STRUCTURAL EQUATION MODELS AND SMALL SAMPLE BIAS REDUCTION WITH APPLICATION TO FISHERY DATA

NOEL G. CADIGAN

ISBN 0-315-86651-9

STRUCTURAL EQUATION MODELS
and small sample bias reduction
with application to fishery data

by

Noel G. Cadigan

A thesis submitted to the school of Graduate
Studies in partial fulfilment of the
requirements for the degree of
Master of Applied Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

1993

St. John's Newfoundland

Abstract

An overview of structural equation models is presented along with an application to fishery data involving estimation and significance testing of the density dependent component of recruitment in 6 cod populations. Estimates and standard errors are based on normal theory and large sample properties of maximum likelihood estimates. The data sets analyzed involve small sample sizes so a sensitivity analysis of the effect (in terms of bias) of small sample sizes and other deviations in model assumptions is conducted. The analysis indicates that sample size is the most influential factor considered on the bias of parameter estimates. The reliability of indicator variables is also important.

Two methods of reducing the bias in estimates are considered, they are the jackknife and a method based on a Taylor's series expansion of the loglikelihood function. The bias reduced estimators are investigated by simulating several confirmatory factor models. Neither the jackknife nor the Taylor's series biased reduced estimator works sufficiently well to warrant their application in practise. Both estimators consistently reduce bias in the maximum likelihood estimates only when little bias exists. A difficulty realized in the investigation is that the expectations of some estimators are unbounded and this makes bias reduction difficult.

Contents

List of Tables

List of Figures

Acknowledgements

I wish to thank my supervisor, Dr. U. Balasooriya (Memorial University of Newfoundland), Dr. R. Bartlett (Memorial University of Newfoundland) and Dr. R. Myers (Department of Fisheries and Oceans) for helpful comments and suggestions about my practicum. I acknowledge the Department of Fisheries and Oceans, Government of Canada, for financial support.

Chapter 1

Covariance analysis of structural equation models

1.1 Introduction

The analysis of covariance with structural equations, also referred to as latent variable analysis, is a relatively new area of research in statistics and is popular in econometrics and psychometrics. Structural equation models are becoming common analytical tools for studying linear relationships among variables. Austin and Wolfle (1991) present a detailed bibliography of technical work related to this subjected. Bollen (1989) gives a short history on the development of structural equation models and also the essential information required to perform such an analysis. The LISREL (Jöreskog 1973) formulation of a structural equation model is used here but others are available (e.g. Bentler 1989).

A structural equation model is usually a linear model relating measured traits and unmeasurable concepts or latent variables. Bollen (1989) defines latent variables as "representations of concepts in measurement models". Common examples are socio-economic status, ambition, IQ, etc. The recruitment or the reproduction of a fish stock is a less abstract measure considered in the following chapter. A simple structural equation model might, for ex-

ample, relate ambition (unobservable) to the variables educational status and hours worked per week (observable). Estimation is in terms of the sample covariance matrix of the observed variables and the population covariance matrix generated by the structural equation model.

The notation used is first presented before constructing modeling techniques. Notation is consistent with LISREL and otherwise with Bollen's (1989) book where possible. Observable and unobservable (latent) scalar random variables are denoted by lower case Roman and Greek letters respectively. Vector and matrix equivalents of the above are in lower and upper case boldface. Parameters are denoted the same as latent variables; data and other constants have the same notation as observable random variables and the distinctions are made clear when confusion exists.

1.2 Structural equation model

Structural equation models typically deal only with the analysis of covariance structures. Mean effects are removed by analyzing the sample covariance matrix, hence, it is assumed that all data considered has been standardized by subtracting means (unless otherwise stated).

A simple example is considered to help fix ideas; it is from part of a model of political democracy and industrialization for developing countries in Bollen (1989). Consider the variables:

y_1 - freedom of press,

y_2 - freedom of group opposition,

x_1 - gross national product per capita, and

x_2 - energy consumption per capita.

The x and y variables measure industrialization and political freedom and they are measured for many countries. A model relating political freedom and industrialization is to be developed.

Variables in structural equation models are classified as either exogenous or endogenous variables. The cause of exogenous variables (similar to independent variables in a regression analysis) is not considered in the structural equation model while the cause for endogenous variables (similar to dependent variables) is. For the example, if industrialization is thought to cause political freedom in the model then political freedom is an endogenous variable and industrialization is an exogenous variable. Exogenous variables are denoted as ξ and endogenous variables are denoted as η.

A structural equation model consists of a measurement model and a latent variable model. The measurement model is a relationship between the observed and latent variables. The measurement model is the same as a confirmatory factor model (as opposed to exploratory factor models, such as principal components analysis, which are not considered here). The measurement model for political freedom (η_1) is:

$$y_1 = \lambda_{y,1}\eta_1 + \epsilon_1,$$
$$y_2 = \lambda_{y,2}\eta_1 + \epsilon_2.$$

The subscript for η is not necessary but is included to facilitate generalizations. The y notation used with the λ parameters differentiates them from similar notation to be developed for exogenous variables. The λ's measure the strength of the relationship between the observed and latent variable. The ϵ's are error terms. The measurement model for industrialization (ξ_1) is:

$$x_1 = \lambda_{x,1}\xi_1 + \delta_1,$$
$$x_2 = \lambda_{x,2}\xi_1 + \delta_2.$$

The measurement models are generalized as follows. Let $\mathbf{x}$ and $\mathbf{y}$ be $d_1 \times 1$ and $d_2 \times 1$ vectors of random variables. Then

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\epsilon}, \\ \mathbf{x} &= \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}. \end{aligned} \tag{1.1}$$

The vectors $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are latent normal random vector variables with dimensions $\leq d_1$ and d_2. Let $\boldsymbol{\Phi}$ be the covariance of $\boldsymbol{\xi}$. The vectors $\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$ are $d_1 \times 1$ and $d_2 \times 1$ random error vectors in the measurment (or modeling) of $\mathbf{x}$ and $\mathbf{y}$. These error vectors are assumed to be independent and from a normal distribution with covariances $\boldsymbol{\Theta}_\epsilon$ and $\boldsymbol{\Theta}_\delta$ respectively. These vectors are also assumed to be independent of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. The matrices $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ define the linear relationship between the latent variables and the observations. The latent variables are sometimes called factors and the coefficients of $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are then called factor loadings.

For the example ($d_1 = d_2 = 2$)

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_{y,1} \\ \lambda_{y,2} \end{bmatrix} \eta_1 + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}.$$

and

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \lambda_{x,1} \\ \lambda_{x,2} \end{bmatrix} \xi_1 + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}.$$

The structural part of the structural equation model is

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{1.2}$$

and links the two equations in (1.1). $\mathbf{B}$ is a matrix of coefficients defining the relationships of the endogenous latent variables on each other. $\boldsymbol{\zeta}$ is a normal random vector of process errors with covariance $\boldsymbol{\Psi}$ and is independent of $\boldsymbol{\xi}$,$\boldsymbol{\epsilon}$

and $\boldsymbol{\delta}$. $\boldsymbol{\Gamma}$ is the coefficient (loading) matrix of the effects of $\boldsymbol{\xi}$ on $\boldsymbol{\eta}$. For the example, $\eta_1 = \gamma_{1,1}\xi_1 + \zeta_1$.

The population covariance matrix for $\mathbf{y}$ and $\mathbf{x}$ ($\boldsymbol{\Sigma}$) may be partioned as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix}.$$

From (1.1),

$$\begin{aligned} \boldsymbol{\Sigma}_{yy} &= \boldsymbol{\Lambda}_y E(\boldsymbol{\eta}\boldsymbol{\eta}')\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon, \text{ and} \\ \boldsymbol{\Sigma}_{xx} &= \boldsymbol{\Lambda}_x \boldsymbol{\Phi} \boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta. \end{aligned}$$

From (1.2), $\boldsymbol{\eta} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$, so that

$$\begin{aligned} \boldsymbol{\Sigma}_{yy} &= \boldsymbol{\Lambda}_y (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon, \text{ and} \\ \boldsymbol{\Sigma}_{yx} &= \boldsymbol{\Lambda}_y (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x'. \end{aligned}$$

For the example, if all error terms are independent, then

$$\begin{bmatrix} VAR(y_1) & COV(y_1, y_2) & COV(y_1, x_1) & COV(y_1, x_2) \\ & VAR(y_2) & COV(y_2, x_1) & COV(y_2, x_2) \\ & & VAR(x_1) & COV(x_1, x_2) \\ & & & VAR(x_2) \end{bmatrix} =$$

$$\begin{bmatrix} \lambda_{y,1}^2(\gamma_{1,1}^2\phi_{1,1} + \psi_{1,1}) + \theta_{\epsilon,1} & \lambda_{y,1}\lambda_{y,2}(\gamma_{1,1}^2\varphi_{1,1} + \psi_{1,1}) & \lambda_{y,1}\lambda_{x,1}\gamma_{1,1}\phi_{1,1} & \lambda_{y,1}\lambda_{x,2}\gamma_{1,1}\phi_{1,1} \\ & \lambda_{y,2}^2(\gamma_{1,1}^2\phi_{1,1} + \psi_{1,1}) + \theta_{\epsilon,2} & \lambda_{y,2}\lambda_{x,1}\gamma_{1,1}\phi_{1,1} & \lambda_{y,2}\lambda_{x,2}\gamma_{1,1}\phi_{1,1} \\ & & \lambda_{x,1}^2\phi_{1,1} + \theta_{\delta,1} & \lambda_{x,1}\lambda_{x,2}\phi_{1,1} \\ & & & \lambda_{x,2}^2\phi_{1,1} + \theta_{\delta,2} \end{bmatrix}.$$

1.3 Estimation

The estimation of $\boldsymbol{\Sigma}$ involves a scalar fit function that measures the discrepancy between the observed variance-covariance matrix ($\mathbf{S}$) and $\boldsymbol{\Sigma}$. The fit function considered here is based on the maximum likelihood method and

is widely used although other functions, such as unweighted and generalized least squares (Bollen 1989), are common. The fit function that is minimized to produce maximum likelihood estimates when the data are from a multivariate normal distribution is:

$$log \mid \boldsymbol{\Sigma} \mid + tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log(\mid \mathbf{S} \mid) - p, \tag{1.3}$$

where tr is the matrix trace and $p = d_1 + d_2$ is the dimension of S. (1.3) is proportional to the loglikelihood (**Appendix B**). If $\boldsymbol{\Sigma} = \mathbf{S}$ then (1.3) = 0.

The parameters that minimize (1.3) are found numerically. Lee and Jennrich (1979) give a good description of some common algorithms. SAS's (1990) PROC CALIS (CALIS is an acronym for **C**ovariance **A**nalysis and **LI**near **S**tructural equations) is used here to compute parameter estimates. The Quasi-Newton optimizing technique is selected to equate (1.3) to zero but other options are left at their default values.

Standard errors are obtained from the inverse of the information matrix (see **Appendix B**). The information matrix is proportional to the expected value of the second order derivatives of (1.3).

The uniqueness of parameter estimates is known as identification and is an important consideration in structural equation models. Λ parameter in $\boldsymbol{\Sigma}$ is identified if it can be expressed as a function of one or more elements in $\mathbf{S}$. A simple example is the relationship $var(y) = \theta_1 + \theta_2$. The θ parameters are not identified; that is, no unique values of θ_1 and θ_2 satisfy the equation unless some constraint is imposed (e.g. $\theta_1 = \theta_2$). Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\theta}$ is an unknown vector of free parameters to be estimated. Bollen (1989) gives the following definition of identification: "If an unknown parameter in $\boldsymbol{\theta}$ can be written as a function of one or more elements in $\boldsymbol{\Sigma}$, that parameter is identified. If all unknown parameters are identified then the model is identified". Alternatively, let $\boldsymbol{\Sigma}_{\hat{\theta}}$ be the covariance matrix constructed from

a $\dot{\theta}$. If a model is identified then for $\dot{\theta}$ and $\ddot{\theta}$.

$$\boldsymbol{\Sigma}_{\dot{\theta}} = \boldsymbol{\Sigma}_{\ddot{\theta}} \Rightarrow \dot{\theta} = \ddot{\theta}.$$

The model is unidentified if these definitions do not hold.

A parameter is overidentified if it can be written as several functions of only known sample covariances. If all parameters in a model are identified and at least one parameter is overidentified then the model is overidentified, otherwise it is exactly identified. A parameter is underidentified if it can not be expressed only as a function of some elements in $\mathbf{S}$. If a model has one underidentified parameter then no unique solution for the unknown parameters exists and the model is unidentified. Bollen (1989) presents necessary and/or sufficient conditions for identification.

In the example, the model is unidentified because there are 11 parameters to estimate and only 10 sample covariances. This remains the case no matter what the sample size is. Determining model identification does not just involve counting parameters; for example, it is possible that one parameter is underidentified and others overidentified such that there are more sample covariances than unknown parameters; however, the model is unidentified because one parameter is underidentified.

