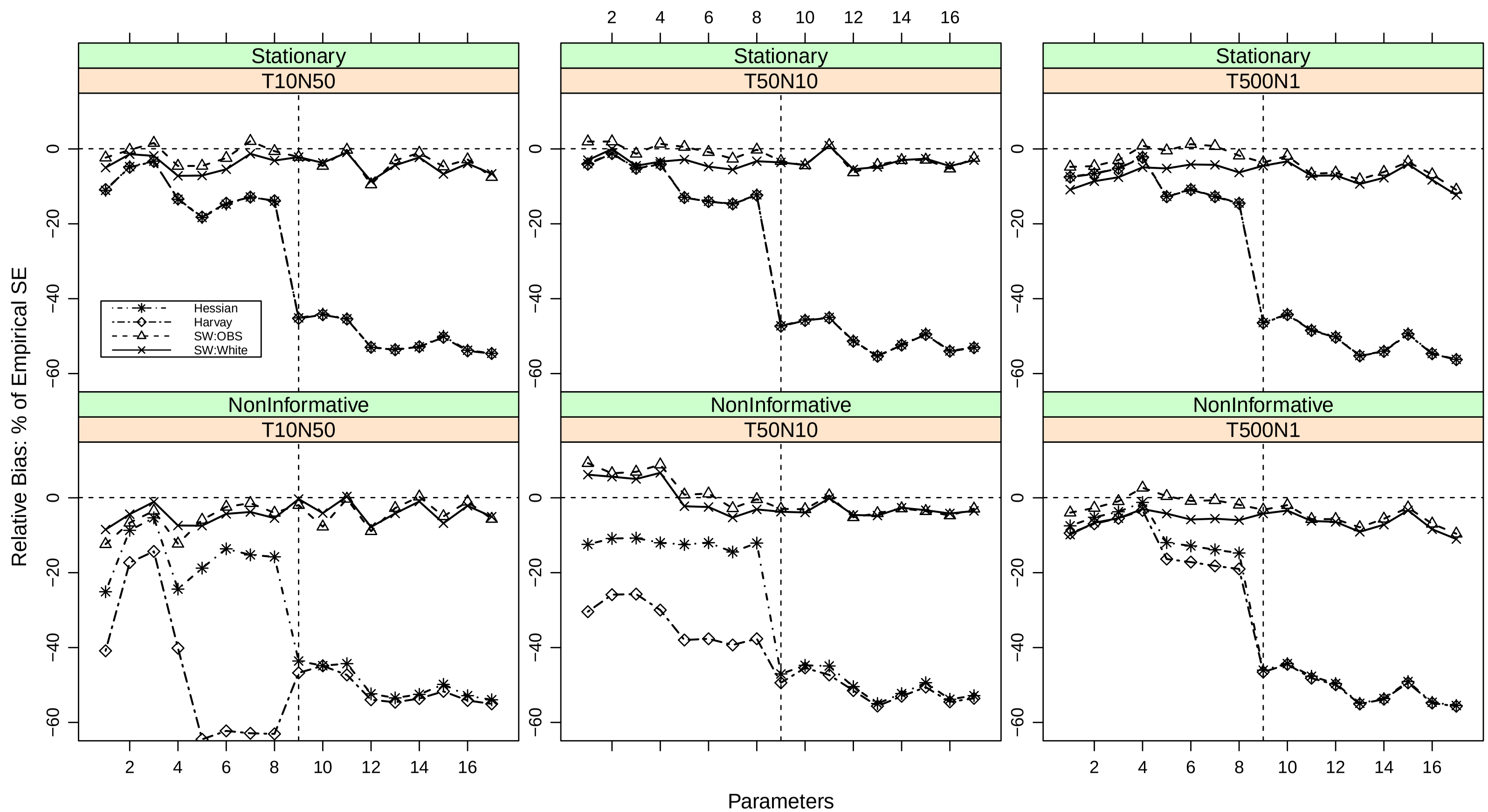

Figure 6.10: Relative bias of SE estimates obtained when fitting model M12 under different length and number-of-participant conditions. Measurement errors and state noise variables in these conditions follow the multivariate contaminated normal distribution respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

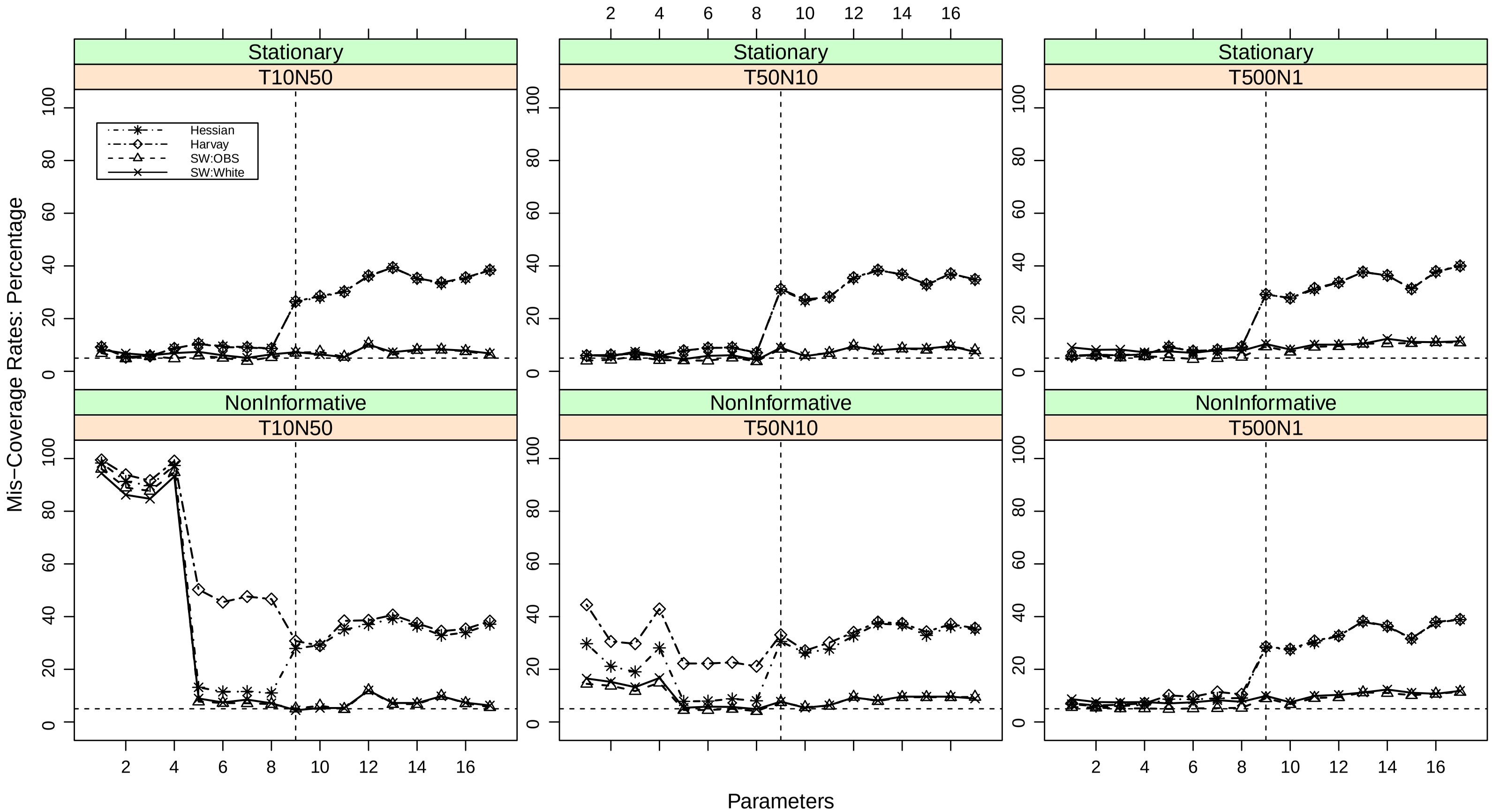

Figure 6.11: Mis-coverage rates under obtained when fitting model M12 different length and number-of-participant conditions. Measurement errors and state noise variables in these conditions follow the multivariate contaminated normal distribution respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

6.3.5 Signal-To-Noise Ratio

The performance of SE estimators is influenced by signal-to-noise ratios. Figure 6.12 displays the relative bias of SE estimates and corresponding mis-coverage rates under different ratio conditions. Results are obtained when the stationary initial settings are used. We can see that the relative bias of SE estimates is smaller and the mis-coverage rates are closer to the nominal level in the conditions with higher signal-to-noise ratios. The effect of signal-to-noise ratio is uniform across other conditions.

And when the signal-to-noise ratio is high, the White's sandwich-type SE estimator, $\widehat{SE}_{SW_W}$, may give more severely biased SE estimates of $\mathbf{\Phi}$ than information based SE estimators. This is especially apparent when the time series length is not sufficiently long.

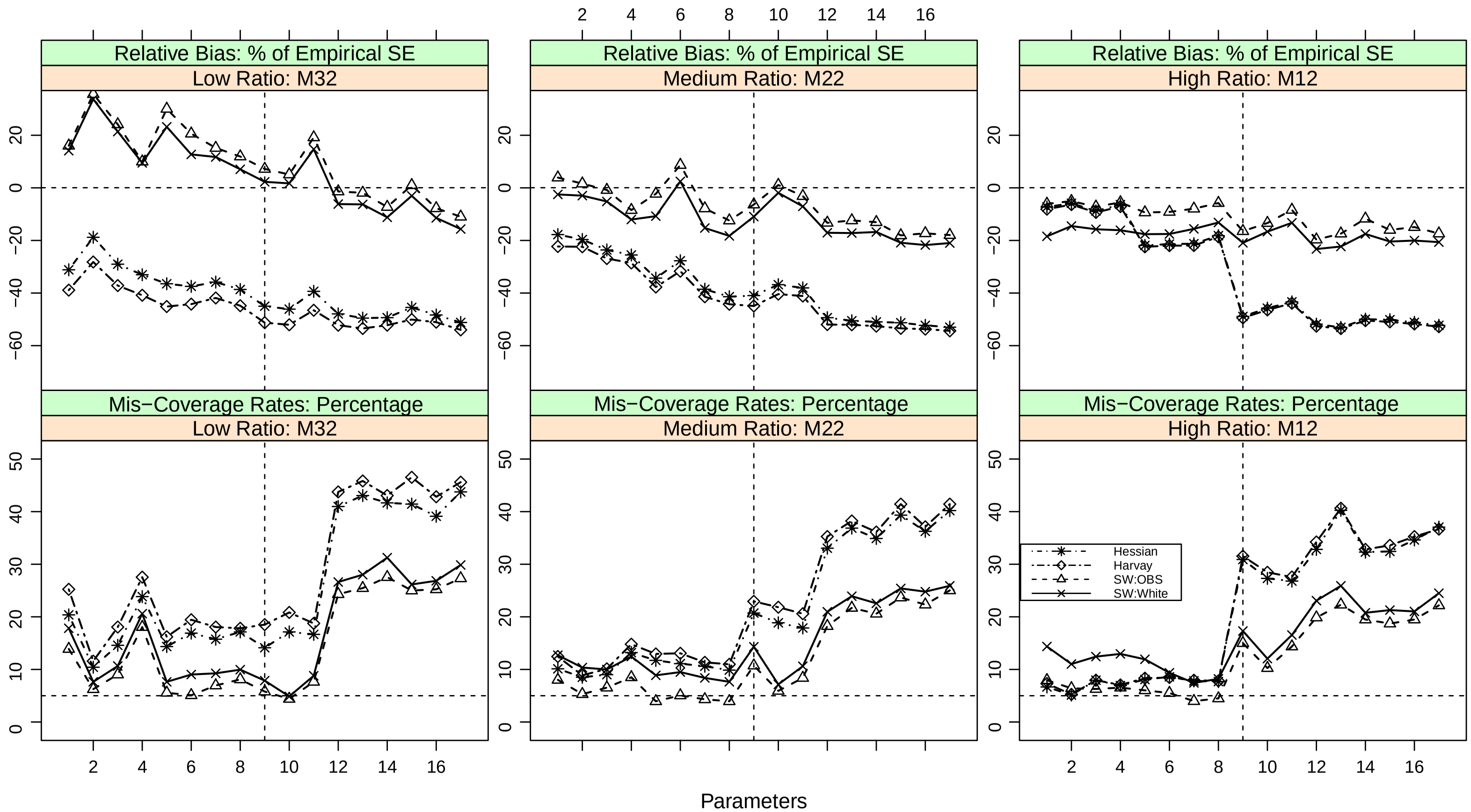

Figure 6.12: Relative bias of SE estimators and mis-coverage rates under different ratio conditions. Results are obtained using the stationary initial setting. The time series length is 100. Measurement errors and state noise variables in these conditions follow the multivariate contaminated normal distributions respectively. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

CHAPTER 7

DISCUSSIONS

In this thesis, the main objective is to examine the finite sample performance of information based SE estimators and sandwich-type SE estimators for dynamic factor analysis. The consequence of violation of the normality assumption is of primary interest. A Monte Carlo simulation is conducted so that the performance of these two types of SE estimators are examined in terms of bias and mis-coverage rates under various distribution, length, initial setting, number-of-subjects and signal-to-noise ratio conditions. Discussions of results are divided into five parts. Section 7.1, section 7.2 and section 7.3 discuss the findings concerning the normal theory based ML estimates, SE estimates and interval estimates respectively. Section 7.4 discusses several limitations of this thesis. Overall implications for applied research are included in section 7.5.

7.1 The Normal Theory Based ML Estimator

7.1.1 The Normal Theory Based ML Estimates Are Less Biased when the Time Series Length Is Sufficiently Long

A major finding of the simulation study in this thesis is that the normal theory based ML estimates become less biased as the length of the time series increases. When the time series length is sufficiently long, parameter values can be accurately recovered by the normal theory based ML estimates irrespective of the data distribution. The estimates of the noncovariance component parameters (the transition matrix $\mathbf{\Phi}$ and the measurement matrix $\mathbf{A}$) show smaller bias than the covariance component parameters (the covariance matrix of the measurement errors $\mathbf{R}$ and the covariance matrix of the state noise $\mathbf{Q}$). That the bias reduces as the time series length increases is in accordance with the previous findings of the asymptotic distribution of the normal theory based ML estimator. It is well established that ML estimates obtained using the state-space approach are consistent and efficient if data are normally distributed and the model is correctly specified (Caines, 1988; Harvey, 1989). With regard to non-normal data, Harvey and Shephard (1996, section 3) claimed, based on the work of Dunsmuir (1979), that the normal theory based ML estimator which ignored the non-normality of data was still consistent and asymptotically normal.