Estimation involves finding values for parameters that best explain the data for overidentified models. If a model is just identified then all estimating techniques should give the same results because a unique solution to $\boldsymbol{\Sigma} = \mathbf{S}$ exists. Underidentified models are usually reparameterized or their parameters are constrained such that the model is identified.

1.4 Inference

Statistical inference, for the most part, involves testing the overall model fit (goodness-of-fit) and the fit of model parameters and equations. Virtually

all overall fit measures involve functions of $\mathbf{S}$ and $\hat{\boldsymbol{\Sigma}}$ ($\hat{}$ denotes a maximum likelihood estimate). These tests gauge the closeness of $\mathbf{S}$ and $\hat{\boldsymbol{\Sigma}}$ and have the advantage of being able to indicate inadequacies not revealed by analyzing parameter estimates. Many measures of model fit are purposed in the literature (Mulaik *et al.* 1989 give a review): for example, SAS's (1990) PROC CALIS routinely computes:

1. Fit criterion,
2. Goodness of Fit Index (GFI),
3. GFI Adjusted for Degrees of Freedom (AGFI),
4. Root Mean Square Residual (RMR),
5. Chi-square,
6. Chi-square p-value,
7. Null Model Chi-square,
8. Bentler's Comparative Fit Index,
9. Normal Theory Reweighted LS Chi-square,
10. Akaike's Information Criterion,
11. Consistent Information Criterion,
12. Schwarz's Bayesian Criterion,
13. McDonald's (1989) Centrality,
14. Bentler & Bonett's (1980) Non-normed Index,
15. Bentler & Bonett's (1980) Normed Index2,

16. James, Mulaik, & Brett (1982) Parsimonious Index,

17. Z-Test of Wilson & Hilferty (1931),

18. Bollen (1986) Normed Index Rho1,

19. Bollen (1988) Non-normed Index Delta2, and

20. Hoelter's (1983) Critical N.

Several papers about model fit are presented in Bollen and Long (1993).

The chi-square statistic is the only measure of fit that can be used for a statistical test of significance however, and it is used in the data analysis in the next chapter. The distribution of the statistic under the null hypothesis $\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}$ is asymptotically χ^2 with degrees of freedom $= p(p+1)/2 - t$ and t is the number of free parameters estimated. The chi-square statistic is N-1 times (1.3). It can be used to test the null hypothesis verses the alternative that $\boldsymbol{\Sigma}$ is a symmetrix matrix with $p(p+1)/2$ free parameters. Wheaton (1987) recommends the chi-square statistic even when used only in a descriptive sense, although not uniformly. The adjusted goodness-of-fit index (**AGFI**) for maximum likelihood estimates (Bollen 1989) is also used in the next chapter for comparative purposes.

The **AGFI** is computed as follows. Let

$$\mathbf{GFI} = 1 - \frac{tr[(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S} - \mathbf{I})^2]}{tr[(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S})^2]}, \qquad \text{then}$$

$$\mathbf{AGFI} = 1 - [\frac{p(p+1)}{2df}][1 - \mathbf{GFI}].$$

The **AGFI** measures the relative amount of variances and covariances in $\mathbf{S}$ that are predicted by $\hat{\boldsymbol{\Sigma}}$, adjusted by the degrees of freedom of the model relative to the number of variables considered, and is a better index for comparing different models. **AGFI** and **GFI** have maximums of one obtained when $\hat{\boldsymbol{\Sigma}} = \mathbf{S}$.

The Z test with linearized asymptotic standard errors and the likelihood ratio procedure are used to test hypotheses about model parameters. Other procedures have been developed, such as a Langrangian Multiplier and Wald test (Bollen 1989), that are computationally easier than the likelihood ratio test.

1.5 Practical issues

Much of the theory developed for structural equation models is based on large sample sizes; however, an understanding of the minimum small sample required for inferences to remain valid is useful. A rule of thumb presented by Bentler and Chow (1987) is that the minimum ratio of sample size to number of free parameters should be 5:1 under normal theory. Geweke and Singleton (1980) studied one and two factor confirmatory factor models and found that the chi-square statistic behaved well for sample sizes of 10 and 30 respectively. Boomsma (1985) suggests not to do a LISREL type analysis with sample sizes less than 100 because maximum likelihood solutions break down. Gerbing and Anderson (1985) tested the validity of the chi-square statistic and also the bias in estimates of parameters and their standard errors and found that procedures worked well for sample sizes as small as 50 except for a confirmatory factor model with two factors and two indicators per factor. La Du and Tanaka (1989) found that the values of several goodness of fit measures depended on sample size and estimation method but they did not make any recommendations about "minimum" sample sizes. Myers and Cadigan (1993b) found that the maximum likelihood estimate of one parameter was, in some cases, very biased in a confirmatory factor model with two factors and two indicators per factor. The standard Z test for this parameter was conservative and had low power for sample sizes of 10 and 20.

The sensitivity of a LISREL analysis to categorical data has been stud-

ied by Homer and O'Brien (1988) and Ethington (1986). Jöreskog (1990) has adapted the LISREL program for crude categorical data. Browne and Shapiro (1988) give analytic results on the robustness of normal theory methods and find that the chi-square test is reasonable to use as long as kurtosis is not excessive. A review of robustness is presented by Satorra (1990). Simulations conducted by Myers and Cadigan (1993a) which are presented in the next chapter show that maximum likelihood estimates are more sensitive to sample size than moderate deviations in some model assumptions.

In practise, the user of structural equation models should be familiar with other issues as well. The scale of variables used in structural equation models can cause difficulties and it may be more appropriate to analyze the correlation matrix than the covariance matrix. Negative variance estimates are possible when the estimation procedure does not constrain the additive variance components to be positive. Negative estimates of variance components are common (Searle *et al.* 1992) and are linked with small sample sizes by Boomsma (1985), Gerbing and Anderson (1987) and also in Chapter 3 and must be interpreted with care; for example, the analysis of identification assumes variance components are positive. A rule of thumb is that the negative estimate is due to sampling error if its 95% confidence interval contains zero (Gerbing and Anderson 1987). Missing values are a problem (see Muthen *et al.* (1987) for an analysis). Model building can be difficult so modification indices have been developed (Bollen 1989) as an aid. Bentler and Chou (1987) deal with these issues very thoroughly.

Chapter 2

Recruitment in cod

2.1 Introduction

Density dependent juvenile mortality in marine demersal fish was investigated by Myers and Cadigan (1993a,b), known hereafter as MCa and MCb, for 6 species and from 1 to 8 stocks per species. Density dependent juvenile mortality, as assumed here (the Ricker model), means that the mortality rate of eggs and juvenile fish is proportional to initial population size (Hilborn and Walters 1992). A single factor confirmatory factor model was developed in Mca but the results were not presented because they were very similar to portions of a more complicated model whose results were presented. This analysis for cod stocks is presented here along with a discussion of the results.

The purpose of the analysis is to determine if density dependent mortality in the juvenile stage of marine demersal fish is significant. Density dependent mortality is measured as an association between mortality and population abundance. In previous studies (e.g. Beverton and Iles 1991a,b and Sundby *et al.* 1989) it has been difficult to determine if the estimated association between mortality and abundance is a statistical artifact (Veer 1986) because measures of mortality and abundance are typically correlated through their method of construction. This problem was addressed in MCa and MCb by

using measurement error models for multiple annual surveys of a stock.

A major problem with the surveys is their length; all surveys occurred over periods less than 25 years. The smallest sample size reported for which the estimates of confirmatory factor models are unbiased (Gerbing and Anderson 1985) is 50. Simulations were conducted to assess the bias and many stocks were analyzed to increase the reliability of the results.

2.2 Data

The data sets analyzed are presented in Tables 2.1-2.6 in **Appendix A**. The data in Tables 2.1 and 2.2 are extracted from ICES (1992a) and (1992b) respectively. The data in Tables 2.3 and 2.4 are presented in Anon. (1991). The Canadian survey data in Table 2.5 is extracted from Baird *et al.* (1992) and the Russian data is extracted from Kuzmin (1992). The data in Table 2.6 is extracted from ICES (1992c). A more detailed account of the data is given in MCa.

The s notation in Tables 2.1-2.6 needs explanation. The surveys are relative estimates of recruitment, which is the number of newborn fish in a year. Denote years as y and cohorts (fish born in the same year) as t. Let $\nu_{y,a}$ be the number of fish of age a surveyed in year y. $\nu_{y,0}$ provides an estimate of recruitment for cohort $t = y$, while $\nu_{y,1}$ provides an estimate of recruitment for cohort $t = y - 1$, $\nu_{y,2}$ for cohort $t = y - 2$, etc. Four surveys (s_1.....s_4) for cohort t are formed from possibly different surveys of numbers at age a in year y, where $y = t + a$. Hence, the data in Tables 2.1-2.6 are in terms of cohorts with the surveys (s's) ordered by length of time since recruitment. For example, in Table 2.1, the recruitment estimate of 20 in the MARCH 1 year old's survey for 1975 actually came from a survey of 1 year old's in March of 1976 and the estimate of 4 in the June 1 year old's survey came from a survey of 1 year old's in June of 1976. These MARCH and JUNE

surveys are denoted as s_2 and s_3 because the JUNE survey of the cohort is at a longer time since recruitment than the MARCH survey.

2.3 Population dynamics

The population dynamics during the juvenile stage is described by:

$$\nu_{t,a} = \nu_{t,0} e^{-\alpha_a},$$

where $\nu_{t,a}$ is the number of fish of age a in cohort t. The parameter α_a represents mortality to age a, and $\alpha_0 = 0$. The density dependent component of mortality is assumed to be proportional to log abundance and results in MCa for North Sea cod suggest this is reasonable. Let $(1 - \lambda_a) \log \nu_{t,0}$ be the density dependent mortality where $\lambda_0 = 1$. The model with density dependent mortality is:

$$\nu_{t,a} = \nu_{t,0} e^{-\alpha_a - (1-\lambda_a) \log \nu_{t,0}}.$$

The natural logarithm transform gives

$$\xi_{t,a} = \lambda_a \xi_{t,0} - \alpha_a,$$

where $\xi_{t,a} = \log \nu_{t,a}$.

Let $x_{t,a,i}$ be the log survey of cohort t based on age a fish from survey i. If $\gamma_{a,i}$ is the log catchability coefficient (catchability is the proportion of the population caught in the survey) of survey i on age a, $\lambda_{a,i}$ is the density dependent survival coefficient, and $\delta_{t,a,i}$ is the observational error of survey i on age a of cohort t, then the observed log transformed abundance from age a and survey i is modeled as

$$\begin{aligned} x_{t,a,i} &= \gamma_{a,i} + \xi_{t,a} + \delta_{t,a,i}, \\ &= \gamma_{a,i} - \alpha_a + \lambda_{a,i} \xi_{t,0} + \delta_{t,a,i}. \end{aligned}$$

The parameters $\gamma_{a,i} - \alpha_a$ are removed by standardizing each $x_{t,a,i}$ (subtracting the sample mean, $\frac{1}{N}\sum_{t=1}^{N} x_{t,a,i}$, where N is the number of years). Hence $x_{t,a,i}$, $\xi_{t,0}$ and $\xi_{t-1,0}$ are all in terms of deviations from their means. The units of the surveys are generally in terms of numbers per unit of fishing but are irrelevant if the surveys are proportional to true recruitment because the surveys are log transformed and have their means removed.

Denote $x_{t,a,i}$ as $x_i = log(s_i)$, $\lambda_{a,i}$ as λ_i, and $\xi_{t,0}$ as ξ. The population dynamics model can be written as:

$$x_i = \lambda_i \xi + \delta_i.$$

ξ represents the yearly deviation in recruitment so that the E(ξ) = 0, but ξ varies from year to year. 1 - λ_i represents density dependent mortality and it is expected that the λ_i's will decrease with i because the surveys are arranged in time since recruitment. The errors (δ's) and ξ are assumed to be independent and normally distributed; consequently, the model is equivalent to (1.1) (a confirmatory factor model). A path diagram of the model is presented in Figure 3.2 in Chapter 3.

A necessary, but not sufficient condition, for identification is met (the t rule) in Bollen's (1989) Table 7.3 because $df = 2$. A sufficient condition is the Three-indicator rule: $\boldsymbol{\Lambda}_x$ in (1.1) has 1 nonzero element in each row, 3 or more x's per factor, and $\boldsymbol{\Theta}_{\boldsymbol{\delta}} = cov(\boldsymbol{\delta}\boldsymbol{\delta}')$ is diagonal. This rule is met so the model is identified.

2.4 Model estimation

The estimation procedures described in the last chapter are used here with a few modifications. Zero's in Tables 2.1-2.6 are replaced by $\frac{1}{10}$ the minimum (over all years) of the nonzero catches for the respective survey because of the log transformations. The procedure was applied replacing $\frac{1}{10}$ with $\frac{1}{2}$ and the

parameter estimates did not change very much. The lagging procedure used to obtain the cohort estimates results in missing values at the beginning and end of the recruitment time series. In addition, some surveys started later than others, and some have been discontinued. Missing values are accommodated by analyzing the pairwise covariance matrix: this matrix uses all nonmissing pairs of observations in computing the sample covariance for any two variables. Another procedure would be to use only that portion of the data with no missing values for all the variables but this would eliminate too much data for some stocks.

The estimation procedure is not constructed to handle missing values nor is it tractable to specify a likelihood for the pairwise covariance matrix so this matrix is treated as if it came from a sample with no missing values. The optimization technique requires the sample covariance matrix to be positive definite and, if not, this matrix is ridged with a constant (added to all the diagonal elements) large enough to make it positive definite. The sample size used in computing standard errors for this procedure is the minimum length of all recruitment time series.

2.5 Model estimates

Estimates are presented in Table 2.7; recall that

- the λ's are the density dependence parameters,
- the θ's are the variances of the errors (δ's), and
- ϕ is the variance of the true recruitment (ξ).

λ_1 is always constrained to equal 1. The joint probability that all unknown $\lambda's$ are ≤ 1 is approximated with 10000 monte-carlo samples obtained using the parameter estimates and their asymptotic distributions. The probability

is obtained by counting the number of times any $\lambda \geq 1$ and dividing this count by the number of replications. This sampling procedure produced 2 digit accuracy when applied to some known results. The joint probabilities for three stocks indicate it is more likely than not that the λ's are less than one. The Gulf of Maine stock is the most likely in which one would conclude that the λ's are not less than one.

The θ's generally range from being approximately equal to $\frac{1}{5}$ of ϕ. One θ estimate is negative for the North Sea stock. $\rho_{s_i s_i}$ is the precision or reliability of survey i (Bollen 1989); it is the squared correlation between the observed survey and the true (latent) recruitment and is interpreted as the proportion of the surveys variance explained by true recruitment.

The Georges Banks cod stock produced the lowest chi-square p-value and adjusted goodness-of-fit index of all stocks examined followed by the Gulf of Maine stock. Only the chi-square p-value is reported because this is scaled between zero and one so it is comparable with the adjusted goodness-of-fit index, the statistic can be obtained as a chi-square ($df = 2$) percentile. Recall that a low adjusted goodness-of-fit index indicates poor model fit. The results for both these stocks should be treated with suspicion. The reliability of some surveys for these stocks are the lowest of all stocks considered.

The normalized residuals (Bollen 1989) were analyzed for each cod stock but they generally did not indicate results different from the overall fit measures (i.e. no components of the models fit particularly worse than others) and are not presented.

The estimated λ's and θ's are plotted verses age in Figures 2.1 and 2.2. The data in these figures is presented in Table 2.7. Age is defined as time since recruitment. Generally there is a decreasing trend in the estimated λ's and θ's with age. The overall decreasing trend in the λ's is evidence of density dependent mortality and is found in most stocks examined.

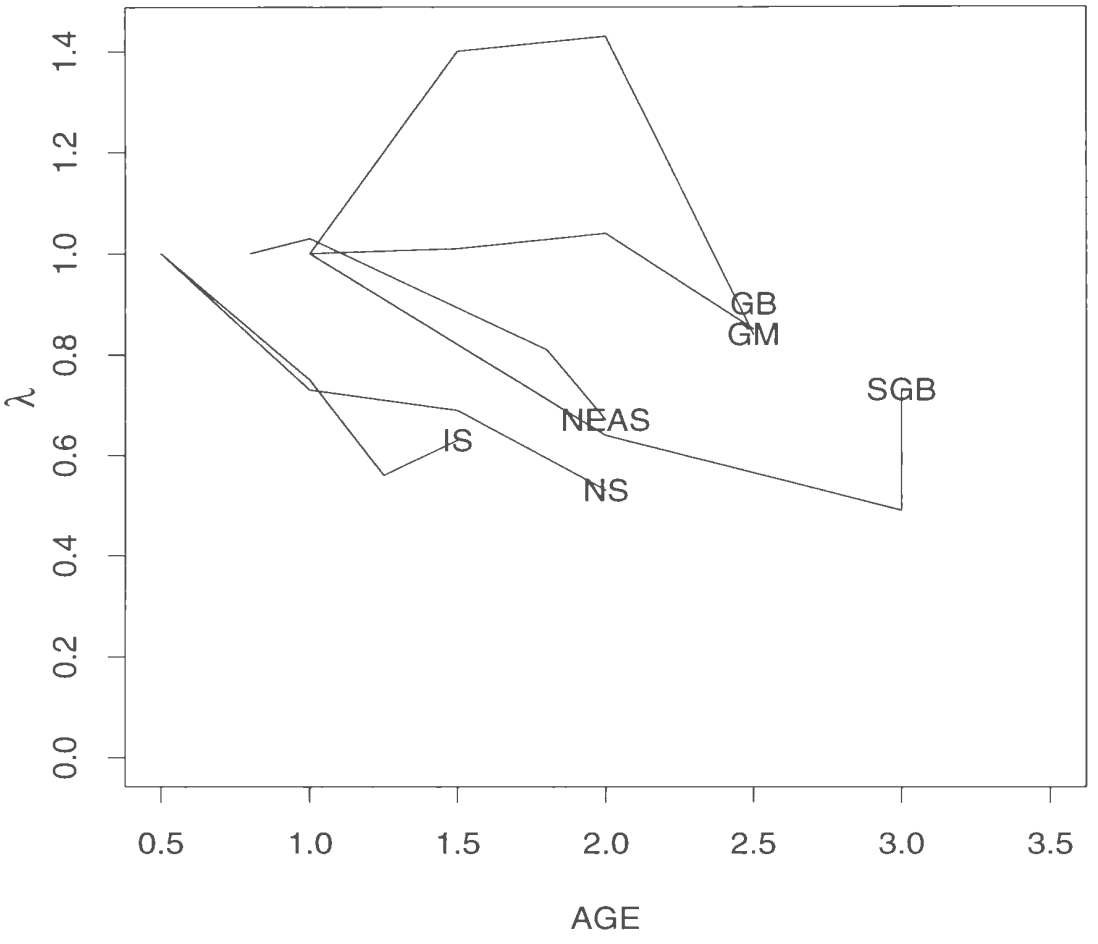

Figure 2.1. The density dependent mortality coefficient, λ, in relation to age for cod. The population locations are abbreviated as follows: GB - Georges Banks, GM Gulf of Maine, IS - Irish Sea, NS - North Sea, NEAS - North East Arctic SGB - Southern Grand Banks.

2.6 Sensitivity analysis

A sensitivity analysis is conducted to assess whether the results could be caused by statistical artifacts; that is, does density dependence appear for the wrong reasons. The sensitivity, in terms of percentage bias, of the

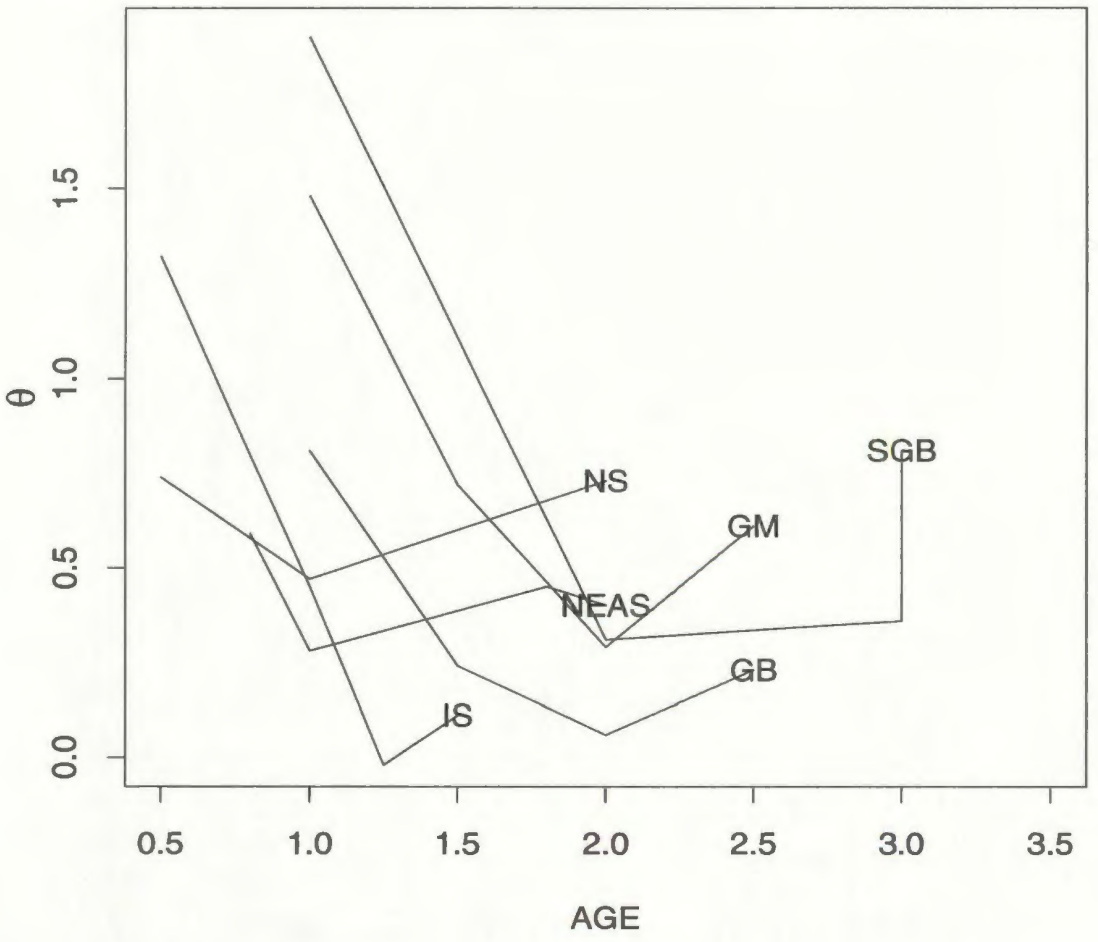

Figure 2.2. Estimates of survey error variance, θ, in relation to age for cod. The population locations are abbreviated the same as in Figure 2.1.

estimators to violations in distributional assumptions and small sample sizes is assessed by simulations. The analysis is taken, in part, from MCa.

The following parameters are fixed in the simulations:

- The density dependence parameters (λ_i) are all equal to one.
- The variability of the true recruitment (ϕ) is equal to one.

Error variances are constrained to be equal ($\theta_1 = ... = \theta_4 = \theta$). In all

simulations θ is equal to and $\frac{1}{5}$ of ϕ, these are rough bounds on estimates in Table 2.7. Sample sizes studied ranged from 10 to 100 but 20 is the norm. 1000 simulated data sets were generated for each effect studied.

The results are presented in Table 2.8. The first two columns represent the simulation number and corresponding sample size. The next two columns indicate the assumed distribution of log recruitment (ξ) and errors (δ's). Columns 5 and 6 show the values of parameters used in the simulation. The notation θ/ϕ is used, even though $\phi = 1$, to indicate that the ratio of variances is more important than their actual values. The mean biases for $\lambda_2, ..., \lambda_4$ and $\theta_{1,1}, ..., \theta_{4,4}$, and their ranges, are reported to simplify the analysis. The last column contains the percentage of simulated solutions that converged with positive variances ($\theta_{1,1}, ..., \theta_{4,4}, \phi \geq 0$).

In the first 6 simulations the effect of the number of years (n) used to estimate parameters is tested. The assumption about the distribution of ξ is investigated in simulations 7-8 by generating it as the natural log of a gamma random variable whose mean and variance equal that of the natural exponent of a standard normal random variable. The gamma distribution is chosen to represent a nonsymmetric distribution. The sensitivity of the analysis to the error distribution is checked in simulations 8-12 by generating right skewed δ's from a gamma random variable with shape and scale parameters 2 and $1/\sqrt{2}$ respectively. The left skewed gamma deviate is the negative of the right skewed deviate.

The effect of correlated errors is tested in simulations 13-14 by adding 0.1 to the first half of the errors (δ_1) and -0.1 to the later half. δ_3 is modified by the same procedure except 0.1 is added to the first half + 1 of the errors. These simulations model changes in measuring techniques halfway through a research survey time series.

The effect of unequal estimation error variances is investigated in the

last four simulations. Of particular concerned is that large estimation error variances at younger ages might lead to biased results for small sample sizes. An estimation error variance of the youngest age, $\theta_{1,1}$, equal to 4 or 9 times the other error variances (which are all assumed equal) is considered.