The simulation study in this thesis also shows that the accuracy (the bias) of the normal theory based ML estimator is influenced by the signal-to-noise ratio. When the signal-to-noise ratio is low, a much longer time series length is needed to obtain the desired accuracy. Thus the signal-to-noise ratio is an important factor that applied researchers need to take into account in the preparation of studies on dynamics. Even though it can be intuitively expected that the performance of parameter estimators is influenced by the quality of data, it is unclear how to quantify the quality of data. The signal-to-noise ratio provides a way to quantify the quality of data and thus it will be helpful in making experimental design decisions. The signal-to-noise ratio is related to the communality in factor analysis and the growth curve reliability in longitudinal analysis[1]. The communality and the growth curve reliability (GCR) measure the proportion of variance in a variable that can be explained by all common factors jointly (or by interindividual differences in change). In time series analysis, the signal-to-noise ratio is related to the proportion of variation in a variable that can be explained by all latent series. This proportion is a monotonically increasing function of the signal-to-noise ratio, i.e., $1 - \frac{1}{r_{s/n}+1}$ where $r_{s/n}$ denotes the signal-to-noise

[1] There are many growth curve reliabilities. Here I discuss the one used in Hertzog, Lindenberger, Ghisletta, and Oertzen (2006).

ratio. Therefore, higher signal-to-noise ratios are associated with higher proportions of variance explained by all latent series. And previous simulation studies showed that communality and the growth curve reliability exhibited similar influences on the parameter estimation and statistical inferences in covariance structure analysis and in longitudinal analysis (MacCallum, Widaman, Preacher, & Hong, 2011; MacCallum, Widaman, Zhang, & Hong, 1999; Hertzog et al., 2006; Hertzog, Oertzen, Ghisletta, & Lindenberger, 2008). Signal-to-noise ratios, communalities and GCRs are related to but different from the reliability. When there are no specific factors or no intraindividual differences, these three indices are indicators of the reliability. [2]

The finite sample performance of the normal theory based ML estimator for non-normal data has been examined in a few previous studies (Harvey & Shephard, 1996; Ruiz, 1994; Sandmann & Koopman, 1998). However, their attentions mainly centered on the sample size that can be often seen in typical economic studies ($T \geq 500$), which is usually much longer than the time series length considered in social and behavioral studies. And the model examined in those papers was a scalar state space model. The generalization of the findings obtained from a scalar model to a multivariate model is questionable. Note that usually, a multivariate state space model is needed in order to do the dynamic factor analysis. In this thesis, the performance of the normal theory based ML estimator is examined using time series lengths that can be often seen in social and behavioral studies, which are much smaller than the sample sizes used in economic studies. And a typical multivariate DFA model is used

[2]Strictly speaking, reliabilities in factor analysis and in longitudinal data analysis are different from communalities and GCRs. In factor analysis, the variance of an observed variable can be decomposed into common variance (variance explained by common factors), specific variance (variance explained by specific factors) and error variance. The reliability is defined as the proportion of variance in an observed variable explained by both the common factors and the specific factors. In longitudinal data analysis, the variance of an observed variable can be decomposed into interindividual differences in change, intraindividual variations and error variance. The corresponding reliability is defined as the proportion of variance in an observed variable that is explained by interindividual differences in change and intraindividual variations. If there are no specific factors or no intraindividual variations, the communality and the GCR can be used as indicators of reliabilities. Similarly, when the disturbance series, $\{\mathbf{w}_t\}$, is purely measurement error series, $1 - \frac{1}{r_{s/n}+1}$ where $r_{s/n}$ denotes the signal-to-noise ratio can be viewed as an indicator of the reliability of time series.

in the simulation study. Thus this thesis is more informative for applied research in social and behavioral area. Song and Ferrer (2009) and Z. Zhang et al. (2008) recently studied the finite sample performance of the ML estimator obtained using the Kalman smoother algorithm in the context of dynamic factor analysis. However, their study was restricted to normally distributed data. And the effect of the signal-to-noise ratio was not emphasized. This thesis extends the results of Song and Ferrer (2009) and Z. Zhang et al. (2008) to non-normal data. And the effect of the signal-to-noise ratio is examined.

7.1.2 Results from Designs that Share the Same Total Number of Observations but Differ in Time Series Length Are Comparable.

This thesis compares two types of experimental designs, one with fewer participants but a longer time series length and the other with more participants but a shorter time series length. Comparisons are made among conditions that share the same total number of observations. Simulation results show that if the stationary initial setting is used, the accuracy of parameter estimates in the multiple-subject conditions with shorter time series lengths is only slightly inferior to that in the single-subject conditions with much longer time series lengths. And the differences become negligible when the signal-to-noise ratio is high. Note that the signal-to-noise ratio in this thesis is lower than the ratio used in previous papers, i.e., Z. Zhang et al. (2008) and Song and Ferrer (2009). We thus expect smaller differences among conditions that share the same total number of participants under the conditions of previous papers. However, one should not overgeneralize the simulation results to experimental designs with time series lengths that are arbitrary short. A reasonable number of occasions is always needed to accurately model the underlying process. Our simulation study only compares conditions with time series lengths that are larger than 10. The pattern of results from data with shorter time series lengths can be different from the pattern found in this thesis. For example, the parameter estimation from data with 100 subjects and 5 waves may be less accurate than the parameter estimation from data with 1 subject and 500 time points. Another issue that requires special caution is that this thesis implicitly assumes, when analyzing the multi-subject data, that all participants are governed by the same process. In real data analysis, a test on the homogeneity of the underlying systems of different participants is always needed.

7.2 Standard Error Estimators

7.2.1 Papanastassiou's Sandwich SE Estimator Gives Better Performance

The simulation study indicates that whereas for normal data, the four SE estimators yield similar performance, for non-normal data, Papanastassiou's sandwich SE estimator, $\widehat{SE}_{SW_{p1}}$, outperforms the other SE estimators in terms of smaller relative bias and better controlled mis-coverage rates. And the advantage of using Papanastassiou's sandwich SE estimator is more substantial for $\mathbf{\Phi}$ and $\mathbf{A}$ than for $\mathbf{R}$ and $\mathbf{Q}$. Even though information based SE estimators work adequately well for $\mathbf{\Phi}$ and $\mathbf{A}$ when the time series length is long and when the signal-to-noise ratio is high, they do not otherwise.

For designs with multiple subjects but shorter time series lengths, the simulation study indicates that when the stationary initial setting is used, similar accuracy of SE estimates and comparable control in mis-coverage rates are observed among conditions that share the same total number of observations. Generally, in multi-subject data conditions (conditions with $N > 1$), sandwich-type SE estimators yield accurate SE estimates and the corresponding mis-coverage rates are well controlled across all ratio conditions. And information SE estimators yield downward biased SE estimates and inflated mis-coverage rates, especially when the signal-to-noise ratio is not high (medium or low).

Thus in general, it is safer to use Papanastassiou's sandwich SE estimator.

Between the two sandwich-type SE estimators, the performance of $\widehat{SE}_{SW_W}$ is generally comparable to $\widehat{SE}_{SW_{P1}}$, except in conditions where the time series length is short and the signal-to-noise ratio is high. In this situation, $\widehat{SE}_{SW_W}$ tends to show larger negative bias than $\widehat{SE}_{SW_{p1}}$. This is against our expectation. $\widehat{SE}_{SW_W}$ is supposed to give better performance than $\widehat{SE}_{SW_{P1}}$ because $\widehat{SE}_{SW_W}$ tries to capture the dependence among $\dot{l}_t$s. To evaluate the dependence among $\dot{l}_t$s, figure 7.1 is included. I use the approximate auto- and cross-correlations to study the dependence among $\dot{l}_t$s (please refer to Appendix E for technical details). Because all models in this thesis have 17 parameters, $\dot{l}_t$ contains 17 univariate derivative series. Figure 7.1 graphs the approximate ACFs of all the 17 derivative series. We can see from figure 7.1 that for non-normal data, $\dot{l}_t$s based on the prediction error decomposition do not appear to be serially correlated or the dependence among $\dot{l}_t$s is negligible. Further discussions on the cross-correlations of $\left\{\dot{l}_t\right\}$ can be found in Appendix E.

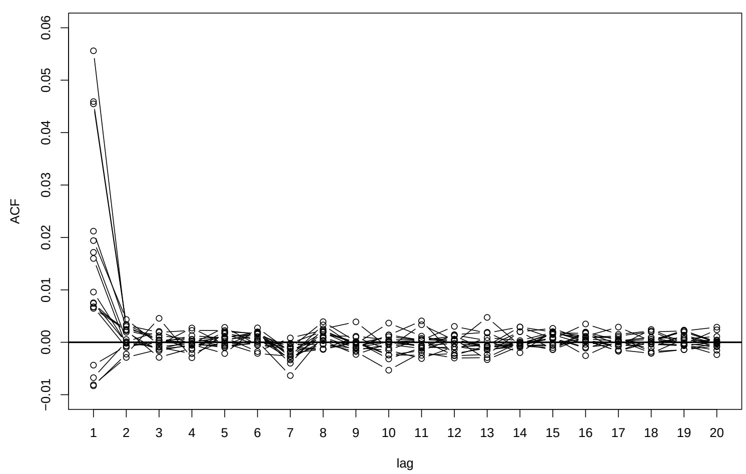

Figure 7.1: All approximate ACFs of $\left\{\dot{l}_t\right\}$ when true parameters are known. Results are obtained based on samples generated from model M12 and under the $T = 500$ and the contaminated normal data condition.

The two information based SE estimators give similar SE estimates in many conditions studied in this thesis. Even though when the noninformative initial setting is used and when the time series length is short, Harvey's SE estimator shows larger downward relative bias than the observed information based SE estimator. The performance of Harvey's SE estimator can be substantially improved by using the stationary initial setting. Therefore, given its computational efficiency, Harvey's SE estimator is a useful alternative to the observed information based SE estimator.

There have been relatively few studies exploring the finite sample performance of standard error estimators for DFA models, especially under the framework of state-space models. The available evidence is limited in scope and cannot be generalized directly to DFA models (Papanastassiou, 2006). This thesis extends the findings of previous papers by examining the finite sample performance of standard error estimators under conditions that can be encountered in typical social and behavioral studies. Dynamical studies in social and behavioral areas usually involve samples with limited number of occasions, severe deviation from the multivariate normal distribution, less-than-perfect signal-to-noise ratios and multiple latent concepts. All these factors are considered in this thesis. Thus this thesis will enrich the understanding of finite-sample properties of SE estimators for the dynamic factor analysis.

7.2.2 Information-Based SE Estimators Work Only When the Time Length Is Long and When the Ratio Is High

The simulation study indicates that for the non-covariance parameters ($\mathbf{A}$ and $\mathbf{\Phi}$), the performance of information based SE estimators depends on the time series length and the signal-to-noise ratio. When the time series length is long and when the signal-to-noise ratio is high, standard errors of $\mathbf{A}$ and $\mathbf{\Phi}$ can be accurately estimated by information based SE estimators and interval estimates of $\mathbf{A}$ and $\mathbf{\Phi}$ yield mis-coverage rates that are close to the nominal level. In medium ratio conditions, the accuracy of information based SE estimates and the controlled on mis-coverage rates improved as the time series length increases. In contrast, for $\mathbf{Q}$ and $\mathbf{R}$, information based SE estimates are uniformly downwards biased and interval estimates using information based SE estimates always produce highly inflated mis-coverage rates. In low ratio conditions, however, information based SE estimators tend to severely underestimate the standard errors across all length conditions considered in this thesis. And the interval estimates using information-based SE estimates produce inflated mis-coverage rates.

Though counterintuitively, this result is in agreement with the findings of Dunsmuir and Hannan (1976) and Dunsmuir (1979, p499). The results of these papers implied that the vector ARMA model, under some regularity conditions, can be separately parameterized. And the sandwich-type asymptotic covariance matrix was blockwise diagonal. And only the asymptotic covariance matrix of covariance component parameters depends on the fourth moments of noise variables. The asymptotic covariance matrix of the autoregressive weights and moving average weights is not influenced by the fourth moments of noise variables. However, Dunsmuir (1979) also noted that this parameterization separation may not always apply to models observed with noise, i.e., the state space model, even though the asymptotic normality is still valid (but the covariance matrix is not blockwise with respect to the partition of parameters) if the fourth moments of the noise variables exist (Dunsmuir, 1979, p499-p500). More thorough work on the asymptotic covariance matrix of the normal theory based ML estimator for the state space model is needed.

7.2.3 The Negative Relative Bias of SE Estimators

The simulation study shows that in most conditions, SE estimates tend to underestimate the standard errors of parameters, even in the normal data conditions. One explanation of the negative relative bias in the normal data conditions is that SE estimates are computed using the formulas derived based on the asymptotic covariance matrix, i.e., the inverse of the Fisher information matrix. This covariance matrix is the smallest covariance matrix that any estimator can attained. Let SE_a denotes the asymptotic standard error based on the asymptotic covariance matrix. The empirical SE estimates estimate the variance of point estimates for certain sample size, namely, SE_T. Since SE_a is based on the lower bound of the covariance matrix of estimators, it is reasonable that the estimates of SE_a are smaller than the estimates of SE_T. In the simulation study, empirical SE estimates are estimates of SE_T and standard error estimates are estimating SE_a. Thus it is natural that SE estimates in the normal data condition produce negative bias.

Another explanation for the negative relative bias in the normal data condition is that all SE estimators considered in this thesis are not unbiased SE estimators. This is because the asymptotic covariance matrix of parameters is scaled by $\frac{1}{T}$ rather than $\frac{1}{T-k}$ where k is the number of parameters in the proposed DFA model. The rationale is analogous to the correction of the degree of freedom when estimating the error variance in regression analysis. k degrees of freedom are lost because the k estimated parameters are used in computing the prediction errors. This explanation also applies to the two sandwich SE estimators in the non-normal data conditions.

In addition, information based SE estimates are negatively biased in non-normal data conditions, especially for $\mathbf{R}$ and $\mathbf{Q}$. This is likely due to the fact that information based SE estimates are asymptotically downward biased, especially for $\mathbf{R}$ and $\mathbf{Q}$. This is because information based SE estimates rely on the Fisher information matrix, which is the lower bound of the covariance matrix of any estimator. When the normality assumption is violated, the covariance matrix of the normal theory based ML estimator is believed to become larger (i.e., the point estimates are less efficient). Thus it is reasonable that information based SE estimates display negative bias in the non-normal data conditions.

7.3 Interval Estimates

For non-normal data, the simulation study indicates that based on the normal theory based ML estimates, the interval estimates using Papanastassiou's sandwich SE estimates display most accurate mis-coverage rates - mis-coverage rates that are closest to the nominal level. Mis-coverage rates of interval estimates (based on Papanastassiou's sandwich SE estimates) approach the nominal level as the time series length increases and as the signal-to-noise ratio increases. And the mis-coverage rates for $\mathbf{\Phi}$ and $\mathbf{A}$ are better controlled than those for $\mathbf{R}$ and $\mathbf{Q}$. This indicates that for $\mathbf{R}$ and $\mathbf{Q}$, longer time series lengths are needed in order to construct confidence intervals that have correct mis-coverage rates.

The mis-coverage rates of interval estimates for $\mathbf{\Phi}$ and $\mathbf{A}$ using information based SE estimates are close to the nominal level when the time series length is long ($T = 500$) and when the signal-to-noise ratio is high (high ratio condition). Interval estimates for $\mathbf{R}$ and $\mathbf{Q}$ using information based SE estimates display high mis-coverage rates across all non-normal data conditions in the simulation study.

For normal data, interval estimates using different SE estimates display similar mis-coverage rates and the mis-coverage rates are close to the nominal level as long as the time series length is sufficiently long.

7.4 Limitations of Current Thesis

Several limitations of current thesis are discussed below. First, the simulation study of this thesis is based on stationary data. Thus a test on stationarity or a method that can remove nonstationary components is needed. The chow test and the forecast scores test detailed in Lütkepohl (2005) are two available methods. In addition, preliminary treatment can be conducted to make the data more stationary. For example, a linear trend in a data set can be identified and thus removed by regressing observed scores on time.

Second, in the simulation study, models are always correctly specified. In reality, however, it is difficult to identify the structure of a latent process. One possible way is to compute the composite scores of each latent factor and treat the composite series as a manifest version of the latent factor series. Then procedures of identifying the structure of manifest series and the procedures of checking the adequacy of models can be applied to the composite series. Nevertheless further research on how to identify the structures of a latent process is needed.

Third, this thesis implicitly assumes that for multi-subject data, there are no individual differences in the underlying system. Thus a method that can identify a homogeneous subgroup or a test examining the homogeneity of the underlying systems of different participants is needed. Unfortunately, we are not aware of any well-established method or test. A method proposed by Nesselroade and Molenaar (1999) has been used in some applied research (Chow et al., 2004). This approach compares participant's lagged covariance matrices via an algorithm that is similar to the cluster analysis. Overall this is a problem that deserves more thorough research.

Finally, in this thesis, I only consider the situation that a latent non-normal process is contaminated by non-normal measurement errors. It is possible that the non-normality in real data only comes from the measurement errors or from the latent series. However, since the models considered in this thesis show larger departure from the standard normal model, it is likely that the influence of non-normality in this thesis is larger than that in those situations. Hence SE estimators that give good performance in this thesis are expected to give better performance in situations where only measurement errors do not follow a normal distribution or only the latent series is not a normal process.

7.5 Implications for Applied Research

This thesis provides useful information for applied researchers who are interested in modeling intraindividual variation. First, model parameters can be accurately recovered by the normal theory based ML estimator as long as the time series length is sufficiently long, even if data do not follow a multivariate normal distribution. How long is sufficiently long depends on the data distribution, the signal-to-noise ratio and the complexity of the DFA model. For non-normal data with low signal-to-noise ratios, a longer time series length is needed to obtain desired level of accuracy. For models considered in this thesis, a time series length of 200 seems to be sufficient for the transition matrix $\mathbf{\Phi}$ and the measurement matrix $\mathbf{A}$ across all distribution and signal-to-noise ratio conditions. For $\mathbf{R}$ and $\mathbf{Q}$, this number increases to 500.

Second, Papanastassiou's sandwich SE estimator, $\widehat{SE}_{SW_{p1}}$, is generally more preferable to the other SE estimators. Mis-coverage rates of interval estimates constructed using Papanastassiou's sandwich SE estimates are closer to the nominal level than those using other SE estimates. In general, for models considered in this thesis, a time series length of 500 is needed so that interval estimates of $\mathbf{A}$ and $\mathbf{\Phi}$ using Papanastassiou's sandwich SE estimates can be trusted across all ratio conditions. For $\mathbf{R}$ and $\mathbf{Q}$, even a length of 500 is not sufficient for interval estimates to be trusted if the signal-to-noise ratio is low. Interval estimates using information based SE estimates are trustworthy only when the time series length is long ($T = 500$) and when the signal-to-noise ratio is high ($r_{s/n} = 5$ or 3.2) and only for $\mathbf{A}$ and $\mathbf{\Phi}$. Generally, longer time series lengths are needed in order to obtain reliable interval estimates when the signal-to-noise ratio is low.

Third, experimental designs with multiple participants but relatively short time series lengths are suitable for the dynamic factor analysis. However, this conclusion depends on the complexity of the model and the quality of data (the signal-to-noise ratio). A reasonable time series length is always needed to accurately recover the underlying dynamical process. And a longer time series length is needed when the complexity of the model increases and the quality of data deteriorates.

Finally, the DFA model is often used to model self-reported data (e.g., Chow et al., 2004; Musher et al., 2002). Thus, the results of this thesis can be applied to self-reported data. For bio-signal data, e.g., fMRI data, however, extensions on the state space models considered in this thesis are needed. For example, if the hemodynamic response function is estimated using the parametric method, the observation matrix $\mathbf{A}$ is known but time varying. Detailed discussions can be found in Ho (2003); Smith, Pillai, Chen, and Horwitz (2010, 2012).

APPENDIX A

DATA GENERATION

Since complicated multivariate non-normal distributions are involved, special programs developed in R are introduced to generate samples from target distributions.

Log-Normal

Suppose the sample $\mathbf{Y}$ from a target log-normal distributions with mean of μ and covariance of $\mathbf{\Sigma}$ is of interests. Note that μ is solely determined by $\mathbf{\Sigma}$. Let $\mathbf{X} = \log \mathbf{Y}$. From probability theories, $\mathbf{X}$ follows a normal distribution, $N\left(\mathbf{u}, \mathbf{D}\right)$. Moreover, μ and $\mathbf{\Sigma}$ are functions of $\mathbf{u}$ and $\mathbf{D}$.

$$
\begin{aligned}
\mu_i &= E\left(y_i\right) \\
&= E\left(e^{\mathbf{t}'\mathbf{X}}\right) && \text{(A.1)} \\
&= exp\left(\mathbf{t}'\mathbf{u} + \frac{1}{2}\mathbf{t}'\mathbf{D}\mathbf{t}\right) && \text{(A.2)} \\
&= exp\left(u_i + .5d_{ii}\right)
\end{aligned}
$$

where $\mathbf{t} = (0, 0, \ldots, 1, \ldots, 0)'$.

$$
\begin{aligned}
\sigma_{ij} &= cov\left(y_i, y_j\right) \\
&= E\left(y_i, y_j\right) - E\left(y_i\right) E\left(y_j\right) \\
&= E\left(e^{x_i} e^{x_i}\right) - \mu_i \mu_j \\
&= E\left(e^{\mathbf{t}'\mathbf{X}}\right) - \mu_i \mu_j && \text{(A.3)} \\
&= exp\left(\mathbf{t}'\mathbf{u} + \frac{1}{2}\mathbf{t}'\mathbf{D}\mathbf{t}\right) - \mu_i \mu_j \\
&= exp\left[u_i + u_j + .5\left(d_{ii} + d_{jj} + 2d_{ij}\right)\right] - \mu_i \mu_j && \text{(A.4)} \\
&= \mu_i \mu_j \left[exp\left(d_{ij}\right) - 1\right]
\end{aligned}
$$

where $\mathbf{t} = (0, 0, \ldots, 1, \ldots, 1, \ldots, 0)'$.

Let $\mathbf{u} = \mathbf{0}$. Then by inverting (A.2) and (A.4) we would obtain that

$$
d_{ii} = 2\log\left(\mu_i\right) \text{ and } d_{ij} = \log\left(\frac{\sigma_{ij}}{\mu_i \mu_j} + 1\right). \tag{A.5}
$$

Thus the sample $\mathbf{Y}$ can be generated using the the following steps:

- Compute $\mathbf{D}$ from $\mathbf{\Sigma}$ using equation (A.5).
- Generate a sample $\mathbf{X} \sim N\left(\mathbf{0}, \mathbf{D}\right)$.
- Compute $\mathbf{Y}$ using the transformation $\mathbf{Y} = e^{\mathbf{X}}$.
- Shift the mean of $\mathbf{Y}$ to $\mathbf{0}$ by subtracting $\mathbf{Y}$ from μ.

t Distribution

Data from a t(df=p) distribution with a zero mean vector and a covariance matrix, $\mathbf{\Sigma}$ can be generated using the following procedure:

S1. Generate $\mathbf{x}$ from $MN\left(\mathbf{0}, \frac{p-2}{p}\mathbf{\Sigma}\right)$ and z^2 from χ^2_p independently.

S2. The data vector, $\mathbf{y}$, equal to $\frac{\mathbf{x}}{\sqrt{z^2/p}}$.

Contaminated Normal

Suppose a $\epsilon-$contaminated normal distribution with zero mean and a target covariance matrix, $\mathbf{\Sigma}$ is of interests. The density function of this distribution has the form (Anderson, 2002, p55),

$$(1-\epsilon)\, N_p\left(\mathbf{0}, \mathbf{D}\right) + \epsilon N_p\left(\mathbf{0}, c\mathbf{D}\right)$$

where $N_p\left(\mathbf{0}, \mathbf{D}\right)$ denotes a $p-$variate normal distribution with mean zero and covariance matrix, $\mathbf{D}$. c is the multiplier of the covariance matrix of one of the two normal distributions and ϵ be the percentage of this normal component. The covariances matrix of contaminated normal distribution is given by $\mathbf{\Sigma} = (1-\epsilon+c\epsilon)\,\mathbf{D}$. Data is generated using the following steps:

S1. Compute a covariance matrix, $\mathbf{D} = {}^{\mathbf{\Sigma}}/_{1-\epsilon+c\epsilon}$.

S2. Generate a random number, u~U(0,1). If $u \geq \epsilon$, draw a data vector, $\mathbf{x}_i$ from $MN\left(\mathbf{0}, \mathbf{D}\right)$ and from $MN\left(\mathbf{0}, c\mathbf{D}\right)$ otherwise.

S3. Save all vectors and the resulting data set is the target sample.

APPENDIX B

STATIONARITY, CONTROLLABILITY AND OBSERVABILITY

Stationarity. A question of interest regarding the performance of estimation is whether or not the system is stationary. In an unstable system, the random shocks may sometimes excite the system and the resulting observations change erratically. Then the theorems of convergence may not apply. A wide sense stationary process, instead, is relatively stable in that its means and covariances across time are the same. Formally speaking, a m-component times series, $\{\mathbf{y}_t\}$, is stationary, if,

$$\begin{aligned} E\left(\sum_{i=1}^{m} y_{t,i}\right) &< \infty, \\ E(\mathbf{y}_t) &= \mu \text{ and} \\ E(\mathbf{y}_t, \mathbf{y}_{t+k}) &= E(\mathbf{y}_s, \mathbf{y}_{s+k}) \end{aligned}$$

for $t, s = 1, 2, \ldots, T$ and $k = 0, 1, \ldots$. In the context of the state space model, the latent state, $\{\mathbf{x}_t\}$, and consequently the observed process, $\{\mathbf{y}_t\}$, is stationary if and only if,

$$|\lambda_i(\mathbf{\Phi})| < 1$$

where $\lambda_i(\mathbf{A})$ is the ith eigenvalue of $\mathbf{A}$.

Controllability. The second condition that is worth mentioning is controllability. In the state space model, the m step ahead latent state can be determined by the current state and the m future random shocks,

$$\begin{aligned}\mathbf{x}_{t+m} &= \sum_{j=0}^{m-1} \mathbf{\Phi}^j \mathbf{w}_t + \mathbf{\Phi}^m \mathbf{x}_t \\ &= \mathbf{C}\mathbf{U}_t + \mathbf{\Phi}^m \mathbf{x}_t\end{aligned}$$

where m is the dimension of the state vector, $\mathbf{x}_t$, and $\mathbf{C} = [\mathbf{I}, \mathbf{\Phi}, \mathbf{\Phi}^2, \ldots, \mathbf{\Phi}^{m-1}]$ and $\mathbf{U}_t = \left(\mathbf{w}'_{t+m}, \ldots \mathbf{w}'_{t+1}\right)'$. The process is controllable if any value of $\mathbf{x}_{t+m}$ can be obtained from any value of $\mathbf{x}_t$ by controlling $\mathbf{U}_t$ (Shumway & Stoffer, 2004). Mathematically, if $\operatorname{rank}(\mathbf{C}) = m$, the system is said to be controllable.

Observability. The third condition concerns the amount of information of $\mathbf{x}_t$ that can be inferred from m future observations. In an observable pure signal system (the system without any noise term), the m future observations can be directly mapped into the space of $\mathbf{x}_t$, that is,

$$\left(\mathbf{y}'_t, \ldots, \mathbf{y}'_{t+m-1}\right) = \mathbf{x}'_t \mathbf{O}'$$

or

$$\mathbf{x}_t = \left(\mathbf{O}'\mathbf{O}\right)^{-1} \mathbf{O}' \left(\mathbf{y}'_t, \ldots, \mathbf{y}'_{t+m-1}\right)'$$

where $\mathbf{O}' = [\mathbf{A}', \mathbf{\Phi}'\mathbf{A}', \ldots, \mathbf{\Phi}'^{m-1}\mathbf{A}']$. The invertibility of $\mathbf{O}'\mathbf{O}$ requires $\mathbf{O}$ to be of full rank. In other words, the process is observable if $\operatorname{rank}(\mathbf{O}) = m$.