Biases in parameter estimates result from small sample sizes and not the deviations in model assumptions considered. The simulations indicate that bias in the maximum likelihood estimates of the model in Figure 3.2 can be up to 30%. Also, if the θ/ϕ ratio (related to reliability, see Chapter 3) is large then the biases may be large. However, the simulated biases are too small and are of the wrong sign for estimates of density dependent mortality to be artifacts of violations in model assumptions, except if the estimation error variance is large.

If the error variance for the youngest age is much larger than for older ages and similar to ϕ, then very large biases may result and density dependent mortality may be estimated to exist when there is none. The estimates of variance parameters in Table 2.7 for most stocks are similar to the parameters in simulations 3, 4 and 16 in Table 2.8. The biases are not large for these simulations, particularly for No.'s 4 and 16. The Gulf of Maine and Georges Banks cod stocks have variance estimates somewhere between those considered in simulations 3 and 15. While these biases as large, they would not result in false conclusions. The variance estimates from all stocks are not in the range of simulation 15. It seems safe to conclude that estimates of density dependent mortality are not statistical artifacts.

2.7 Discussion

Density dependent juvenile mortality appears to be significant in most of the cod stocks examined and this study is the first to demonstrate this. The confirmatory factor model used here has so far had little application in

population ecology and fisheries science and has proven to be an effect tool in investigating difficult hypotheses.

The results appear to be robust in that the conclusion that density dependent mortality is important is not affected by reasonable violations in the model assumptions. If anything, the methods under-estimate the strength of density dependent mortality. The estimates of observational error variance and inter-annual recruitment variance are negatively and positively biased respectively, but usually not more than 10%.

The degree of density dependent mortality is often strong; for example, λ is estimated to be 0.53 for age 2 North Sea cod (Table 2.7). The presence of such strong density dependent mortality within cohorts may explain why there is often little relationship between stock and recruitment; that is, large recruitment that occurs when stock size is large may suffer higher mortality during the juvenile stage.

A problem with this analysis is that errors (δ's) for recruitment surveys obtained from different ages of a numbers at age survey may be correlated; however, the model assumes errors are independent. There are generally insufficient degrees of freedom to test this assumption unless all correlations are assumed to be equal, which is unreasonable. Another problem is that the data are not observations but the result of surveys and have estimation error variances associated with them. It is plausible that the inclusion of standard errors might affect the results.

Unbiasedness is important in this type of analysis because statistical inference is based more on the distribution of estimates from independent samples than from any particular sample, e.g. no attempt is made to consider standard errors in Figure 2.1. However, repeating a biased estimate does not decrease bias. It would be useful to estimate the models with a biased reduced estimator as a check of the effect of small sample bias on conclusions.

Chapter 3

Bias reduction using the jackknife

3.1 Introduction

Previous research about bias in structural equation models (Boomsma 1985; Gerbing and Anderson 1985) has not dealt with sample sizes as small as those considered in the previous chapter. The simulations in Chapter 2 indicate that there is a potential for large biases when sample sizes are small. Also, the maximum likelihood estimate of a particularly interesting parameter in MCb has large bias in small samples. A biased reduced estimator that consistently reduces bias without large increases in standard errors is desirable. Previous research indicates the jackknife (and the bootstrap) does not reduce bias in structural equations models (Boomsma 1986); however, only a two factor model was considered and estimates were based on the correlation matrix. It is not clear if these results extend to all models or estimates based on the covariance matrix. The jackknife is investigated as a biased reduced estimator through simulations in addition to investigating the bias of several confirmatory factor models more rigorously than in Chapter 2.

The jackknife is a statistical technique first purposed by Quenouille (1956) and Tukey (1958) to estimate the bias and variance of an estimator. Que-

nouille developed the jackknife as a simple nonparametric method of estimating the bias of a maximum likelihood estimator when a parametric approach proved too difficult. Tukey recognized the usefulness of the jackknife in terms of bias reduction and variance estimation. Other authors such as Schucany et al (1971), Miller (1974), Efron (1982) and Wu (1986) have made important contributions. A review up to the mid 1980's is found in Frangos (1987).

3.1.1 Delete-one jackknife

The notation used loosely follows Efron (1982). For simplicity, the univariate case is considered; however, the extension to multivariate problems is straightforward. Let $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters and not necessarily error variances. Consider a random sample $x_1, ..., x_n \quad iid \quad F$. Let $\hat{\boldsymbol{\theta}}_{(i)}$ denote the maximum likelihood estimate of $\boldsymbol{\theta}$ based on the sample with x_i removed: $x_1, ..., x_{i-i}, x_{i+i}, ..., x_n$ and let $b(\hat{\boldsymbol{\theta}})$ be the bias of $\hat{\boldsymbol{\theta}}$.

For many estimators, including maximum likelihood estimates, $E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{a_1(F)}{n} + \frac{a_2(F)}{n^2} + ...$ where the functions $a_i(.)$, i = 1,..., do not depend on n and the bias is $O(\frac{1}{n})$ (Efron 1982). $O(\frac{1}{n})$ bias means that the bias is $< K/n$, where K is a constant. Define the ith pseudo-value (Tukey 1958) as

$$\tilde{\boldsymbol{\theta}}_i = \hat{\boldsymbol{\theta}} - (n-1)(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}). \tag{3.1}$$

The bias of $\tilde{\boldsymbol{\theta}}_i$ is $O(n^{-2})$; that is

$$E(\tilde{\boldsymbol{\theta}}_i) = \boldsymbol{\theta} - a_2(F)(\frac{1}{n(n-1)}) + a_3(F)(\frac{1}{n^2} - \frac{1}{(n-1)^2}) + ...$$

If $\hat{\boldsymbol{\theta}}$ is taken as an approximation for $\boldsymbol{\theta}$ then the jackknife bias estimate is given by (Quenouille 1956)

$$\tilde{b}(\hat{\boldsymbol{\theta}}) \quad = \quad -\frac{1}{n}\sum_{i=1}^{n}(\tilde{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}).$$

The average is taken to maximize the efficiency of the procedure. The jackknife bias reduced estimate of $\boldsymbol{\theta}$ is

$$\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \tilde{b}(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{\theta}}_i. \tag{3.2}$$

The bias in $\tilde{\boldsymbol{\theta}}$ is $O(n^{-2})$, therefore the jackknife reduces bias by a factor of $1/n$ provided first moments exist and the $a_i(F)$ are bounded for all i.

Tukey's jackknife variance approximation is

$$\tilde{V}(\hat{\boldsymbol{\theta}}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\tilde{\boldsymbol{\theta}}_i - \tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}}_i - \tilde{\boldsymbol{\theta}})' \tag{3.3}$$

(3.3) can also be used as an estimate of $V(\tilde{\boldsymbol{\theta}})$ and it is for this purpose that the jackknife is often used. Tukey suggested that

$$\frac{\hat{\theta}_i - \theta_i}{\tilde{V}(\hat{\theta}_i)}$$

($\hat{\theta}_i$ is a scalar and is the ith element of $\hat{\boldsymbol{\theta}}$) has approximately a t distribution with $n-1$ degrees of freedom; however, Miller (1964) has shown that this is generally not the case except for linear estimators.

3.1.2 Delete-g jackknife

A grouped jackknife estimator is defined by computing pseudovalues from a sample with a block of observations removed. Suppose $n = gh$. Remove $x_1, ..., x_h$ and compute $\hat{\boldsymbol{\theta}}_{(1)}$ from the remaining observations. Next, remove $x_{h+1}, ..., x_{2h}$ and compute $\hat{\boldsymbol{\theta}}_{(2)}$, etc. Compute $\tilde{\boldsymbol{\theta}}_i = \hat{\boldsymbol{\theta}} - (g-1)\hat{\boldsymbol{\theta}}_{(i)}$ and then use these pseudovalues in (3.2) and (3.3) replacing n with g. The groups need not be independent either: for any $h < n - p$ it is possible to form $\begin{pmatrix} n \\ h \end{pmatrix}$ groups and use the average of all pseudovalues as a jackknife estimate. Miller (1964) shows that the delete-1 jackknife variance estimator (3.3) is consistent only for smooth estimators; however, Wu (1986) suggest that the delete-g jackknife works better. The maximum likelihood estimates considered are smooth so only the delete one jackknife is considered here.

3.1.3 Weighted jackknife

In unbalanced data all observations are not equidistant from the center of the design; for example, data from an analysis of variance with unequal sample sizes in each treatment is unbalanced. Hinkley (1977) points out some of the shortcomings of the jackknife in a regression analysis with unbalanced data. These shortcomings apply to many nonlinear or nonnormal problems. Hinkley (1977) considers the regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}.$$

where the $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and the $COV(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$. He shows that

1. $\tilde{\boldsymbol{\theta}} \neq \hat{\boldsymbol{\theta}}$,

2. $V(\tilde{\boldsymbol{\theta}}) > V(\hat{\boldsymbol{\theta}})$ (guaranteed by the Gauss-Markhov theorem) except for balanced designs,

3. the jackknife variance estimator (3.3) is biased unless the diagonal elements (w_i) of $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ are all equal.

Hinkley suggests the bias is because "pseudovalues [(3.1)] are defined symmetrically with respect to the observations, whereas the model is generally unbalanced".

Hinkley (1977) purposes a weighted modification to the jackknife that is free from the above problems. $(n-1)(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)})$ in (3.1) is the sample influence curve for case i and is proportional to the change in $\hat{\boldsymbol{\theta}}$ when case i is deleted (Cook and Weisberg 1982). For the multiple regression case, Hinkley weights $(n-1)(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)})$ so that it refects the distance from $\mathbf{x}_i$ to the centroid of $\mathbf{x}_1, ..., \mathbf{x}_n$. He replaces the sample influence curve in (3.1) with the emperical influence curve - the rational being that the later is less

sensitive to remote cases. Define weighted pseudo-values as:

$$\begin{aligned}
\tilde{\boldsymbol{\theta}}^{H}_{w_i} &= \hat{\boldsymbol{\theta}} + n(1-w_i)(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)}), \quad \text{where} \\
w_i &= \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i, \quad \text{and} \\
\mathbf{X}'_{\mathbf{p\times n}} &= [\mathbf{x}_1, ..., \mathbf{x}_n].
\end{aligned}$$

Hinkley's weighted jackknife estimate of $\boldsymbol{\theta}$ is: $\tilde{\boldsymbol{\theta}}^{H}_{w} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{\theta}}^{H}_{w_i}$. The estimator has the property that $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$ and therefore $V(\tilde{\boldsymbol{\theta}}) = V(\hat{\boldsymbol{\theta}})$. $\tilde{V}^{H}_{w}(\hat{\boldsymbol{\theta}})$ is unbiased even if the w_i's are not all equal.

The methodology may not extend easily to some problems because the computation of the emperical influence curve can be difficult. Also, it is not obvious how the method extends to the delete-g ($g > 1$) jackknife. An alternative method presented by Wu (1986) does extend easily. Let

$$\begin{aligned}
\tilde{\boldsymbol{\theta}}^{W}_{w_i} &= \hat{\boldsymbol{\theta}} + n(n-p)w_i^{*}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)}), \quad \text{where} \\
w_i^{*} &\propto |\mathcal{I}_{(i)}|, \quad \sum w_i^{*} = 1.
\end{aligned}$$

$\mathcal{I}_{(i)}$ is the fisher information matrix for the sample reduced by the ith observation. Wu's weighted jackknife estimate of $\boldsymbol{\theta}$ is: $\tilde{\boldsymbol{\theta}}^{W}_{w} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{\theta}}^{W}_{w_i}$. Wu shows that $\tilde{V}^{W}_{w}(\hat{\boldsymbol{\theta}})$ is robust against error variance heterogeneity. For linear models:

$$\begin{aligned}
w_i^{*} &= \frac{|\mathbf{X}_i'\mathbf{X}_i|}{\sum_i |\mathbf{X}_i'\mathbf{X}_i|}, \\
&= \frac{|\mathbf{X}'\mathbf{X}|(1-w_i)}{\sum_i |\mathbf{X}'\mathbf{X}|(1-w_i)}, \\
&= \frac{1-w_i}{n-p}.
\end{aligned}$$

so that $\tilde{\boldsymbol{\theta}}^{H}_{w} = \tilde{\boldsymbol{\theta}}^{W}_{w}$.

The robustness of the jackknife is also considered by Hinkley (1978) in terms of estimating a correltion coefficient. Hinkley and Wang (1980) purpose

a trimmed jackknife (the pseudovalues are trimmed) as a naive type of robust estimate.

Wu's procedure is appealing for structural equation models because it excludes subsample solutions that do not converge because $\mathcal{I}_{(i)}$ is usually singular in this case. This topic is not addressed further except to say that in some preliminary unreported analyses (simulations using homoscedastic data) Wu's weighted jackknife performed very poorly, followed by Hinkley's weighted jackknife and the unweighted jackknife.

3.2 Methods

The jackknife is modified so that if a subsampled solution does not converge it is discarded and the sample size in the bias and variance formulae is reduced by 1. This is usually what Wu's weighted jackknife would do for nonconvergent cases. If the full sample solution converges but none of the subsampled solutions converge then the jackknife estimate is said to not converge.

Three confirmatory factor models are considered in assessing how the jackknife works as a bias reduced estimator. These models are depicted in Figures 3.1-3.3 and are denoted as M1, M2 and M3. The jackknife is assessed by simulating known models and comparing estimates with their known values.

M1 is chosen because it is a simple, exactly identified model that has closed form expressions for parameter estimates. The population covariance matrix is:

$$\begin{bmatrix} \phi + \theta_1 & \lambda_2\phi & \lambda_3\phi \\ & \lambda_2^2\phi + \theta_2 & \lambda_2\lambda_3\phi \\ & & \lambda_3^2\phi + \theta_3 \end{bmatrix}.$$

There are 6 parameters to estimate from 6 sample covariances. Let the observed sample covariance matrix $\mathbf{S} = \{s_{i,j}\}$.

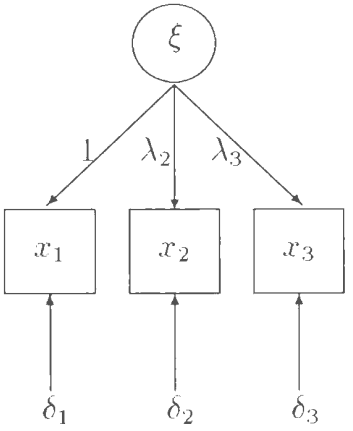

Figure 3.1. Path diagram of a confirmatory factor model with one factor and three variables (M1).

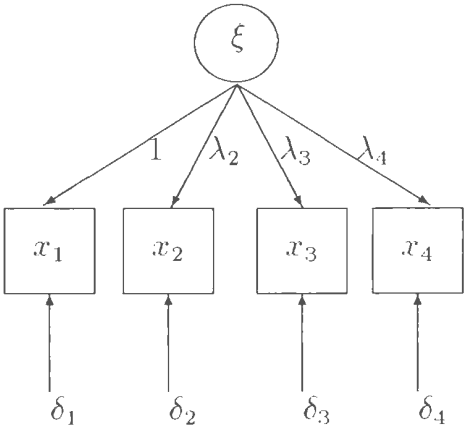

Figure 3.2. Path diagram of a confirmatory factor model with one factor and four variables (M2).

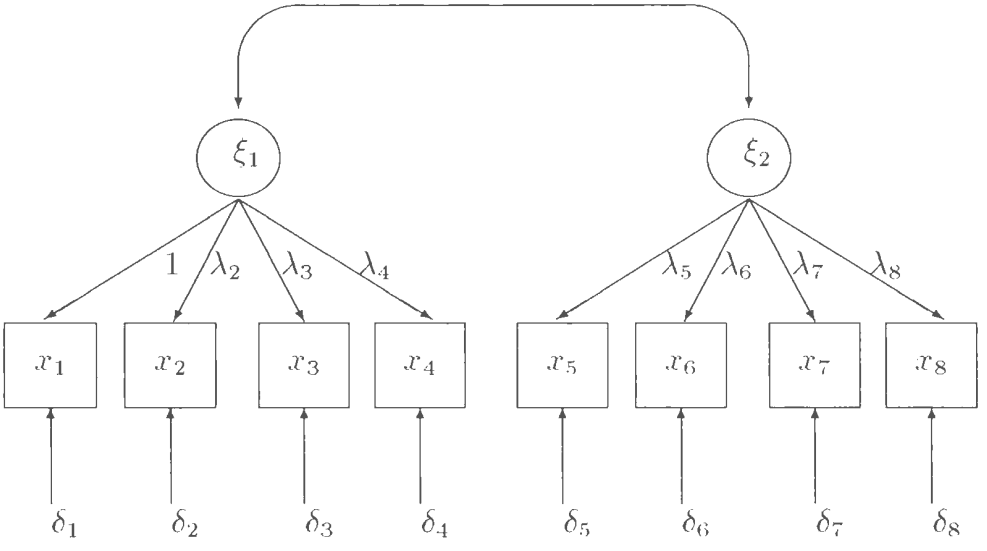

Figure 3.3. Path diagram of a confirmatory factor model with two factors and four variables per factor (M3).

The parameter estimators are

$$\begin{aligned}
\hat{\phi} &= \frac{s_{1,2}s_{1,3}}{s_{2,3}}, \\
\hat{\lambda}_2 &= \frac{s_{2,3}}{s_{1,3}}, \\
\hat{\lambda}_3 &= \frac{s_{2,3}}{s_{1,2}}, \\
\hat{\theta}_1 &= s_{1,1} - \frac{s_{1,2}s_{1,3}}{s_{2,3}}, \\
\hat{\theta}_2 &= s_{2,2} - \frac{s_{1,2}s_{2,3}}{s_{1,3}}, \\
\hat{\theta}_3 &= s_{3,3} - \frac{s_{1,3}s_{2,3}}{s_{1,2}}.
\end{aligned}$$

These are maximum likelihood estimates because $\hat{\boldsymbol{\Sigma}} = \mathbf{S}$.

M2 is considered because it is the model estimated in the previous chapter and it is desirable for the results obtained here to apply to those estimates. Also, M2 can be used with M1 to assess the effect of adding an indicator variable to a model. M3 is chosen to assess if the results obtained from the analysis of M1 and M2 extend to a more complex model, and to investigate the biases in parameters solely related to latent variables.

Many combinations of parameters and sample sizes can be considered in generating simulated data, even for the model in Figure 3.1. It is necessary to constrain those parameter combinations and sample sizes to the minimum required to understand their effect. Small sample size effects are the primary interest so sample sizes of $n = 10$, 30 and 50 are considered for all models. θ's that are one-tenth or equal to the latent variable variance (ϕ) are considered; this results in differential reliability (see the **Discussion** for more detail) or correlation between observed and latent variables. Factor variances and loadings are all fixed as one. The correlation among latent variables is 0.9 in M3.

1000 simulations for each model are performed. The latent variables and errors are generated as normal random deviates with predetermined

covariances and the data then generated using (1.1). The following measures are considered in simulations:

- The percentage bias in the maximum likelihood estimates of model parameters ($\hat{\boldsymbol{\theta}}$).
- The percentage bias in the jackknife estimates of model parameters ($\tilde{\boldsymbol{\theta}}$).
- The ratio of the variance of $\tilde{\boldsymbol{\theta}}$ ($V(\tilde{\boldsymbol{\theta}})$) and the variance of $\hat{\boldsymbol{\theta}}$ ($V(\hat{\boldsymbol{\theta}})$).

Percentage bias is the bias divided by the true value. The variances are approximated by the sample variance of the 1000 simulated estimates. The ratio of variances is considered because a small reduction in bias may result in a large decrease in precision which is not useful.

The Quasi-Newton option is used in estimating parameters with SAS's (1990) PROC CALIS because this option resulted in slightly better convergence rates compared to the default option. The computational complexity of the simulations should not be underrated. For one model and one set of parameters, considering the 3 sample sizes, a total of 93000 optimizations are performed, this requires considerable computer time and storage.

3.3 Results

There are three different results obtained in this analysis. The first is the potential biases in maximum likelihood estimates of the parameters in some structural equation models. The second result is whether the jackknife is an effective tool in reducing the bias of the maximum likelihood estimates. The third result deals with the extent of improper (negative) estimation of variance components by maximum likelihood and jackknifing.

The simulated % bias and variance ratio of the maximum likelihood and jackknife estimates are presented in Tables 3.1 (M1), 3.4 (M2) and 3.7 (M3) for the λ and ϕ parameters and in Tables 3.2, 3.5 and 3.8 for the θ parameters. The percentage of converged solutions and the percentage of converged solutions with positive variances is presented in Tables 3.3, 3.6 and 3.9. The values of error variances and the sample size used in the simulations are indicated in the first four rows.

3.3.1 Bias of maximum likelihood estimates

Some consistent patterns in the results are:

- The percentage bias of maximum likelihood estimates in Tables 3.1, 3.2, 3.4, 3.5, 3.7 and 3.8 generally decreases with sample size and indicator reliability.

- The bias in λ's and φ's is generally positive except for $\lambda_5, .., \lambda_8$ in M3; it is negative for θ's.

- The inclusion of a variable (e.g. the difference between M1 and M2) in the single factor model did not affect the bias of parameter estimates.

- The percentage bias of θ is relatively large compared to other parameters when $\theta = 0.1$ than when $\theta = 1.0$.

- If $n = 10$, less reliable observed variables seems to result in relatively more bias in the maximum likelihood estimate of ϕ compared to other parameters, although this result is not consistent throughout all models.

What is not evident in these tables is that the frequency distribution of simulated estimates is different if the reliability of observed variables is low

than if it is high. Histograms of simulated estimates look well behaved and "normal" for $n = 50$ if $\theta = 0.1$ but a disturbing feature if $\theta = 1.0$ is the presence of a small number of very unreasonable estimates. For example, when $n = 10$ and $\theta = 0.1$ is used in generating simulated data in M1, the maximum and minimum estimates for $\theta_{3,3}$ are -0.171 and 0.387 (Figure 3.4) but when $\theta = 1.0$ the maximum and minimum are -179 and 33.8 (Figure 3.5). The latter estimates are clearly atypical in the histograms.

3.3.2 Bias reduction with the jackknife

Consistent patterns in the results are:

- The percentage bias decreases with sample size and reliability.
- The jackknife reduces bias almost always if $\theta = 0.1$ and $n = 30$ or $n = 50$, and usually if $n = 10$.
- If $\theta = 1.0$ the jackknife generally increases bias in M1, decreases bias only if $n = 50$ or for error variances (all sample sizes) in M2, and decreases bias for all parameter estimators in M3.
- Where bias reduction exists, the increase in the variance of jackknife parameter estimates compared to maximum likelihood estimates is small for $n \geq 30$.

Some jackknife estimates are even more atypical than the maximum likelihood estimates. For example, if $n = 10$ and $\theta = 0.1$ in M1, the maximum and minimum estimates for $\theta_{3,3}$ are -0.110 and 0.464 (Figure 3.4) but if $\theta = 1.0$ the maximum and minimum estimates are -643 and 5910 (Figure 3.5).

3.3.3 Convergence and improper estimates

In the first two rows of Table 3.3 (M1) all solutions converged because these

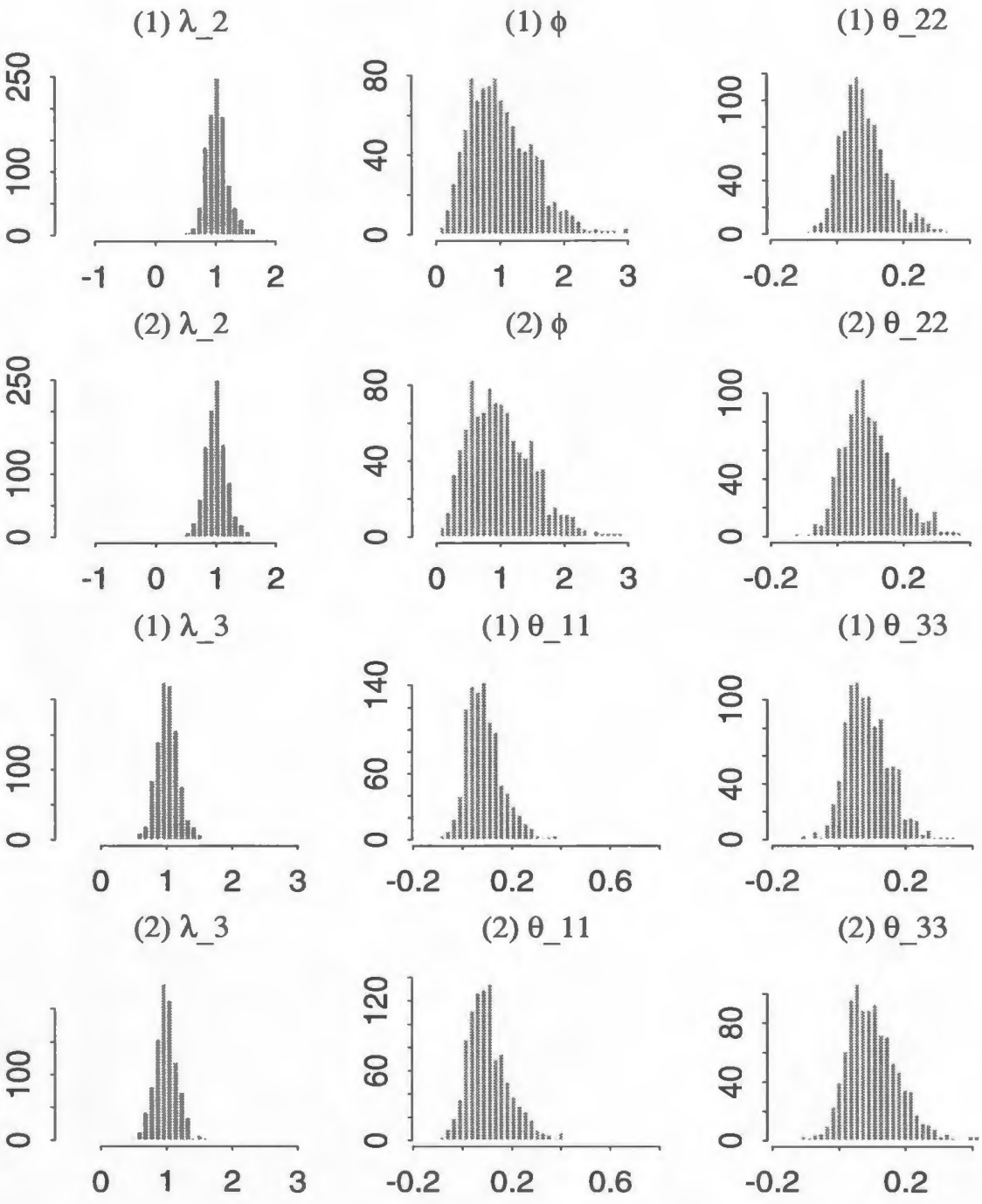

Figure 3.4. Frequency histograms of the simulated maximum likelihood (1) and jackknife (2) parameter estimates of M1. Simulated data was generated with $\theta = 0.1$ and $n = 10$.