APPENDIX C

CONDITIONS FOR ASYMPTOTIC NORMALITY

Caines listed three major regularity conditions. First, the observed process has to be exponentially stable. A process is said to be exponentially stable, if the dependence of the moments on preceding observations decays exponentially. This condition allows the use of the main convergence theorems. Second, the predictors of the observations need to be a family of exponentially stable predictors. This family of predictors produces exponentially stable estimates if the input process is exponentially stable. Third, the criterion function, such as likelihood function, needs to be appropriately bounded. For formal definitions of these conditions, please refer to Caines(1988, chapter 8).

This set of regularity conditions has a weak version and a strong version. In the strong version, it also requires the second derivative of the predictors of observations to be a family of exponentially stable predictors. However, in the weak version, only the predictor and its first derivative are required to satisfy this condition. The weak version of the regularity conditions is adequate to establish the consistency of estimates (Caines, 1988, p.517). A standard state space model (a state space model that is stationary, controllable and observable) satisfies the weak version of the regularity conditions (see discussions in Caines, 1988 on p.488-489, p.494-496 and p.498-500). Thus normal theory based ML estimates for the standard state space model are consistent. The strong version of the regularity conditions permits the asymptotic normality. However, it is unclear whether or not the standard state space model satisfies the strong version of conditions.

In an unpublished paper, White (1984) derived the same asymptotic normality using a different set of conditions. For example, instead of using the concepts of exponentially stable, he used mixing conditions to define the stability required for convergence theorems. No discussions on whether the standard state space model will meet the conditions are found, however.

Dunsmuir and Hannan (1976) established the asymptotic normality for normal theory based ML estimates of a close related model. This model contains all times series models that can be transferred to a vector MA model with infinity order. They also extended their results to the autoregressive signal plus noise model (Dunsmuir, 1979). The state space model can be viewed as the autoregressive signal plus noise model. From their results, normal theory based ML estimators had the asymptotic normality similar to the one defined in (4.4). However, their results were discussed from the frequency perspective and thus were less straightforward.

APPENDIX D

INFLUENCES OF HOW THE SIGNAL-TO-NOISE RATIO IS MANIPULATED

To illustrate that in most conditions, how the signal-to-noise ratio is manipulated would not influence the performance of the normal theory based ML estimator and SE estimators, figure D.1 and figure D.2 are included. Figure D.1 and figure D.2 graph the relative bias of normal theory based ML estimates and SE estimates under conditions that vary in how the signal-to-noise ratio is manipulated. Model M11, M12 and M13 share the same signal-to-noise ratio. However, they differ in variances of measurement errors and variances of latent factors. The same applies to model M21, M22, M23 and model M31, M32, M33. From figure D.1, we can see that the relative bias of normal theory based ML estimates is comparable among conditions that have the same signal-to-noise ratio, indicating that how the signal-to-noise ratio is manipulated does not influence the accuracy of the normal theory based ML estimator. From the first and the third columns of figure D.2, we can see that how the signal-to-noise ratio is manipulated does not affect the relative bias of SE estimators as well. In the second column, the relative bias of covariance component parameters (points on the right to the vertical dashed line) differs across different models. This is likely due to the fact that the skewness and the kurtosis of these three models are also changing in the second column (when noise variables follow the log-normal distribution). Note that the skewness and the kurtosis of the log-normal distribution are related to the first and second moments, that is, the mean and the variance. Thus when the signal-to-noise ratio is manipulated via adjusting the ratio of variances of measurement errors to variances of latent factors, the skewness and the kurtosis are also manipulated. Consequently, the differences in the relative bias of SE estimates across models are observed in the second column. For some other distributions, however, the skewness and the kurtosis are independent of the first two moments, for example, normal distribution and contaminated normal distribution. Hence for the first and the third columns, no substantial differences are observed among model M11, M12 and M13.

Figure D.3 graphs the mis-coverage rates of conditions illustrated in figure D.2. Mis-coverage rates are compared across models that differ in how the signal-to-noise ratio is manipulated. We can see from this figure that as long as the signal-to-noise ratio and other conditions are fixed, how the signal-to-noise ratio is manipulated does not affect mis-coverage rates.

Results of many other conditions agree with the findings mentioned above. However, there are some exceptions. Exceptions are observed only when examining the influences of ways of manipulating the signal-to-noise ratio on SE estimators. The observed influences on the performance of the normal theory based ML estimator are trivial across all conditions. Those exceptions in which the influences of ways of manipulating the signal-to-noise ratio are observed can be grouped into three classes. For the first class of exceptions, a few problematic samples with large sandwich-type SE estimates are observed. Figure D.4 displays one example of this type of exceptions. The signal-to-noise ratio in figure D.4 is lower than in figure D.2. We can see by comparing the three plots in the first column (the normal data conditions) of figure D.4 that the relative bias of sandwich-type SE estimates differ across model M31, M32 and M33.

Figure D.5 displays an example of another class of exceptions. For this class of exceptions, a few problematic samples with large point estimates are observed. The time series length in figure D.5 is longer compared with figure D.2 (500 vs 50). As shown in the second column (the log-normal data condition) of figure D.5, the relative bias of SE estimates appears to be different across models in which the signal-to-noise ratio is manipulated via different ways.

Figure D.6 displays an example of the third class of exceptions. Different from figure D.4, the time series length is shorter in figure D.6 (10 vs 50) but the number of participants is larger (50 versus 1). And the noninformative initial setting is used in figure D.6. By comparing the rows of figure D.6, we can see that the relative bias of SE estimators (except for Harvey's SE estimators) of transition matrix, **A**, increases from the first row to the last row. Examinations on the histograms of point estimates and SE estimates show that no problematic samples are observed. This indicates that the observed information based SE estimator and the two sandwich-type SE estimators are influenced by how signal-to-noise ratio is manipulated when time series length is as short as 10. Specifically, they underestimate the empirical SE estimates of **A** when both factor variances and measurement error variances are low, that is, when the total observation variances are low.

Table D.1 summarizes conditions that different classes of exceptions are observed. 0 means the condition is clean. No influences of how signal-to-noise ratio is manipulated are observed. 1 to 3 indicate which class of exceptions is observed. We can find from this table that, first, more conditions involve problematic samples in the low signal-to-noise ratio conditions. Second, in the median signal-to-noise ratio condition, problematic samples are observed in conditions with short time series length. Third, more conditions involve problematic samples when the noninformative initial setting is used. This indicates that, first, increasing signal-to-noise ratio helps to reduce the chances of observing problematic samples. Second, increasing time series length also helps to avoid observing problematic samples. Third, using appropriate initial settings is beneficial in eliminating the chances of observing problematic samples.

Figure D.7, figure D.8 and figure D.9 visualize mis-coverage rates of conditions in which the three classes of exceptions discussed above are observed. In all these three figures, mis-coverage rates are compared across models that differ in how signal-to-noise ratio is manipulated. We can see from figure D.7 and figure D.8 that when the first two classes of exceptions are observed, mis-coverage rates are roughly comparable across conditions varying in how the signal-to-noise ratio is manipulated. This indicates that when the first two classes of exceptions are observed, mis-coverage rates are still not influenced by how signal-to-noise ratio is manipulated. However, in case the third class of exceptions are observed, mis-coverage rates are influenced by how the signal-to-noise ratio is manipulated. Figure D.9 illustrates one example. As shown in table D.1, the third class of exceptions are observed in multi-subject conditions with time series lengths that are shorter than 20 and only when the noninformative initial setting is used. Even though differences among conditions varying in how the signal-to-noise ratio is manipulated are observed when the noninformative initial setting is used, it does not affect the conclusions of subsection 6.3.3 and subsection 6.3.4, that is, the stationary initial setting is preferable to the noninformative initial setting and Harvey's SE estimator is more sensitive compared with other SE estimators. Therefore discussions on the third class of exceptions are not included in subection 6.3.3 and subsection 6.3.4.

TABLE D.1

CONDITIONS OBSERVING INFLUENCES OF HOW THE SIGNAL-TO-NOISE RATIO IS MANIPULATED

Initial	Stationary									noninformative								
STN Ratio	Low			Median			High			Low			Median			High		
	N	LN	CN	N	LN	CN	N	LN	CN	N	LN	CN	N	LN	CN	N	LN	CN
T50N1	1	0	1	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0
T100N1	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0
T200N1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T500N1	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
T10N50	0	0	0	0	0	0	0	0	0	3	2,3	3	0	2	0	0	0	0
T20N25	0	0	0	0	0	0	0	0	0	3	2,3	3	0	0	0	0	0	0
T25N20	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
T50N10	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0

Note: The quantity in the table indicates which type of problematic samples are observed. 0 means the condition is clean. 1 means problematic samples with large sandwich-type SE estimates are observed. 2 means problematic samples with large point estimates are observed. 3 means no problematic samples are observed; but results show that in this condition how the signal-to-noise ratio is manipulated influences the performance of SE estimators.

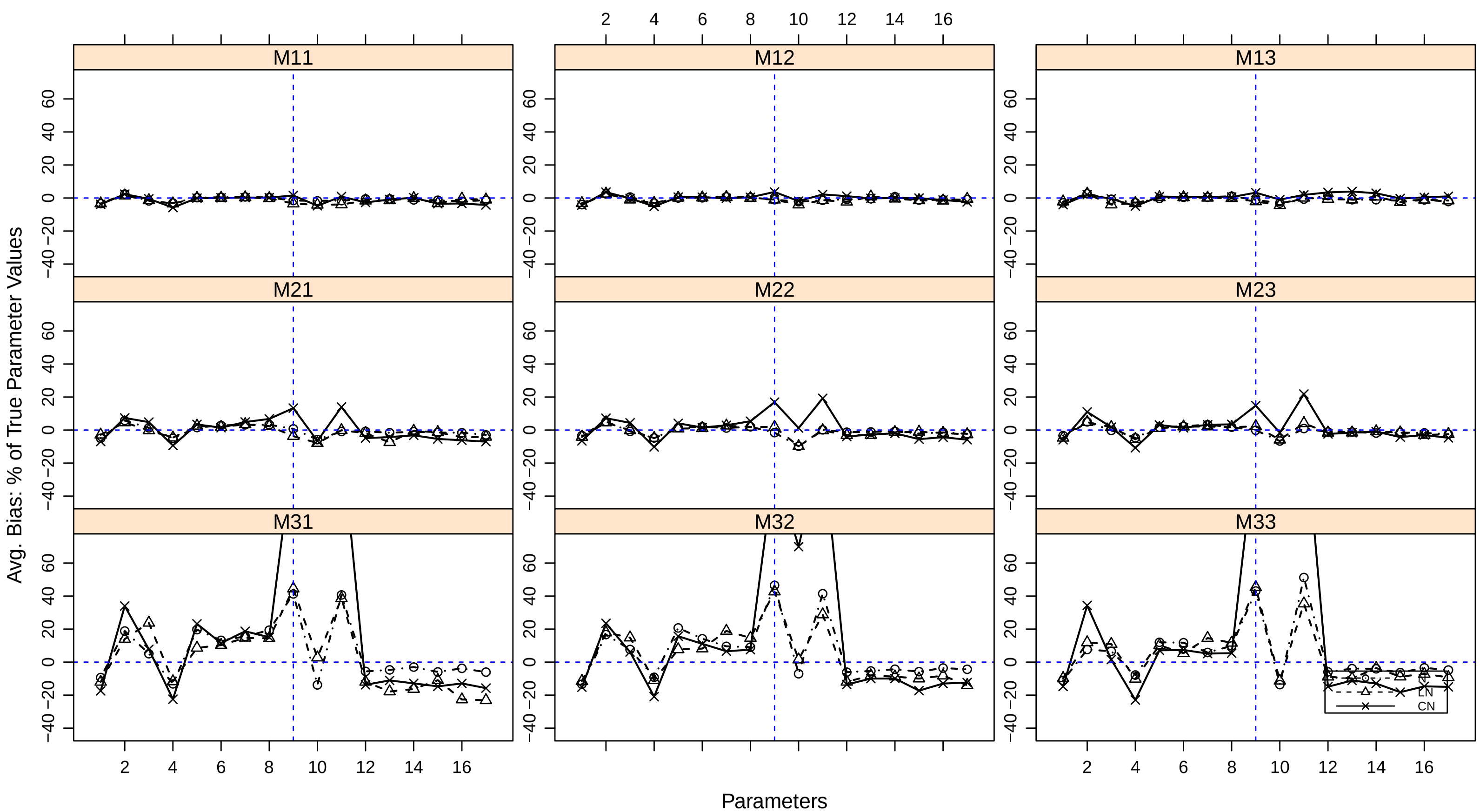

Figure D.1: Relative bias of normal theory based ML estimates obtained when the stationary initial setting is used. The relative bias of normal theory based ML estimates is given by $rbias_i = M^{-1}\sum_{j=1}^{M}\left(\hat{\theta}_{i,j} - \theta_i\right)/\theta_i$. Data are with a length of 100 and are collected from one participant. Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

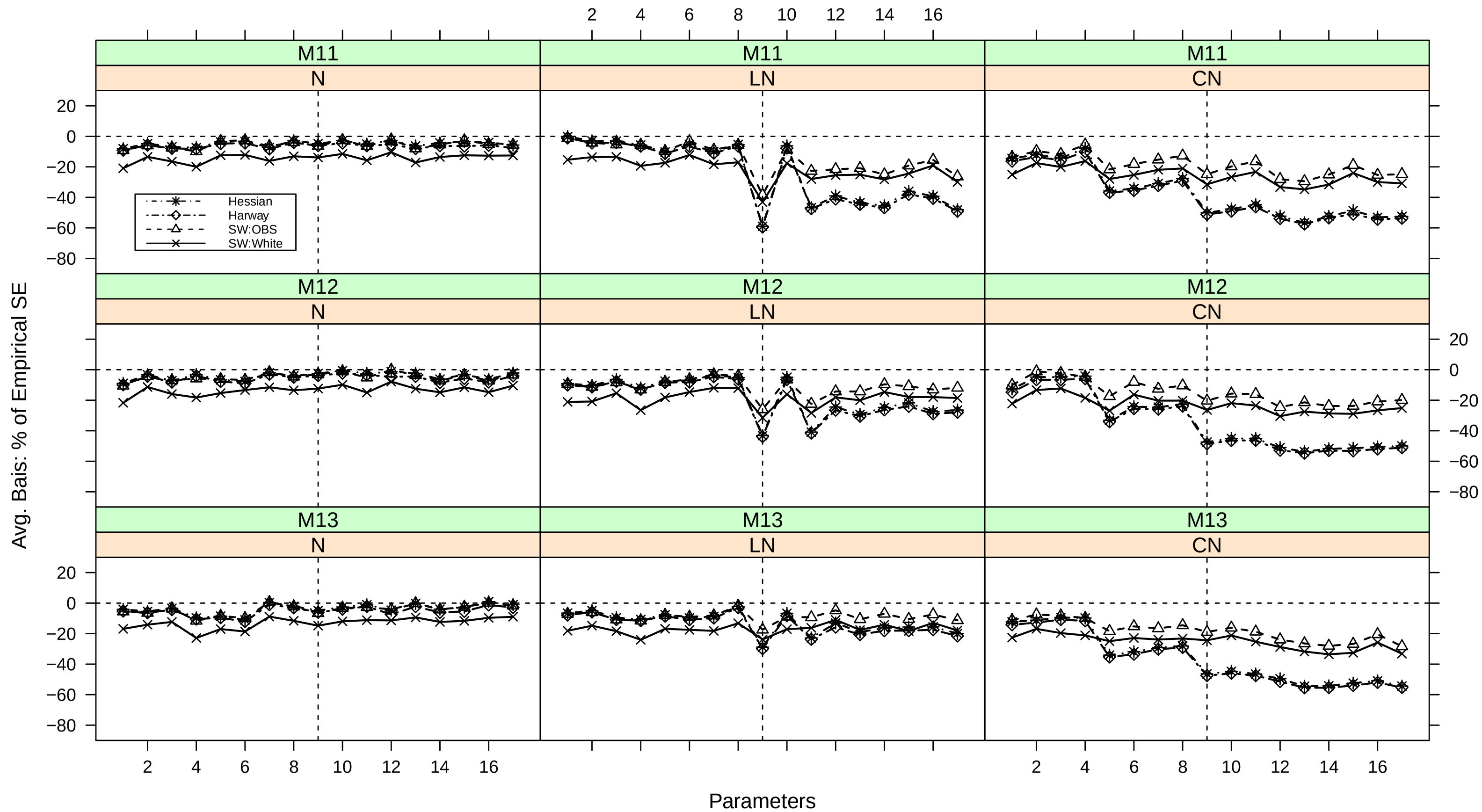

Figure D.2: Relative bias of SE estimates obtained when the stationary initial setting is used. Data are with a length of 50 and are collected from one participant. All models in the figure have high signal-to-noise ratios. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

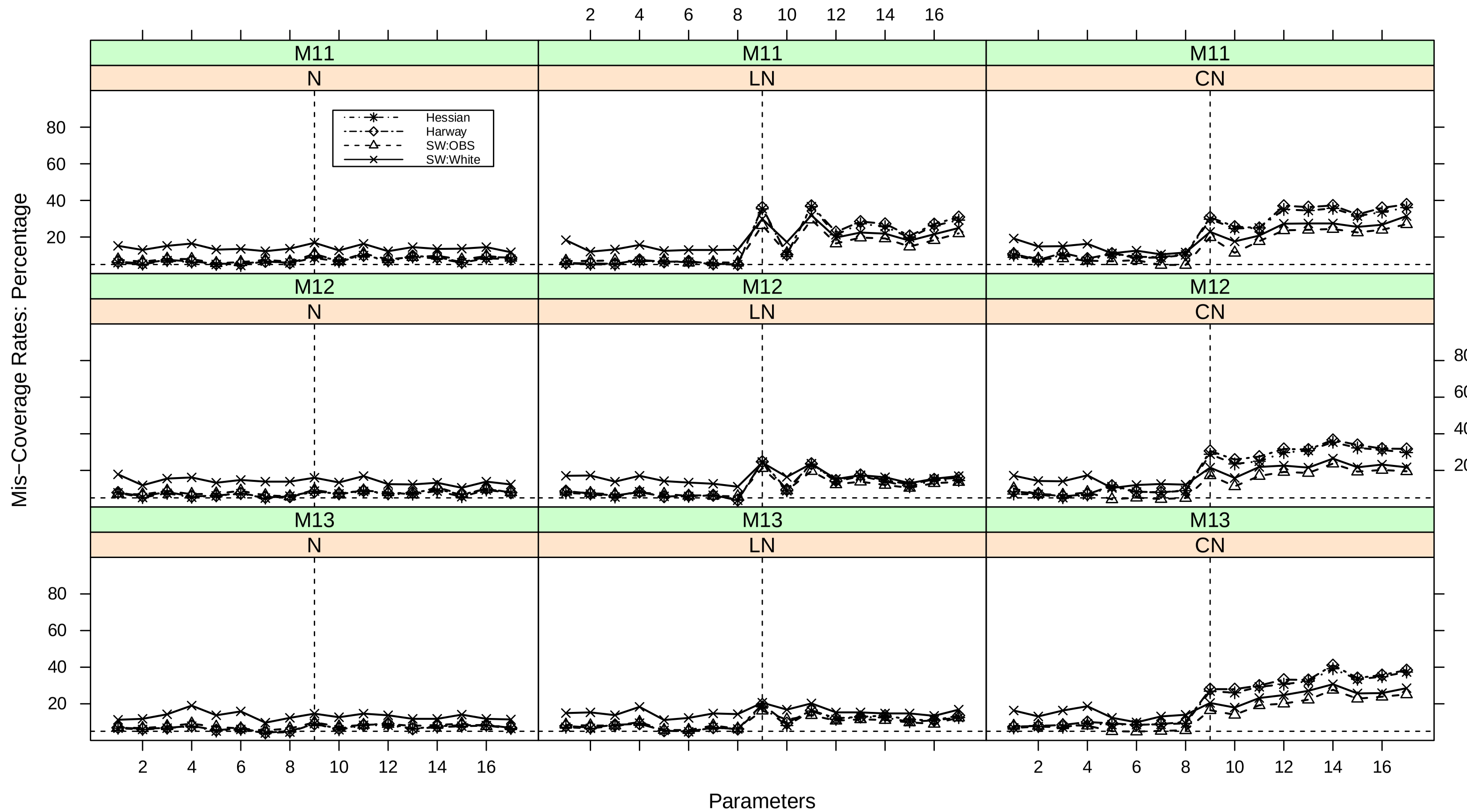

Figure D.3: Mis-coverage rates obtained when the stationary initial setting is used. Data are with a length of 50 and are collected from one participant. All models in the figure have high signal-to-noise ratios. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

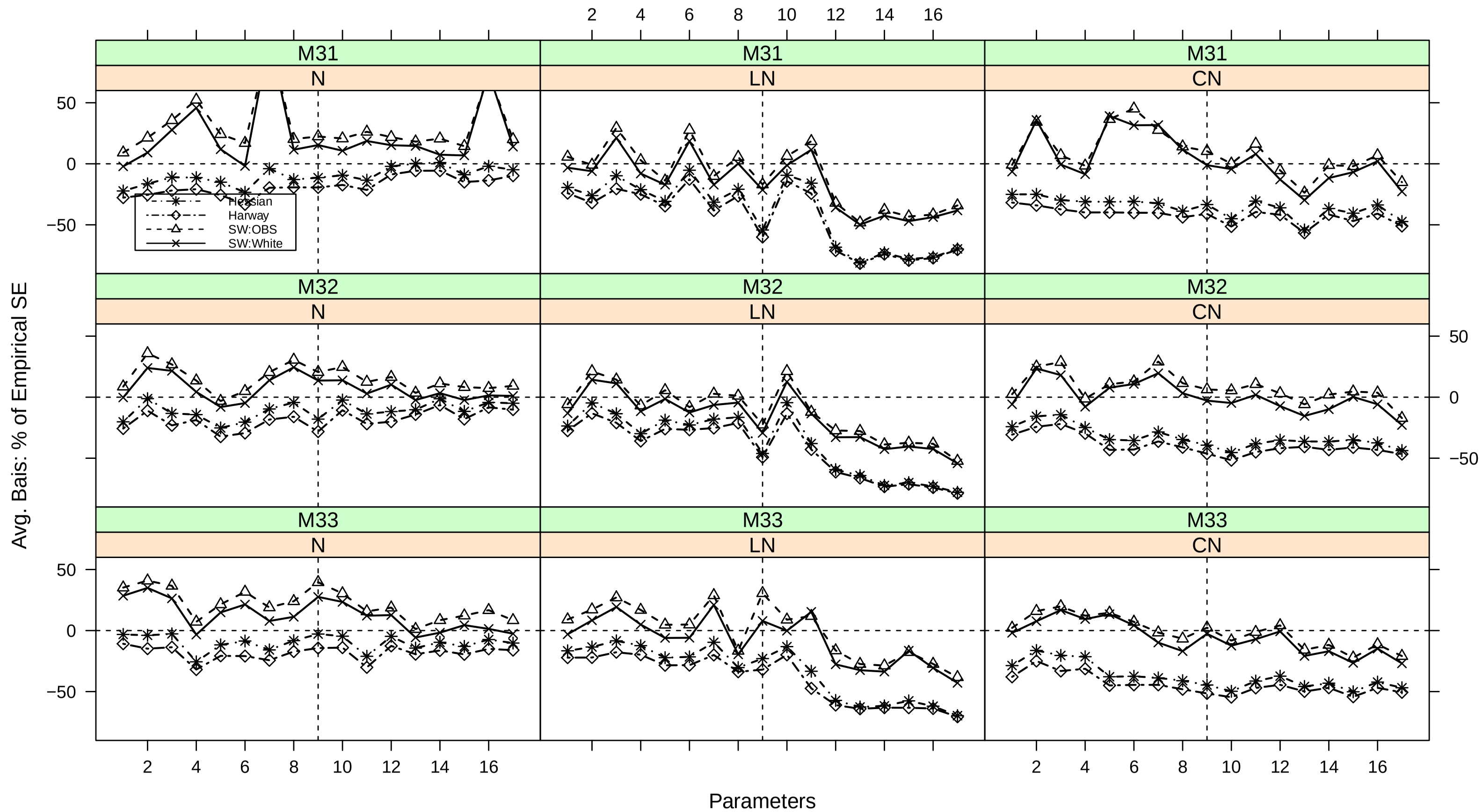

Figure D.4: Relative bias of SE estimates obtained when the stationary initial setting is used. Data are with a length of 50 and are collected from one participant. All models in the figure have low signal-to-noise ratios. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

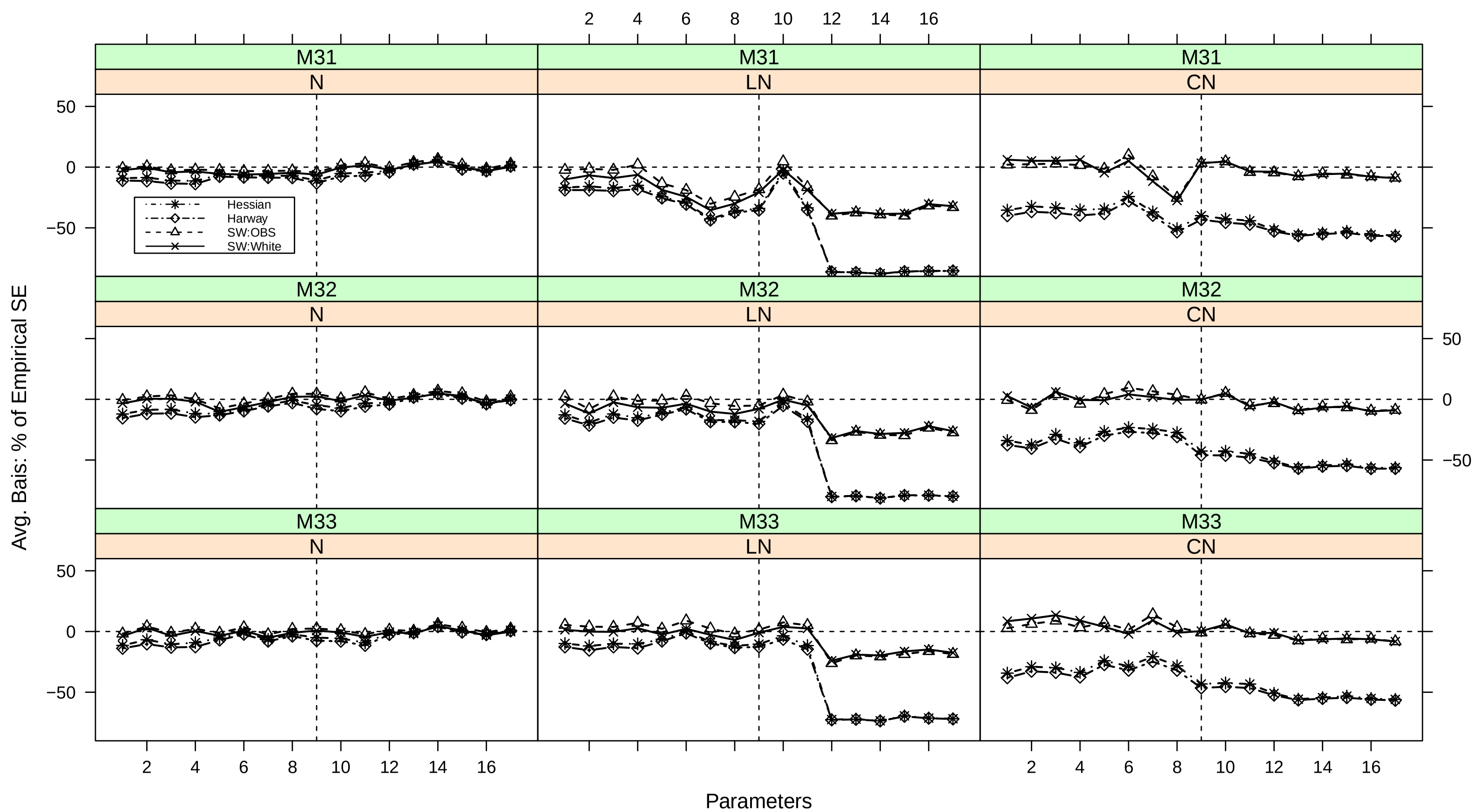

Figure D.5: Relative bias of SE estimates obtained when the stationary initial setting is used. Data are with a length of 500 and are collected from one participant. All models in the figure have low signal-to-noise ratios. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