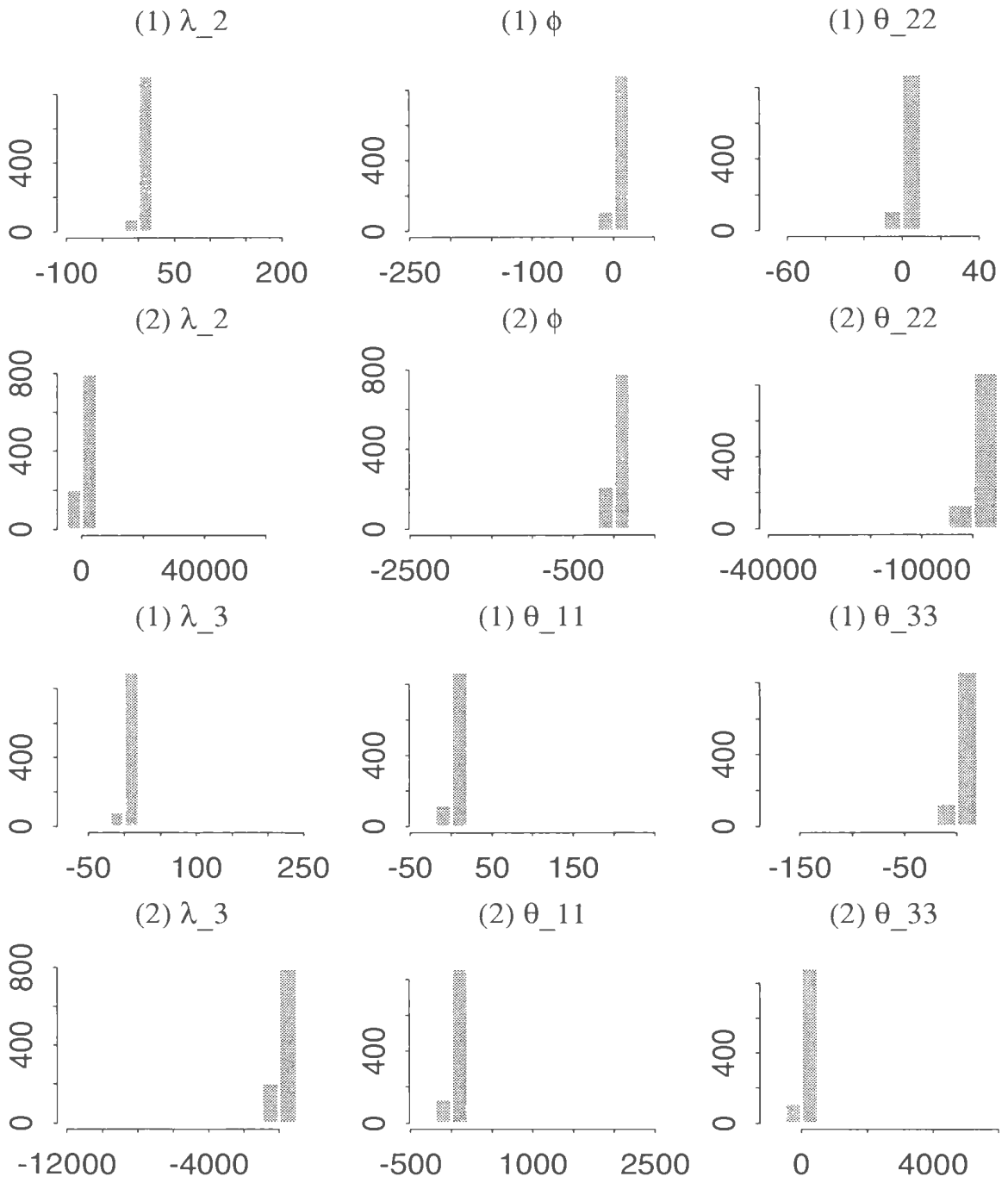

Figure 3.5. Frequency histograms of the simulated maximum likelihood (1) and jackknife (2) parameter estimates of M1. Simulated data was generated with $\theta = 1.0$ and $n = 10$.

estimates come from closed form expressions; however, this part of the table is included for consistency. Iterative solutions in Tables 3.6 (M2) and 3.9 (M3) do not converge only when the sample size is small ($n = 10$) and reliability is low ($\theta = 1.0$). The percentage of solutions converged is always the same for maximum likelihood estimates and jackknife estimates; they could only differ if a full sample solution converged but all subsampled solutions did not.

Improper solutions (negative estimates of error variances (θ's) and latent variable variances (ϕ's)) are more frequent with less reliable surveys and small sample sizes. More improper estimates occur in M3 than in M2 (more variances estimated) but this is not the case for M2 and M1. Jackknifing produces more improper estimates than maximum likelihood estimates in M2 and M3 but not M1. This is consistent with the results in Boomsma (1986).

3.4 Discussion

The primary outcome in this chapter is a quantification of the potential bias in small sample maximum likelihood estimates of the parameters in some confirmatory factor models and also whether the jackknife provides bias reduced estimates. A secondary outcome is a quantification of the occurrence of improper (negative) estimates of model variances.

The biases are small ($< 20\%$) when indicators are reliable, even if $n = 10$, and this may be acceptable. The reliability of an observed variable is the correlation of the observed variable and its corresponding latent variable. For M1, the reliability of x_i is given by

$$(1 + \frac{\theta_i}{\phi\lambda_i^2})^{-1}.$$

If ϕ and all λ's equal one then varying θ as 0.1 or 1.0 results in reliabilities

of 0.5 or 0.91. If, instead, ϕ and the θ's are fixed as ones then the same reliabilities could be achieved by taking $\lambda_i = 1$ or $\sqrt{10}$. The results should be similar in models with equivalent indicator reliabilities even if true parameters differ. The results may not extend to all latent variable models; for example, the simulations indicate the bias in λ's is positive for M1 and M2 but not for λ_5 $, \ldots . \lambda_8$ in M3.

The jackknife consistently reduces bias only when the reliability is high. When the reliability is low the presence of a small percentage of very unreasonable estimates has a large effect on the simulated bias approximation, and this is true for both maximum likelihood and jackknife estimates. The presence of unreasonable estimates in M1 is related to unbounded expectations of the parameter estimators. Division by zero or near zero sample covariances in the formulas in section 3.2 results in extreme estimates. Bollen (1989), pp. 250, mentions the problem of zero sample covariances for this same model. Kenny (1979), pp. 40, refers to this problem as empirical underidentification and suggests that it is often the result of two few observed variables (indicators).

The theory of jackknifing does not apply for M1 and it seems that any estimation procedure based on subsampling data will not work because of the increased chance of getting unreasonable estimates. The estimation of covariances improves as the sample size increases, and the frequency of occurrence of unreasonable values decreases; however, the expectations are still unbounded for all finite sample sizes. The problem is not unique to M1 either, many examples where zero sample covariances yield undefined estimates are given in Bollen (1989) and also Jöreskog (1978).

Atypical values occur in simulations of M2 and M3 though not as extreme as for M1. Whether the maximum likelihood estimators are bounded for M2 and M3 has not been shown. An analysis of some simulation iterations with

atypical estimates reveal that the resulting estimated population covariance matrix is reasonable as are goodness of fit statistics. If a parameter estimate is extreme it is usually multiplied by another estimate that is near zero in the population covariance matrix so that the estimate of a sample covariance is reasonable. In these cases an analysis of the parameter estimates correlation matrix (computed from the linearized covariance matrix of maximum likelihood estimates defined in the next chapter) reveals a high (> 0.99) correlation between the large and near zero parameters. Excluding these cases and also cases where the 95% confidence intervals for negative estimates of variance parameters do not contain zero removes many of the unreasonable estimates, but not all of them.

The problem of atypical estimates has been noted by Kelly (1984) for the errors in variables problem (a type of latent variable model) where "extreme values" of parameters estimates occurred in bootstrap replications causing the estimates of standard errors to be positively biased when $n = 20$. Boomsma (1986) also found that both the jackknife and the bootstrap were substantially influenced by "extreme outliers due to improper solutions" when $n = 25$ and the reliability of observed variables was either 0.36 or 0.64. Further research into the problems of unbounded estimates and methods for dealing with them is necessary. For example, a statistic that indicates that a data set has a potential to produce seriously biased estimates would be useful.

A serious bias in estimating the variance of the maximum likelihood estimate for λ_5 compared to the other λ's exists in M3. The results are not reported because of the uncertainty about the impact of atypical estimates; however, the bias for variances of parameter estimators are generally large and negative for sample sizes less than 50, even with reliable observed variables. The exception is the bias in the linearized variance approximation of

the maximum likelihood estimate of λ_5 in M3, it is very large and increases with sample size. The true variance of λ_5 (obtained from the simulations) is about 40 times less than the variance for the other λ's when $n = 50$ and $\theta = 0.1$; however, the average variance approximation of $\hat{\lambda}_5$ from all simulations does not differ from the other λ's, and is identical when the known (those used in generating simulated data) parameters are used to compute the covariances of the maximum likelihood estimates. The jackknife variance approximation ($\tilde{V}(\hat{\boldsymbol{\theta}})$) seems to work better because the simulated average is about 20 times less for λ_5 than for the other λ's.

The analysis of the jackknife has not provided sufficient evidence to warrant its application to the data sets in Chapter 2. The jackknife reduces bias in maximum likelihood estimates only when the maximum likelihood estimates have little bias, at least for the confirmatory factor models considered here

Chapter 4

Bias reduction by Taylor's series

4.1 Introduction

The results in Chapter 3 suggest that the jackknife does not reduce bias in maximum likelihood estimates of the parameters of some simple structural equation models sufficiently to warrant its use, although this problem may be related more to the estimators than the jackknife itself, hence an alternative method is investigated. A parametric approach for estimating the bias in maximum likelihood estimates, based on a Taylor's series expansion of the loglikelihood, is presented in Cox and Snell (1968) and in Cox and Hinkley (1974). The methodology has been applied to nonlinear regression models with normal errors (Box 1971; Cook *et al.* 1986), generalized linear models (Cordeiro and McCullagh 1991) and nonlinear exponential family regression models (Paula 1992). Cordeiro (1993) presents 2 examples involving multiplicative heteroscedastic regression models.

The bias reduction method has not been applied to structural equation models, therefore it is developed first generally, and then for structural equation models. The method is tested with the results from the simulation study of M2.