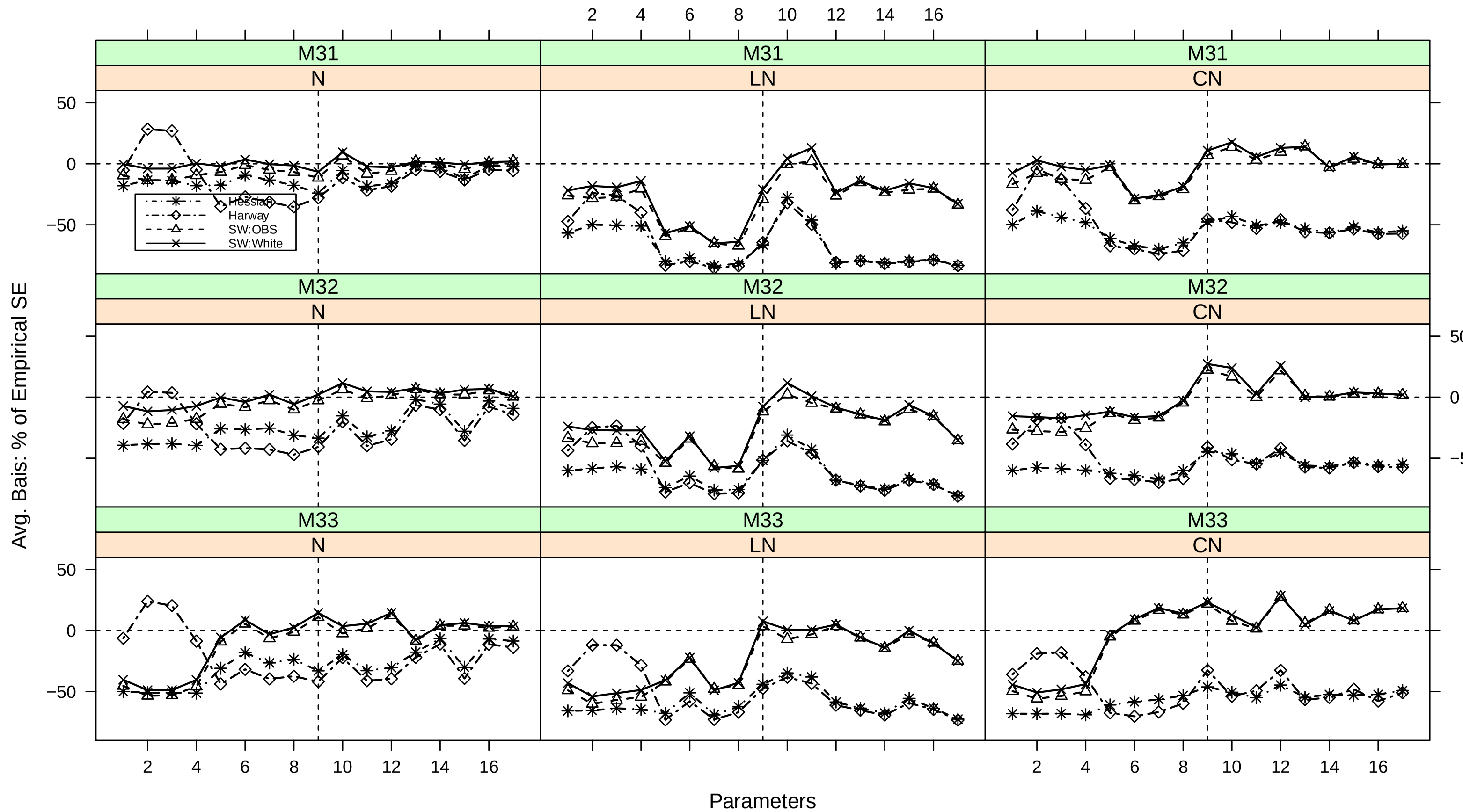

Figure D.6: Relative bias of SE estimates obtained when the stationary initial setting is used. Data are with a length of 10 and are collected from 50 participant. All models in the figure have low signal-to-noise ratios. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

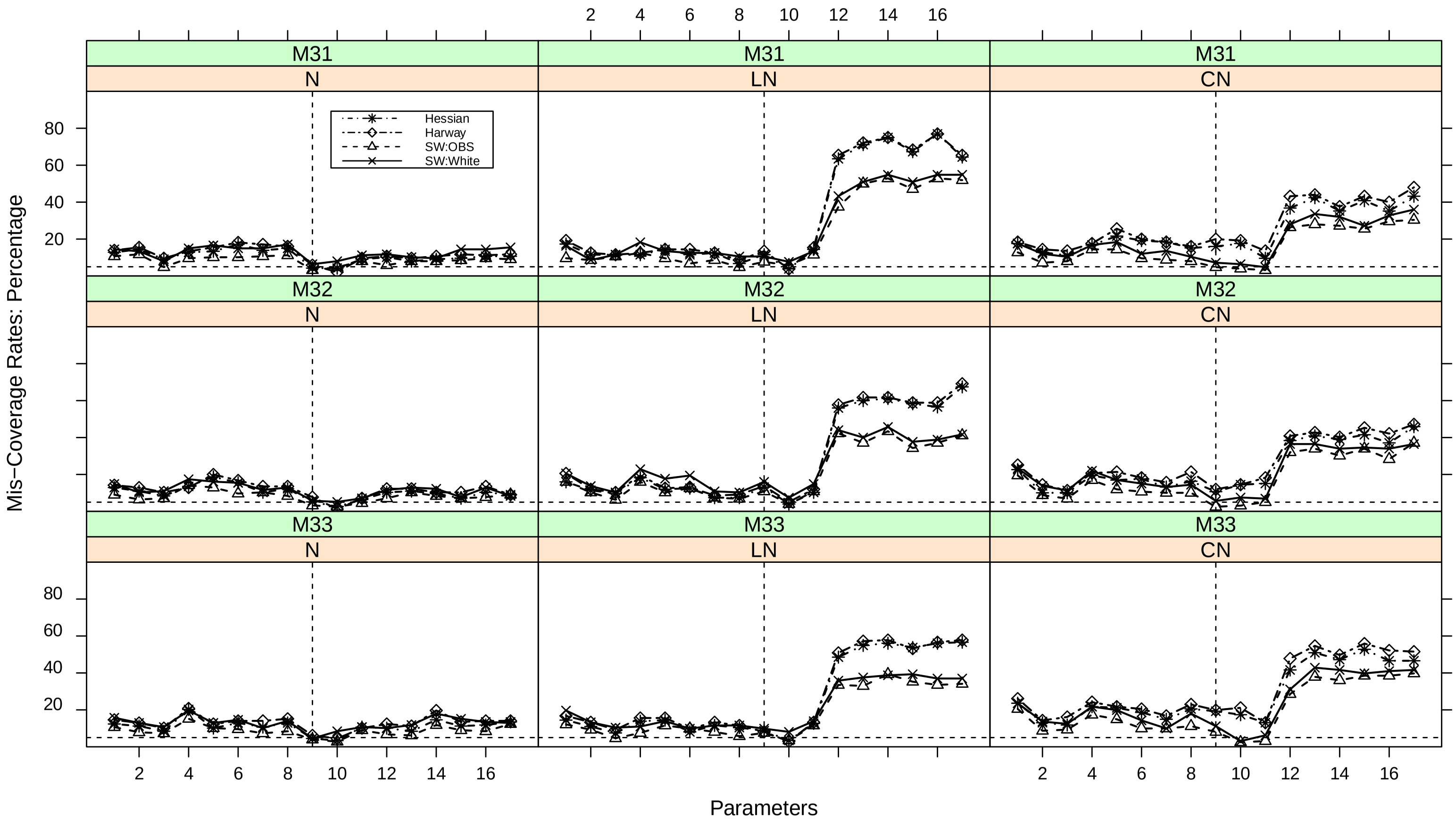

Figure D.7: Mis-coverage rates obtained when the stationary initial setting is used. Data are with a length of 50 and are collected from one participant. All models in the figure have low signal-to-noise ratios. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

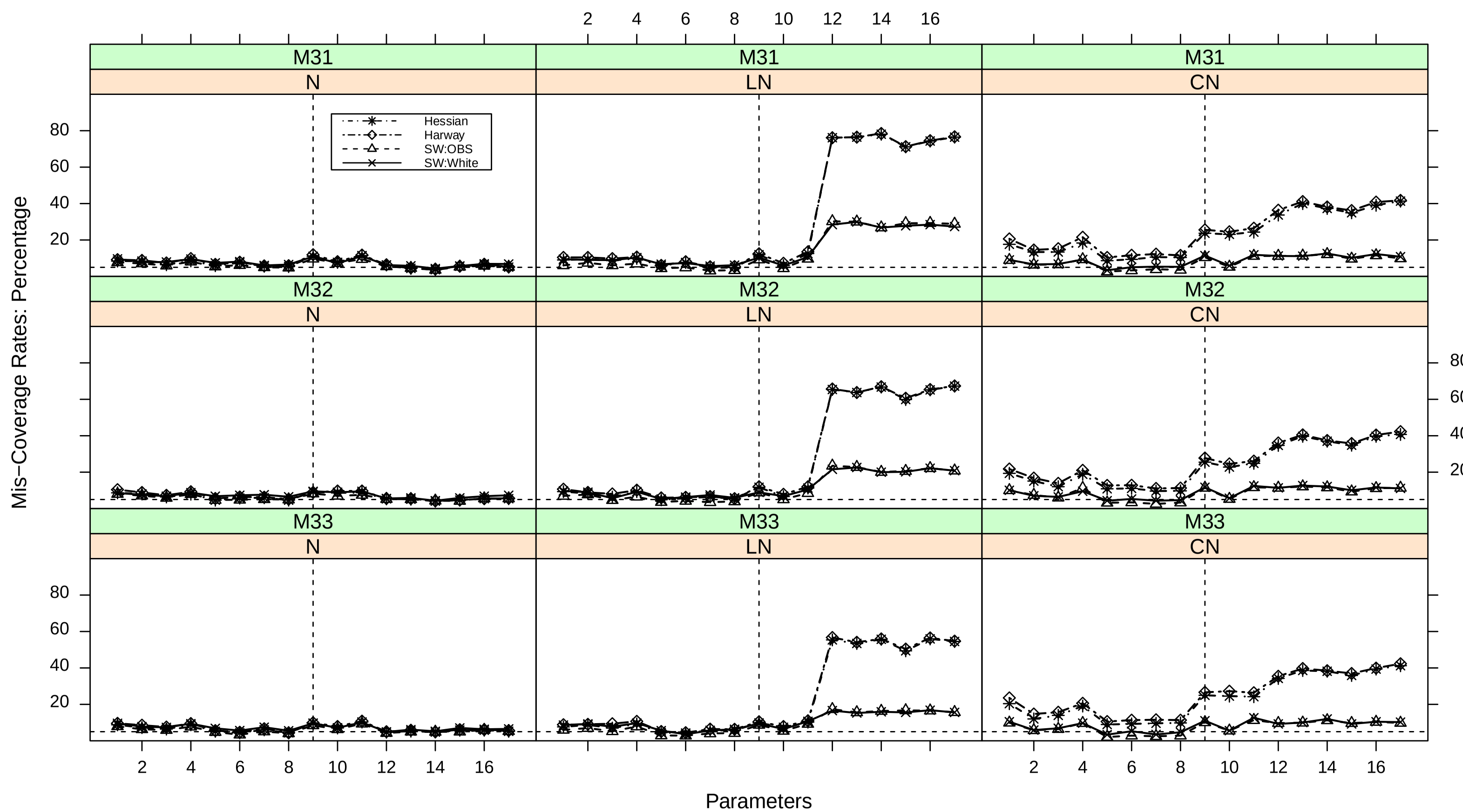

Figure D.8: Mis-coverage rates obtained when the stationary initial setting is used. Data are with a length of 500 and are collected from one participant. All models in the figure have low signal-to-noise ratios. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