4.2 Methods

4.2.1 First order Taylor's series approximation of bias

Much of the statistical theory developed for structural equation models comes from the large sample statistical theory of general maximum likelihood estimates which is based on a Taylor's series expansion of the loglikelihood function. Let L denote the loglikelihood function where $L = L(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}$ is a $p \times 1$ vector of parameters to estimate. The notation used is similar to Cox and Hinkley's (1974). Let $u_i(\boldsymbol{\theta}) = \frac{\partial L}{\partial \theta_i}$ and $\mathbf{u}(\boldsymbol{\theta}) = \frac{\partial L}{\partial \boldsymbol{\theta}}$. The first order Taylor's series approximation of $\mathbf{u}(\hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}$ is

$$\mathbf{u}(\hat{\boldsymbol{\theta}}) \approx \mathbf{u}(\boldsymbol{\theta}) + \frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

Let $\mathcal{I} = -\mathrm{E}\left[\frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\boldsymbol{\theta}'}\right] = \mathrm{E}[\mathbf{u}(\boldsymbol{\theta})\mathbf{u}(\boldsymbol{\theta})']$; $\mathcal{I}$ is known as Fisher's information matrix. Substituting $\mathcal{I}$ for $\frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\boldsymbol{\theta}'}$ gives

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \approx \mathcal{I}^{-1}\mathbf{u}(\boldsymbol{\theta}),$$

where $\mathbf{u}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$; that is, because $\hat{\boldsymbol{\theta}}$ is a maximum likelihood estimate which is the solution of $\mathbf{u}(\boldsymbol{\theta}) = \mathbf{0}$. The maximum likelihood estimate of $\boldsymbol{\theta}$ is unbiased using the first order approximation because the $\mathrm{E}(\mathbf{u}(\boldsymbol{\theta})) = \mathbf{0}$. A variance approximation of $\boldsymbol{\theta}$ (refered to as the linearized variance estimator) is

$$\mathrm{E}\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\right] \approx \mathcal{I}^{-1}.$$

The bias approximation is o(1); that is, as $n \to \infty$ the bias $\to 0$ (Cox and Hinkley 1974, 294). Consequently, maximum likelihood estimates are generally unbiased and the variance approximation exact only as the sample size approaches infinity. The finite sample results developed in Chapters 2 and 3 also suggest that the estimators for structural equation models are biased, hence a better approximation is developed.

4.2.2 Second order Taylor's series approximation of bias

The same technique is used as in the previous section; however, the Taylor's series expansion has one more term (second order approximation) which gives

$$\mathbf{0} = \mathbf{u}(\hat{\boldsymbol{\theta}}) \approx \mathbf{u}(\boldsymbol{\theta}) + \begin{bmatrix} \frac{\partial u_1(\boldsymbol{\theta})}{\boldsymbol{\theta}'} \\ \frac{\partial u_2(\boldsymbol{\theta})}{\boldsymbol{\theta}'} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial u_p(\boldsymbol{\theta})}{\boldsymbol{\theta}'} \end{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + 1/2 \begin{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \frac{\partial u_1(\boldsymbol{\theta})}{\boldsymbol{\theta}\boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \frac{\partial u_2(\boldsymbol{\theta})}{\boldsymbol{\theta}\boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ \cdot \\ \cdot \\ \cdot \\ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \frac{\partial u_p(\boldsymbol{\theta})}{\boldsymbol{\theta}\boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \end{bmatrix},$$

or

$$\mathbf{0} \approx \mathbf{u}(\boldsymbol{\theta}) + \dot{\mathbf{u}}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + 1/2 \sum_{j=1}^{p} \mathcal{I}\boldsymbol{\iota}^j \frac{\partial \mathrm{vec}(\dot{\mathbf{u}}(\boldsymbol{\theta}))'}{\partial \theta_i} \mathrm{vec}\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\right).$$

where $\boldsymbol{\iota}^j$ is the jth column of $\mathcal{I}^{-1}$ and $\dot{\mathbf{u}}(\boldsymbol{\theta}) = \frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\boldsymbol{\theta}'} = \frac{\partial^2 L}{\partial \boldsymbol{\theta}\boldsymbol{\theta}'}$.

Taking expectations yields

$$\begin{aligned} \mathbf{0} \approx\ & \mathrm{E}(\dot{\mathbf{u}}(\boldsymbol{\theta}))b(\hat{\boldsymbol{\theta}}) + \mathrm{COV}(\dot{\mathbf{u}}(\boldsymbol{\theta})\hat{\boldsymbol{\theta}}) + 1/2 \sum_{j=1}^{p} \mathcal{I}\boldsymbol{\iota}^j \left[\mathrm{E}\left(\frac{\partial \mathrm{vec}(\dot{\mathbf{u}}(\boldsymbol{\theta}))'}{\partial \theta_i} \right) \mathrm{vec}(\mathcal{I}^{-1}) \right. \\ & \left. + \mathrm{COV}\left(\frac{\partial \mathrm{vec}(\dot{\mathbf{u}}(\boldsymbol{\theta}))'}{\partial \theta_i} \mathrm{vec}\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\right) \right) \right]. \end{aligned}$$

$b(\hat{\boldsymbol{\theta}})$ is the bias in $\hat{\boldsymbol{\theta}}$. The covariance terms are evaluated as follows

$$\begin{aligned} \mathrm{COV}(\dot{\mathbf{u}}(\boldsymbol{\theta})\hat{\boldsymbol{\theta}}) &= \mathrm{E}[(\dot{\mathbf{u}}(\boldsymbol{\theta}) + \mathcal{I})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]. \\ &\approx \mathrm{E}[\dot{\mathbf{u}}(\boldsymbol{\theta})\mathcal{I}^{-1}\mathbf{u}(\boldsymbol{\theta})]. \end{aligned}$$

The $\approx$ approximation comes from the substitution of the first order approximation $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \mathcal{I}^{-1}\mathbf{u}(\boldsymbol{\theta})$. The last COV is $\mathrm{o}(n^{-1})$ (Cox and Hinkley 1974, 309).

Using these covariances yields

$$\mathbf{0} \approx -\mathcal{I}b(\hat{\boldsymbol{\theta}}) + \mathrm{E}[\dot{\mathbf{u}}(\boldsymbol{\theta})\mathcal{I}^{-1}\mathbf{u}(\boldsymbol{\theta})] + 1/2 \sum_{j=1}^{p} \mathcal{I}\boldsymbol{\iota}^j \mathrm{E}\left(\frac{\partial \mathrm{vec}(\dot{\mathbf{u}}(\boldsymbol{\theta}))'}{\partial \theta_i} \right) \mathrm{vec}(\mathcal{I}^{-1}).$$

Hence the bias estimator is

$$b(\hat{\boldsymbol{\theta}}) \approx \mathcal{I}^{-1}\mathrm{E}[\dot{\mathbf{u}}(\boldsymbol{\theta})\mathcal{I}^{-1}\mathbf{u}(\boldsymbol{\theta})] + 1/2\sum_{j=1}^{p}\boldsymbol{\iota}^{j}\mathrm{E}\left(\frac{\partial \mathrm{vec}(\dot{\mathbf{u}}(\boldsymbol{\theta}))'}{\partial\theta_i}\right)\mathrm{vec}(\mathcal{I}^{-1}).$$

$$\approx \mathcal{I}^{-1}\left\{\sum_{j=1}^{p}\mathrm{E}[u_i(\boldsymbol{\theta})\dot{\mathbf{u}}(\boldsymbol{\theta})]\boldsymbol{\iota}^{j}\right\} + 1/2\sum_{j=1}^{p}\mathrm{tr}\left(\mathcal{I}^{-1}\mathrm{E}\left[\frac{\partial\dot{\mathbf{u}}(\boldsymbol{\theta})}{\partial\theta_i}\right]\right)\boldsymbol{\iota}^{j}. \quad (4.1)$$

This estimator is $o(n^{-1})$ compared to $o(1)$ for the first order approximation.

4.2.3 Bias in structural equation models

The following results are required for the second order Taylor's series bias approximation (4.1)

- $\mathcal{I}^{-1}$,
- $\mathrm{E}[u_i(\boldsymbol{\theta})\dot{\mathbf{u}}(\boldsymbol{\theta})]$,
- $\mathrm{E}\left[\frac{\partial\dot{\mathbf{u}}(\boldsymbol{\theta})}{\partial\theta_i}\right]$.

These quantities are developed in **Appendix B**.

The algorithm used to compute the bias is as follows. Let

$$\begin{aligned}
\mathbf{B}_{1,j} &= n\boldsymbol{\Delta}'\mathbf{W}(\mathbf{I}\otimes\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Delta},\\
\mathbf{B}_{2,j} &= \frac{n}{2}\dot{\boldsymbol{\Delta}}_j'\mathbf{W}\boldsymbol{\Delta},\\
\mathbf{B}_{3,j} &= \frac{n}{2}[\dot{\boldsymbol{\Delta}}_1',\ldots,\dot{\boldsymbol{\Delta}}_p']\mathbf{W}\boldsymbol{\delta}_j.
\end{aligned}$$

where $\boldsymbol{\Delta}, \mathbf{W}, \dot{\boldsymbol{\Delta}}_j$ and $\boldsymbol{\delta}_j$ are all defined in **Appendix B**. The bias formula is

$$\begin{aligned}
b(\hat{\boldsymbol{\theta}}) &= \mathcal{I}^{-1}\left\{\sum_{j=1}^{p}(\mathbf{B}_{3,j}-\mathbf{B}_{1,j})\boldsymbol{\iota}^{j}\right\}\\
&\quad + 1/2\sum_{j=1}^{p}\mathrm{tr}\left(\mathcal{I}^{-1}(2\mathbf{B}_{1,j}-\mathbf{B}_{2,j}-\mathbf{B}_{2,j}'-\mathbf{B}_{3,j})\right)\boldsymbol{\iota}^{j}. \quad (4.2)
\end{aligned}$$

4.3 Results

4.3.1 Evaluation of the bias approximation

The accuracy of the bias approximation (1.2) for the maximum likelihood estimates of M2 can be tested using the average simulated bias obtained in Chapter 3. The technique is not used for M1 because the maximum likelihood estimates are unbounded and the methodology does not apply. It is not used for M3 because the computation of derivatives is very difficult. The test with M2 suffices in a preliminary analysis.

The results are presented in Table 4.1 (note that they are multiplied by 100). The Taylor's series bias approximation is based on the known parameters used in generating the simulated data and is not an estimate. The bias approximation corresponds very well with the simulated estimates in terms of the sign and magnitude of the bias, this indicates that the second order Taylor's series expansion is a good approximation.

4.3.2 Bias corrected estimate

The analysis in the previous section requires known parameters values but in practise these values are unknown. A maximum likelihood estimate of the bias can be obtained using the estimates of the parameters and the invariance property of maximum likelihood estimators (Bain and Engelhardt 1987). A bias corrected estimate is formed by subtracting the estimated bias from the parameter estimate. The procedure is tested using the simulated estimates of M2.

The results are presented in Table 4.2 and should be analyzed in conjunction with Tables 3.4-3.6. Some patterns in the results are:

- The percentage bias of the Taylor's series biased reduced estimator usually decreases with sample size and reliability, although small increases

sometimes occur.

- Bias is reduced for all realizations of M2, except when $\theta = 1.0$ and $n = 10$ and occasionally when $n = 50$ in which case the percentage bias is very small.

- When bias reduction does occur the increase in sampling variability of the bias reduced estimate compared to the maximum likelihood estimate is small and is sometimes less by as much as 50%.

- The Taylor's series bias reduced estimator has slightly less bias than the jackknife estimator, except when $\theta = 1.0$ and $n = 10$. Their biases and variance ratio's are nearly identical when $n = 50$.

- The percentage of converged solutions with positive variances are greater for the biased reduced estimator than for the maximum likelihood estimator.

The presence of a atypical simulated estimates is again problematic. For example, when $n = 10$ and $\theta = 0.1$ the maximum and minimum biased reduced estimates for θ_1 are -0.075 and 0.486. When $n = 10$ and $\theta = 1.0$ the maximum and minimum estimates are -76.8 and 447.0.

4.4 Discussion

The Taylor's series bias reduced estimator consistently reduces bias with little increase (and sometimes a decrease) in sampling variability except when $n = 10$ and observed variables are not highly correlated with model factors (reliability = 0.5). Another advantage of this estimator is that is reduces the occurrence of improper (negative) estimates of population variances. The biased reduced estimator is often more efficient than the jackknife (at least for M2) in that it has less bias and lower sampling variability when n is small

except with low indicator reliability; however, the two techniques are approximately equivalent when $n = 50$. Akahira (1983) found that the biased-adjusted maximum likelihood estimator is asymptotically equivalent to the jackknife in the variance, which corresponds with the results here. The above findings may not apply to every structural equation model, so analyses of different types of models would be useful to test this.

The conclusion from this analysis is that the Taylor's series biased reduced estimator works slightly better than the jackknife in reducing bias; however, like the jackknife, reduction in bias only occurs when the bias in maximum likelihood estimates is small. It seems likely that the atypical estimates have an effect here. For this reason the Taylor's series biased reduced estimator does not seem to be a viable procedure for reducing bias in the maximum likelihood estimates of the parameters in structural equation models. Further research dealing with atypical estimates may result in a feasible procedure. A better approach may be to consider a bias reduced estimation scheme such as that developed by Firth (1993) rather than merely correcting the bias of maximum likelihood estimates.

Bibliography

Akahira, M. 1983. Asymptotic deficiency of the jackknife estimator. Austral. Jor. Statist. 25: 123-129.