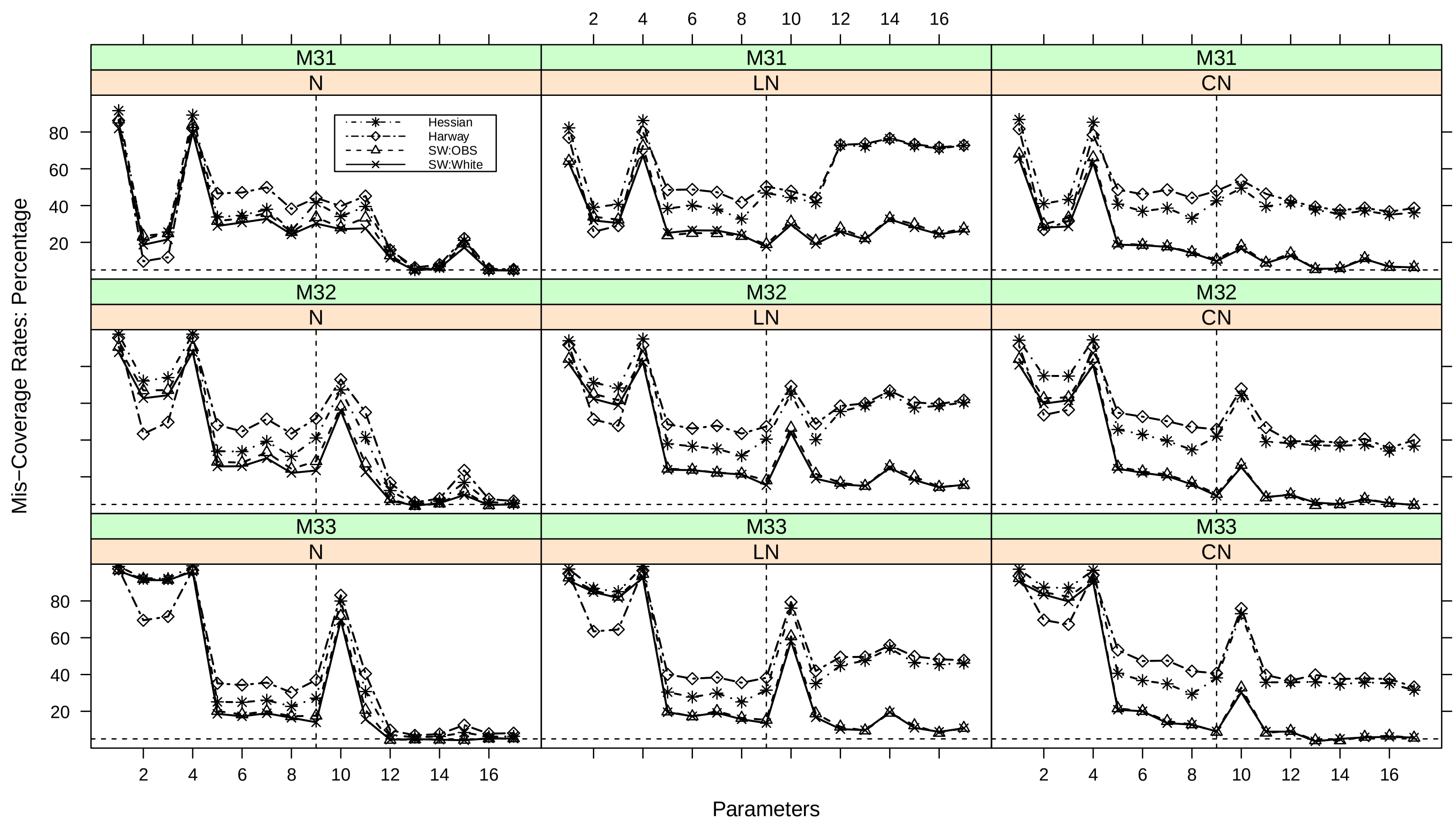

Figure D.9: Mis-coverage rates obtained when the stationary initial setting is used. Data are with a length of 10 and are collected from 50 participant. All models in the figure have high signal-to-noise ratios. The four SE estimators considered are the observed information SE estimator (Hessian or $\widehat{SE}_O$), the Harvey's SE estimator (Harvey or $\widehat{SE}_H$), Papanastassiou's sandwich SE estimator (SW:OBS or $\widehat{SE}_{SW_{P1}}$) and White's sandwich SE estimator (SW:White or $\widehat{SE}_{SW_W}$). Parameters are arranged as below: 1$\mathbf{\Phi}_{11}$, 2$\mathbf{\Phi}_{12}$,3$\mathbf{\Phi}_{21}$,4$\mathbf{\Phi}_{22}$, 5$\mathbf{A}_{21}$, 6$\mathbf{A}_{31}$, 7$\mathbf{A}_{52}$, 8$\mathbf{A}_{62}$, 9$\mathbf{Q}_{11}$, 10$\mathbf{Q}_{21}$, 11$\mathbf{Q}_{22}$, 12$\mathbf{R}_{11}$, 13$\mathbf{R}_{22}$, 14$\mathbf{R}_{33}$, 15$\mathbf{R}_{44}$, 16$\mathbf{R}_{55}$, 17$\mathbf{R}_{66}$.

APPENDIX E

DIFFERENCES BETWEEN PAPANASTASSIOU'S SANDWICH SE ESTIMATOR AND WHITE'S SANDWICH SE ESTIMATOR

Papanastassiou's sandwich SE estimator and White's sandwich SE estimator differ in whether or not the estimates of the lagged auto- and cross-covariance matrices, $\text{Cov}\left(\dot{l}_t, \dot{l}'_s\right)$ $(t \neq s)$, are included. As shown in subsection 7.2.1, the approximate lagged autocorrelations of $\left\{\dot{l}_t\right\}$ are close to zero. This suggests that the diagonal elements of $\text{Cov}\left(\dot{l}_t, \dot{l}'_s\right)$ are close to zero or are ignorable when $t \neq s$. In this appendix, we further evaluate the off-diagonal elements of $\text{Cov}\left(\dot{l}_t, \dot{l}'_s\right)$ numerically. Specifically, we show that when the true parameter values are known, the approximate lagged cross-correlations of $\left\{\dot{l}_t\right\}$ are also negligible.

The log-likelihood function at time t of a stationary state space model is given by $-2l_t = \log|\mathbf{\Sigma}_t| + \mathbf{e}'_t\mathbf{\Sigma}_t^{-1}\mathbf{e}_t$ where $\mathbf{e}_t$ and $\mathbf{\Sigma}_t$ are defined in section 3.2. Thus the first-order derivative of the log-likelihood function at time t is

$$\dot{l}'_t = \frac{\partial l}{\partial \theta'} = -.5\left[\left(\sigma_t^{-1}\right)'\dot{\sigma}_t + 2\mathbf{e}'_t\mathbf{\Sigma}_t^{-1}\dot{\mathbf{e}}_t - \mathbf{s}'_t\left(\mathbf{\Sigma}_t^{-1}\otimes\mathbf{\Sigma}_t^{-1}\right)\dot{\sigma}_t\right] \tag{E.1}$$

where $\dot{\sigma}_t = \frac{\partial \text{vec}(\mathbf{\Sigma}_t)}{\partial \theta'}$, $\dot{\mathbf{e}}_t = \frac{\partial \mathbf{e}_t}{\partial \theta'}$, $\sigma_t = \text{vec}\left(\mathbf{\Sigma}_t\right)$ and $\mathbf{s}_t = \text{vec}\left(\mathbf{e}_t\mathbf{e}'_t\right)$. Since the expectation of $\dot{l}_t$ is zero, $\text{Cov}\left(\dot{l}_t, \dot{l}'_s\right) = E\dot{l}_t\dot{l}'_s$. Therefore, we only need to evaluate $E\dot{l}_t\dot{l}'_s$. To see why the expectation of $\dot{l}_t$ equals zero, we first express $\dot{l}'_t$ as

$$E\dot{l}'_t = -.5\left[\left(\sigma_t^{-1}\right)'\dot{\sigma}_t + 2E\left(\mathbf{e}'_t\mathbf{\Sigma}_t^{-1}\dot{\mathbf{e}}_t\right) - E\left(\mathbf{s}'_t\right)\left(\mathbf{\Sigma}_t^{-1}\otimes\mathbf{\Sigma}_t^{-1}\right)\dot{\sigma}_t\right].$$

We known that $E\mathbf{s}_t = \text{vec}\left(E\left(\mathbf{e}_t\mathbf{e}_t'\right)\right)$ and $E\left(\mathbf{e}_t\mathbf{e}_t'\right) = \mathbf{A}E\left[\left(\mathbf{x}_t - \mathbf{x}_t^{t-1}\right)\left(\mathbf{x}_t - \mathbf{x}_t^{t-1}\right)'\right]\mathbf{A}'+$ $\mathbf{R}$. Because $\mathbf{P}_t^{t-1}$ is the unconditional error variance of $\mathbf{x}_t^{t-1}$(Anderson & Moore, 1979), $\mathbf{P}_t^{t-1} = E\left[\left(\mathbf{x}_t - \mathbf{x}_t^{t-1}\right)\left(\mathbf{x}_t - \mathbf{x}_t^{t-1}\right)'\right]$. And thus $E\left(\mathbf{e}_t\mathbf{e}_t'\right)$=$\mathbf{A}\mathbf{P}_t^{t-1}\mathbf{A}' + \mathbf{R}$= $\mathbf{\Sigma}_t$ and $E\mathbf{s}_t = \sigma_t$. Hence the expectation of $\dot{l}_t$ can be simplified as

$$\begin{aligned} E\dot{l}_t &= -.5\left[\left(\sigma_t^{-1}\right)'\dot{\sigma}_t + 2E\left(\mathbf{e}_t'\mathbf{\Sigma}_t^{-1}\dot{\mathbf{e}}_t\right) - \sigma_t'\left(\mathbf{\Sigma}_t^{-1}\otimes\mathbf{\Sigma}_t^{-1}\right)\dot{\sigma}_t\right] \\ &= -E\left(\mathbf{e}_t'\mathbf{\Sigma}_t^{-1}\dot{\mathbf{e}}_t\right). \end{aligned}$$

To see why $E\left(\mathbf{e}_t'\mathbf{\Sigma}_t^{-1}\dot{\mathbf{e}}_t\right) = 0$, we can examine the derivatives, $\dot{\mathbf{e}}_t$, one by one, that is, examining $E\left(\mathbf{e}_t'\mathbf{\Sigma}_t^{-1}\frac{\partial\mathbf{e}_t}{\partial\theta_i}\right)$ $(i = 1,..., n.p)$. Note $\frac{\partial\mathbf{e}_t}{\partial\theta_i} = \frac{\partial}{\partial\theta_i}\left(\mathbf{y}_t - \mathbf{A}\mathbf{x}_t^{t-1}\right)$= $-\frac{\partial\mathbf{A}\mathbf{x}_t^{t-1}}{\partial\theta_i}$. It is a linear function of past observations, $\mathbf{Y}_{t-1} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1}\}$.

$$\begin{aligned} E\left(\mathbf{e}_t'\mathbf{\Sigma}_t^{-1}\frac{\partial\mathbf{e}_t}{\partial\theta_i}\right) &= \text{tr}\left(\mathbf{\Sigma}_t^{-1}E\left(\frac{\partial\mathbf{e}_t}{\partial\theta_i}\mathbf{e}_t'\right)\right) \\ &= -\text{tr}\left(\mathbf{\Sigma}_t^{-1}E\left(\frac{\partial\mathbf{A}\mathbf{x}_t^{t-1}}{\partial\theta_i}\mathbf{e}_t'\right)\right) \\ &= -\text{tr}\left(\mathbf{\Sigma}_t^{-1}E\left[\frac{\partial\mathbf{A}\mathbf{x}_t^{t-1}}{\partial\theta_i}\left(\mathbf{y}_t - \mathbf{A}\mathbf{x}_t^{t-1}\right)'\right]\right). \end{aligned}$$

Since the Kalman filter estimator, $\mathbf{A}\mathbf{x}_t^{t-1}$, is the best linear estimator of $\mathbf{y}_t$, the prediction error $\mathbf{e}_t$ is orthogonal to the space spanned by $\mathbf{Y}_{t-1}$ (Anderson & Moore, 1979, see Chapter 5). Thus $E\left[\frac{\partial\mathbf{A}\mathbf{x}_t^{t-1}}{\partial\theta_i}\left(\mathbf{y}_t - \mathbf{A}\mathbf{x}_t^{t-1}\right)'\right] = 0$. And as a result, $E\left(\mathbf{e}_t'\mathbf{\Sigma}_t^{-1}\frac{\partial\mathbf{e}_t}{\partial\theta_i}\right) =$ 0. This can be applied to the derivative with respect to any parameter. Thus, $E\left(\mathbf{e}_t'\mathbf{\Sigma}_t^{-1}\dot{\mathbf{e}}_t\right) = 0$. This immediately follows that,

$$\begin{aligned} E\dot{l}_t &= -E\left(\mathbf{e}_t'\mathbf{\Sigma}_t^{-1}\dot{\mathbf{e}}_t\right) \\ &= 0. \end{aligned}$$

In short, we only need to evaluate $E\dot{l}_t\dot{l}_s'$ for $t \neq s$.

Since the expectation operation involved in $E\dot{l}_t\dot{l}_s'$ is difficult to compute, we use an average to approximate the expectation operation. Specifically, we use the approximate auto- and cross-covariance matrix, $\frac{1}{N_{sim}T}\sum_{j=1}^{N_{sim}}\sum_{k=1}^{T-h}\dot{l}_{k,j}\dot{l}_{k+h,j}'$, where $\dot{l}_{k,j}$ is the first derivative at time k for the jth sample and $h = t - s$. (E.1) is used to compute $\dot{l}_t$. In (E.1), $\mathbf{e}_t$ and $\mathbf{\Sigma}_t$ can be computed using the Kalman filter algorithm (see section 3.2). The following algorithm can be used to compute $\dot{\mathbf{e}}_t$ and $\dot{\sigma}_t$ for $t = 1, \ldots, T$,

$$\begin{aligned}
\dot{\mathbf{x}}_t^{t-1} &= \left(\mathbf{x}_{t-1}^{t-1\prime} \otimes \mathbf{I}\right)\dot{\phi} + \mathbf{\Phi}\dot{\mathbf{x}}_{t-1}^{t-1} \\
\dot{\mathbf{e}}_t &= -\left(\mathbf{x}_t^{t-1\prime} \otimes \mathbf{I}\right)\dot{\mathbf{a}} - \mathbf{A}\dot{\mathbf{x}}_t^{t-1} \\
\dot{\mathbf{p}}_t^{t-1} &= \left(\mathbf{I} + \mathbf{C}_{p,p}\right)\left[\left(\mathbf{\Phi}\mathbf{P}_t^{t-1}\right) \otimes \mathbf{I}\right]\dot{\phi} + \left(\mathbf{\Phi} \otimes \mathbf{\Phi}\right)\dot{\mathbf{p}}_t^{t-1} + \dot{\mathbf{q}} \\
\dot{\sigma}_t &= \left(\mathbf{I} + \mathbf{C}_{k,k}\right)\left[\left(\mathbf{A}\mathbf{P}_t^{t-1}\right) \otimes \mathbf{I}\right]\dot{\mathbf{a}} + \left(\mathbf{A} \otimes \mathbf{A}\right)\dot{\mathbf{p}}_t^{t-1} + \dot{\mathbf{r}} \\
\dot{\mathbf{k}} &= \left[\left(\mathbf{\Sigma}_t^{-1}\mathbf{A}\right) \otimes \mathbf{I}\right]\dot{\mathbf{p}}_t^{t-1} + \left(\mathbf{\Sigma}_t^{-1} \otimes \mathbf{P}_t^{t-1}\right)\mathbf{C}_{k,p}\dot{\mathbf{a}} - \left(\mathbf{\Sigma}_t^{-1} \otimes \mathbf{K}_t\right)\dot{\sigma}_t \\
\dot{\mathbf{x}}_{t-1}^{t-1} &= \dot{\mathbf{x}}_t^{t-1} + \left(\dot{\mathbf{e}}_t' \otimes \mathbf{I}\right)\dot{\mathbf{k}}_t + \mathbf{K}_t\dot{\mathbf{e}}_t \\
\dot{\mathbf{p}}_t^{t} &= \dot{\mathbf{p}}_t^{t-1} - \left[\left(\mathbf{P}_t^{t-1}\mathbf{A}'\right) \otimes \mathbf{I}\right]\dot{\mathbf{k}} - \left(\mathbf{P}_t^{t-1} \otimes \mathbf{K}_t\right)\dot{\mathbf{a}} - \left[\mathbf{I} \otimes \left(\mathbf{K}_t\mathbf{A}\right)\right]\dot{\mathbf{p}}_t^{t-1}
\end{aligned}$$

where $\mathbf{C}_{k,p}$ is the commutation matrix for any $k \times p$ matrix $\mathbf{M}$ and for any vector $\mathbf{m}$ or matrix $\mathbf{M}$,$\dot{\mathbf{m}}$ denotes the first-order derivative of $\mathbf{m}$ with respect to the parameter vector, θ, that is, $\dot{\mathbf{m}} = \frac{\partial \mathbf{m}}{\partial \theta'}$ (or $\dot{\mathbf{m}} = \frac{\partial \text{vec}(\mathbf{M})}{\partial \theta'}$ when taking the first-order derivative of a matrix $\mathbf{M}$). If the stationary initial setting is used, the initial factor covariance matrix is defined as $\text{vec}\left(\mathbf{P}_0\right) = \left[\mathbf{I} - \mathbf{\Phi} \otimes \mathbf{\Phi}\right]^{-1}\text{vec}\left(\mathbf{Q}\right)$. Then,

$$\begin{aligned}
\dot{\mathbf{p}}_0 &= \left[\mathbf{I} - \mathbf{\Phi} \otimes \mathbf{\Phi}\right]^{-1}\dot{\mathbf{q}} \\
&\quad + \left\{\left[\left[\mathbf{I} - \mathbf{\Phi} \otimes \mathbf{\Phi}\right]^{-1}\text{vec}\left(\mathbf{Q}\right)\right]' \otimes \left[\mathbf{I} - \mathbf{\Phi} \otimes \mathbf{\Phi}\right]^{-1}\right\} \\
&\quad \times \left(\mathbf{I} \otimes \mathbf{C}_p \otimes \mathbf{I}\right)\left(\mathbf{I}_{p^4} + \mathbf{C}_{p^2}\right)\left(\mathbf{I}_{p^2} \otimes \text{vec}\left(\mathbf{\Phi}\right)\right)\dot{\phi}.
\end{aligned}$$

Thus with true parameter values and a given sample, we can approximate the auto- and cross-covariance matrix via the aforementioned algorithm and the Kalman filter algorithm. Since the meaning of a covariance matrix depends on its measurement scale, we display the results of correlation matrix. Specifically, the auto- and cross-correlation matrix is computed by,

$$\widehat{\text{Cor}}\left(\dot{l}_t, \dot{l}'_{t+h}\right) = \frac{1}{N_{sim}\left(T-h\right)} \sum_{j=1}^{N_{sim}} \sum_{k=1}^{T-h} \mathbf{D}_{\mathbf{S}}^{-1/2} \dot{l}_{k,j} \dot{l}'_{k+h,j} \mathbf{D}_{\mathbf{S}}^{-1/2}$$

where $\mathbf{D}_{\mathbf{S}}$ is a diagonal matrix whose diagonal elements are the same as $\mathbf{S} = \frac{1}{N_{sim}T} \sum_{j=1}^{N_{sim}} \sum_{k=1}^{T} \dot{l}_{k,j} \dot{l}'_{k,j}$ and $h = s - t$. We then can obtain the approximate auto- and cross-correlation matrix, $\overline{\widehat{\text{Cor}}\left(\dot{l}_t, \dot{l}'_{t+h}\right)}$, by averaging the sample auto- and cross-correlation matrix across 1000 simulated replications.

For models considered in this thesis, there are 17 parameters. And thus $\left\{\dot{l}_t\right\}$ contains 17 univariate derivative series and this follows that there are 136=$\frac{17\times 16}{2}$ cross-correlation functions. Figure E.1 graphs all the 136 CCFs of $\left\{\dot{l}_t\right\}$. We can see that approximate cross-correlations are very close to zero when lags are greater than 1. The approximate lag-1 cross-correlations are relatively larger. The largest approximate lag-1 cross-correlation is the one between $\frac{\partial l_t}{\partial \mathbf{\Phi}_{22}}$ and $\frac{\partial l_t}{\partial \mathbf{R}_{11}}$ and it is smaller than .09.

In sum, we show numerically that the lagged auto- and cross-correlation matrices and hence the lagged auto- and cross-covariance matrices are nearly zero matrices. And thus the linear dependence among $\dot{l}_t$s is ignorable.

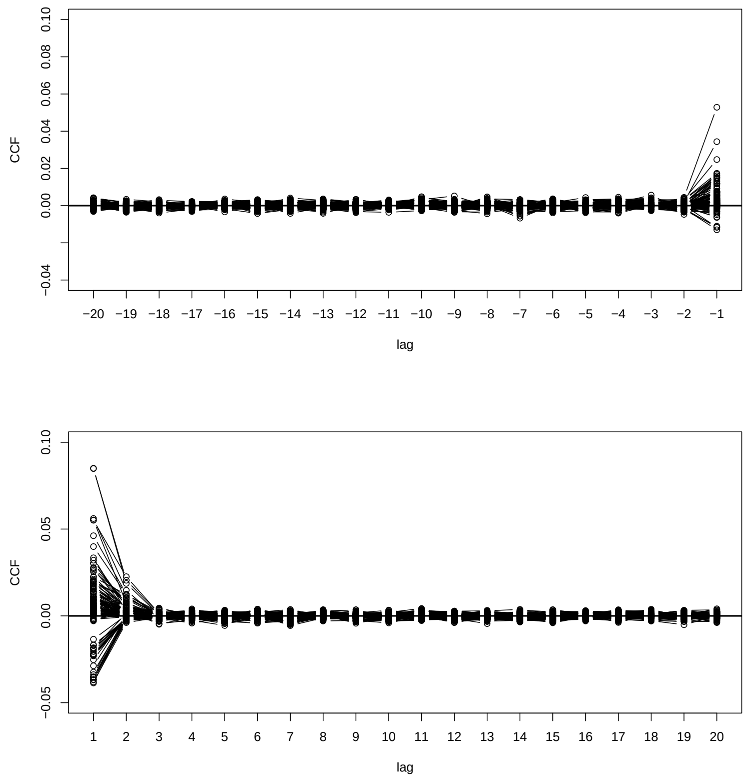

Figure E.1: All approximate CCFs of $\left\{\hat{l}_t\right\}$ when true parameters are known. Results are obtained based on samples generated from model M12 and under the $T = 500$ and the contaminated normal data condition.

BIBLIOGRAPHY

Andersen, T. G., Chung, H. J., & Sorensen, B. E. (1999). Efficient method of moments estimation of a stochastic volatility model: a monte carlo study. *Journal of Econometrics*, *91*, 61-87.

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering.* Englewood Cliffs: Prentice-Hall.

Ansley, C. F., & Kohn, R. (1985). Estimation, filtering and smoothing in state space models with incompletely sepcified initial conditions. *Annal of Statistics*, *13*, 1286-1316.

Bauwens, L., & Veredas, D. (2004). The stochastic conditional duration model: a latent variable model for the analysis of financial durations. *Journal of Econometrics*, *119*, 381-412.

Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, *34*, 181-197.

Breidt, F. J., & Carriquiry, A. L. (1996). Improved quasi-maximum likelihood estimation for stochastic volatility models. In A. Zellner & J. Lee (Eds.), *Modeling and prediction: honouring seymour geisel.* Springer.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: some promising uses and extensions of autogressive modving average time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Advances in psychometrics: a festschrift to roderick p. mcdonald* (p. 415-452). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Caines, P. E. (1988). *Linear stochastic systems*. Wiley: John Wiley & Sons.

Carlin, B. P., Polson, N. G., & Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of American Statistical Association*, *87*, 493-500.

Cattell, R. B., Cattell, A. K. S., & Rhymer, R. M. (1947). P-technique demonstrated in determining psychologphysical nonlinear dynamical systems models. *Psychometrika*, *12*, 267-288.

Cavanaugh, J. E., & Shumway, R. H. (1996). On computing the expected fisher information matrix for state-space model parameters. *Statistics and Probability Letters*, *26*, 347-355.

Chow, S. M., Ho, M. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 303-332.

Chow, S. M., Nesselroade, J. R., Shiren, K., & McArdle, J. (2004). Dynamic structure of emotions among individuals with parkinson's disease. *Structural Equation Modeling*, *11*, 560-582.

Creal, D. D. (2008). Analysis of filtering and smoothing algorithms for Levy-driven stochastic volatility models. *Computational Statistics & Data Analysis*, *52*, 2863-2876.

De Jong, P. (1988). The likelihood for a state space model. *Biometrika*, *75*, 165-169.

Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *The Annals of Statistics*, *7*, 490-506.

Dunsmuir, W., & Hannan, E. J. (1976). Vector linear time series models. *Advances in Applied Probablity*, *8*, 339-364.

Dunsmuir, W., & Hannan, E. J. (1978). Vector linear time series models: corrections and extensions. *Advances in Applied Probablity*, *10*, 360-372.

Durbin, J., & Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, *3*, 669-684.

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

Fruhwirth-Schnatter, S. (1994). Applied state space modeling of non-Gaussian time series using integration-based kalman filtering. *Statistics and Computing*, *4*, 259-269.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2002). On the nature of SEM estimates of ARMA parameters. *Structural Equation Modeling*, *9*, 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 N: raw data maximum likelihood. *Structural Equation Modeling*, *10*, 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. *Multivariate Behavioral Research*, *40*, 207-233.

Harvey, A. C. (1989). *Forecasting structural time series models and the kalman filter.* New York: Cambridge University Press.

Harvey, A. C., & Shephard, N. (1996). Estimation of asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistic*, *14*, 429-434.

Hertzog, C., Lindenberger, U., Ghisletta, P., & Oertzen, T. von. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *3*, 244-252.

Hertzog, C., Oertzen, T. von, Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 541-563.

Ho, M. R. (2003). *Modelling effective connectivity in functional magnetic resonance imaging data by state space models.* Unpublished doctoral dissertation, University of Illinoise at Urbana-Champaign.

Jacquier, E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistic*, *12*, 371-389.

Kawakatsu, H. (2007). Numerical integration-based gaussian mixture filters for maximum likelihood estimatino of asymmetric stochastic volatility models. *Econometrics*, *10*, 342-358.

Kirby, C. (2006). Linear filtering for asymmetric stochastic volatility models. *Economics Letters*, *92*, 284-292.

Kitagawa, G. (1987). Non-Gaussian state space modeling of nonstationary time series. *Journal of the American Statistical Association*, *82*, 1032-1042.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Berlin: Springer.

MacCallum, R., Widaman, K., Preacher, K. J., & Hong, S.-H. (2011). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*, 611-637.

MacCallum, R., Widaman, K., Zhang, K.-B., & Hong, S.-H. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99.

Meng, X., & Rubin, D. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM lagorithm. *Journal of the American Statistical Association*, *86*, 899-909.

Micceri. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156-166.

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

Molenaar, P. C. M., & Nesselroade, J. R. (1998). A comparison of psudo-maximum likelihood and asymptotically distribution-free dynamic factor analysis parameter estimation in fitting covariance-structure models to block-toeplitz matrices representing single-subject multivariate time series. *Multivariate Behavioral Research*, *33*, 313-342.

Musher, E., Nesselroade, J., & Schmitz, B. (2002). Perceived control and academic performance: a comparison of high-and low-performing children on within-person change patterns. *International Journal of Behavior Development*, *27*, 49-63.

Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. (2002). Alternative dynamic factor models for multivariate time series analyses. In D. M. Moscowitz & S. L. Hersh-berger (Eds.), *Modeling intraindividual variability with repeated measures data: Advances and techniques* (p. 235-265). Mahwah, NJ: Lawrence Erlbaum Associates.

Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time-series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (p. 224-251). Newbury Park, CA: Sage.

Nesselroade, J. R., & Paolo, G. (2003). Structuring and measuring change over the life span. In U. Lindenberger U. & Staudinger (Ed.), *Understanding human developments: dialogues with lifespan psychology* (p. 317-337). Kluwer acadeimc publishers.

Papanastassiou, D. (2006). Computing the covariance matrix of QML estimators for a state space model. *Statistics and Probability Letters*, *76*, 1001-1006.

Parkinson, A., Laura, C., & Jackson, S. R. (2010). Parietal cortex coding of limb posture: in search of the body schema. *Neuropsychologia*, *48*, 3228-3234.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, *63*, 289-306.

Sandmann, G., & Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, *87*, 271-301.

Shumway, R. H., & Stoffer, D. S. (2004). *Time series analysis and its applications.* New York: Springer-Verlag.

Smith, J. F., Pillai, A., Chen, K., & Horwitz, B. (2010). Identification and validation of effective connectivity networks in functional magnetic responance imaging using switching linear dynamic systems. *NeuroImage*, *52*, 1027-1040.

Smith, J. F., Pillai, A., Chen, K., & Horwitz, B. (2012). Effective connectivity modeling for fmri: six issues and possible solutions using linear dynamic systems. *Frontiers In Systems Neurosicence*, *5*, 1-17.

Song, H., & Ferrer, E. (2009). State-space modeling of dynamic psychological processes via the kalman smoother algorithm: rationale, finite sample properties, and applications. *Structural Equation Modeling*, *16*, 338-363.

Strickland, C. M., Forbes, C. S., & Martin, G. M. (2006). Bayesian analysis of the stochastic conditional duration model. *Computational Statistics & Data Analysis*, *50*, 2247-2267.

Tarvainen, M. P., Georgiadis, S. D., Ranta-aho, P. O., & Karjalainen, P. (2006). Time-varying analysis of heart rate variability signals with a kalman smoother algorithm. *Physiological Measurement*, *27*, 225-239.

White, H. (1984). Maximum-likelihood estimation of misspecified dynmaic-models. *Lecture Notes in Economics and Mathematical Systems*, *237*, 1-19.

Yang, M.-S., & Chow, S.-M. (2010). Using state-space models with regime switching to represent the dynamics of facial electromyography (emg) data. *Psychometrika: Application and Case Studies*, *74*, 744-771.

Yuan, K.-H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, *51*, 289-309.

Yuan, K.-H., & Hayashi, K. (2006). Standard errors in covariance structure models: asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology*, *59*, 397-417.

Zarchan, P., & Musoff, H. (2000). *Fundamentals of Kalman filtering: a practical approach.* VA: American institute of aeronautics and astronautics, Inc.

Zhang, G., & Browne, M. W. (2010). Bootstrap standard error estimates in dynamic factor analysis. *Multivariate Behavioral Research*, *45*, 453-482.

Zhang, G., & Chow, S. (2010). Standard error estimation in stationary multivariate time series models using residual-based bootstrap procedures. In P. Molenaar & K. Newell (Eds.), *Pathway to individual change: statistical models for analyzing learning and development.* American Psychological Association.

Zhang, G., Chow, S. M., & Ong, A. D. (2011). A sandwich-type standard error estimator of SEM models with multivariate time series. *Psychometrika*, *76*, 77-96.

Zhang, Z., Hamaker, E. L., & Nesselroade, J. R. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 377-402.