Anon. 1991. Report of the twelfth Northeast regional stock assessment workshop. Northeast Fisheries Science Reference Document 91-03, Woods hole, MA. 02543, USA.

Austin, J. T. and L. M. Wolfle. 1991. Annotated bibliography of structural equation modelling: Technical work. Brit. Jor. Math. Stat. Psy. 44: 93-152.

Bain, L. J. and M. Engelhardt. 1987. Introduction to probability and mathematical statistics. United States of America: Wadsworth, Inc., 565 pp.

Baird, J. W., C. A. Bishop, and E. F. Murphy. 1992. An assessment of the cod stock in NAFO Div. 3NO. NAFO SCR Doc. 92/75. Ser. No. N2130. 28 pp.

Bentler, P. M. 1989. EQS structural equations program manual. Los Angeles: BMDP Statistical Software.

Bentler, P. M. and C. Chou. 1987. Practical issues in structural equation modeling. Soc. Meth. Res. 16: 78-117.

Beverton, R.J.H. and T. C. Iles. 1991a. Mortality rates of O-group plaice (*Pleuronectes platessa* L.), dab (*Limanda limanda* L.) and turbot (*Scophthalmus maximus* L.) in European waters. II. Comparison of mortality rates and construction of life-table for O-group plaice. Neth. J. Sea Res. 29: 49-59.

Beverton, R.J.H. and T. C. Iles. 1991b. Mortality rates of O-group plaice (*Pleuronectes platessa* L.), dab (*Limanda limanda* L.) and turbot (*Scophthalmus maximus* L.) in European waters. III. Density-dependence of mortality rate in O-group plaice, and some demographic implications. Neth. J. Sea Res. 29: 61-79.

Bollen, K. A. 1989. Structural equations with latent variables. New York: Wiley Inc., 541 pp.

Bollen, K. A. and J. S. Long. 1993. Testing structural equation models. Newbury Park: SAGE Publications, Inc., 320 pp.

Boomsma, A. 1985. Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. Psychometrika 50: 229-242.

Boomsma, A. 1986. On the use of bootstrap and jackknife in covariance structure analysis, pp. 205-210. *In* F. De Antoni, N. Lauro and A. Rizzi [ed.] COMPSTAT 1986. Proceedings in computational statistics. Heidelberg: Physica-Verlag, 512 pp.

Box, M. J. 1971. Bias in nonlinear estimation. JRSS B 32: 171-201.

Browne, M. W. and A. Shapiro. 1988. Robustness of normal theory methods in the analysis of linear latent variable models. Brit. Jor. Math. Stat. Psy. 44: 193-208.

Cook, R. D. and S. Weisberg. 1982. Residuals and influence in regression. New York: Chapman and Hall, 230 pp.

Cook, R. D., C.-L. Tsai and B. C. Wei. 1986. Bias in nonlinear regression. Biometrika 73: 615-623;

Cordeiro, G. M. and P. McCullagh. 1991. Bias correction in generalized linear models. JRSS B 53: 629-643.

Cordeiro, G. M. 1993. Bartlett corrections and bias correction for two heteroscedastic regression models. Commun. Stat. - Theory Meth. 22: 169-188.

Cox, D. R. and E. J. Snell. 1968. A general definition of residuals. JRSS B 30: 248-275.

Cox, D. R. and D. V. Hinkley. 1974. Theoretical statistics. New York: Chapman and Hall, 511 pp.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. Philadelphia: SIAM, 92 pp.

Ethington, C. A. 1986. The robustness of LISREL estimates in structural equation models with categorical variables. Jor. Exp. Educ. 56: 80-88.

Firth, D. 1993. Bias reduction of maximum likelihood estimates. Biometrika 80: 27-38.

Frangos, C. C. 1987. An updated bibliography on the jackknife method. Commun. Stat. - Theory Meth. 16: 1543-1584.

Gerbing, D. W. and J. C. Anderson. 1985. The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. Mult. Beh. Res. 20: 255-271.

Gerbing, D. W. and J. C. Anderson. 1987. Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. Psychometrika 52: 99-111.

Geweke, J. F. and K. J. Singleton. 1980. Interpreting the likelihood ratio statistic in factor models when the sample size is small. JASA 75: 133-137.

Graham, A. 1981. Kronecker products and matrix calculus: with applications. New York: Wiley Inc., 130 pp.

Hilborn, R. and C. J. Walters. 1992. Quantitative fisheries stock assessment. New York: Chapman and Hall, 570 pp.

Hinkley, D. V. 1977. Jackknifing in unbalanced situations. Technometrics 19: 285-292.

Hinkley, D. V. 1978. Improving the jackknife with special reference to correlation estimation. Biometrika 65: 13-21.

Hinkley, D. V. and H.-L. Wang. 1980. A trimmed jackknife. JRSS B 42: 347-356.

Homer, P. and R. M. O'Brien. 1988. Using LISREL models with crude rank category measures. Qual. Quant. 22: 191-201.

ICES 1992a. Report of the roundfish working group. ICES, Doc. C.M. 1992/Assess: 4.

ICES 1992b. Report of the Irish Sea and Bristol Channel working group. ICES, Doc. C.M. 1992/ Assess: 1.

ICES 1992c. Report of the Arctic fisheries working group. ICES, Doc. C.M. 1992/Assess: 3.

Jöreskog, K. G. 1973. Analysis of covariance structures. In P. R. Krishnaiah (Ed.), Multivariate Analysis-III. New York: Academic Press, Inc., pp. 263-285.

Jöreskog, K. G. 1990. New developments in LISREL: analysis of ordinal variables using polychoric correlations and weighted least squares. Qual. Quant. 24: 387-404.

Kelly, G. 1984. The influence function in the errors in variables problem. Ann. Statist. 12: 87-100.

Kenny, D. A. 1979. Correlation and causality. New York: John Wiley and Sons, Inc., 277 pp.

Kuzmin, S. A. 1992. Assessment of the cod stock in NAFO Subarea 3 by the data from the 1990 trawl-acoustic survey. NAFO SCR Doc. 92/13. Ser. No. N2055. 11 pp.

Lee, S. Y. and R. I. Jennrich. 1979. A study of algorithms for covariance structure analysis with specified comparisons using factor analysis. Psychometrika 44: 99-113.

Miller, R. G. 1964. A trustworthy jackknife. Ann. Math. Statist. 35: 1594-1605.

Miller, R. G. 1974. The jackknife - a review. Biometrika 61: 1-15.

Myers, R. A., W. Blanchard, and K. R. Thompson. 1990. Summary of north Atlantic fish recruitment 1942-1987. Can. Tech. Rep. of Fisheries and Aquatic Sciences, No 1743. 108 pp.

Myers, R. A. and N. G. Cadigan. 1993a. Density-dependent juvenile mortality in marine demersal fish. Can. Jor. Fish. Aquat. Sci. *In Press.*

Myers, R. A. and N. G. Cadigan. 1993b. Is juvenile mortality in marine demersal fish variable?. Can. Jor. Fish. Aquat. Sci. *In Press.*

Mulaik, S. A., L. R. James, J. Van Alstine, N. Bennett, S. Lind and C. D. Stilwell. 1989. Evaluation of goodness-of-fit indices for structural equation models. Quant. Meth. Psy. 105: 430-445.

Muthén, B., D. Kaplan and M. Hollis. 1987. On structural equation modeling with data that are not missing completely at random. Psychometrika 52: 431-462.

Paula, G. A. 1992. Bias correction for exponential family nonlinear models. Jor. Statist. Comput. Simul. 40: 43-54.

Quenouille, M. H. 1956. Notes on bias in estimation. Biometrika 43: 353-360.

SAS Institute Inc. 1990. SAS/IML Software: Usage and Reference, Version 6, First Edition, Cary, NC. 501 pp.

SAS Institute Inc. 1990. SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC. 890 pp.

Satorra, A. 1990. Robustness issues in structural equation modelling: a review of recents developments. Qual. Quant. 24: 367-386.

Schucany, W. R., H. L. Gray and D. B. Owen. 1971. On bias reduction in estimation. JASA 66: 524-533.

Searle, S. R., G. Casella and C. E. McCulloch. 1992 Variance components. New York: Wiley Inc., 501 pp.

Sundby, S., H. Bjorke., A. V. Soldal and S. Olsen. 1989. Mortality rates during the early life stages and yearclass strength of north east Arctic

cod (*Gadus morhus* L). Rapp. P.-v Réun. Cons. perm. int. Explor. Mer. 191: 351-358.

Tukey, J. 1958. Bias and confidence in not quite large samples (abstract). Ann. Math. Stat. 29: 614.

Veer, H. W. van Der. 1986. Immigration, settlement and density- dependent mortality of a larval and early post-larval O-group (*Pleuronectes platessa*) population in the western Wadden Sea. Mar. Ecol. Prog. Ser. 29: 223-236.

Wheaton, B. 1987. Assessment of fit in overidentified models with latent variables. Soc. Meth. Res. 16: 118-154.

Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Ann. Stat. 14: 1261-1295.

Chapter 5

Appendix A Tables

Table 2.1. Data for the Irish Sea cod stock from the North Whales groundfish survey in October, March and June. Units are in numbers per 16 hours fished.

cohort	OCTOBER 0 yr. old's	MARCH 1 yr. old's	JUNE 1 yr. old's	OCTOBER 1 yr. old's
	s_1	s_2	s_3	s_4
1975	.	20	4	18
1976	21	54	20	48
1977	11	30	18	3
1978	19	292	51	10
1979	491	777	267	173
1980	51	77	212	98
1981	5	10	7	8
1982	113	240	31	29
1983	168	155	11	11
1984	36	132	53	17
1985	1	6	17	5
1986	458	686	210	412
1987	50	.	51	20
1988	1	.	7	2
1989	24	.	18	44
1990	1143	.	161	.

Table 2.2. Data for the North Sea cod stock from the international youngfish survey (IYFS) and English groundfish survey (EGFS). Units are in numbers per hour fished.

cohort	EGFS 0 yr. old's s_1	IYFS 1 yr. old's s_2	EGFS 1 yr. old's s_3	IYFS 2 yr. old's s_4
1970	.	98.3	.	34.5
1971	.	4.1	.	10.6
1972	.	38	.	9.5
1973	.	14.7	.	6.2
1974	.	40.3	.	19.9
1975	.	7.9	.	3.2
1976	.	36.7	62.7	29.3
1977	13.9	12.9	22.8	9.3
1978	12.6	9.9	24.2	14.8
1979	18.6	16.9	50.8	25.5
1980	10.2	2.9	11.4	6.7
1981	74.2	9.2	32.4	16.6
1982	2.5	3.9	15.4	8
1983	95.1	15.2	61.2	17.6
1984	0.4	0.9	4.3	3.6
1985	8.3	17	34.4	28.8
1986	1.2	8.8	14.2	6.1
1987	0.4	3.6	8.4	6.3
1988	16.8	13.1	22.8	15.2
1989	6	3.4	6.1	4.1
1990	3.9	2.4	7.5	.
1991	48.4	.	.	.

Table 2.3. Data for the Gulf of Maine cod stock from the U.S.A. Northeast Fisheries Center spring and fall surveys. Units are in numbers per tow.

cohort	SPRING 1 yr. old's s_1	FALL 1 yr. old's s_2	SPRING 2 yr. old's s_3	FALL 2 yr. old's s_4
1961	.	.	.	0.865
1962	.	0.416	.	0.078
1963	.	0.059	.	0.564
1964	.	0.545	.	0.41
1965	.	0.131	.	0.138
1966	.	0.083	0.791	0.115
1967	0.393	0.023	0.023	0.079
1968	0	0.038	0.079	0.17
1969	0.102	0.603	0.091	0.153
1970	0.016	0.114	0.098	0.78
1971	0.226	3.576	2.724	1.393
1972	0.022	0.21	0.036	0.121
1973	0.305	0.72	0.448	1.966
1974	0.06	0.094	0.195	0.134
1975	0.027	0.156	0.191	0.291
1976	0.016	0.018	0.067	0.301
1977	0.022	1.111	1.045	0.361
1978	0.343	0.236	0.357	2.111
1979	0.057	1.026	0.537	0.245
1980	0.623	0.397	0.827	2.014
1981	0.273	0.449	0.627	0.626
1982	0.401	1.064	0.662	0.27
1983	0.097	0.246	0.0952	0.364
1984	0.0112	0.1512	0.132	0.196
1985	0.1668	0.1204	0.2552	0.5296
1986	0.0196	0.2396	0.4212	0.898
1987	0.2652	0.7804	0.3288	1.2104
1988	0.0116	0.2104	0.0964	0.1856
1989	0	0.0148	.	.

Table 2.4. Data for the Georges Banks cod stock from the U.S.A. Northeast Fisheries Center spring and fall surveys. Units are in numbers per tow.

cohort	SPRING 1 yr. old's s_1	FALL 1 yr. old's s_2	SPRING 2 yr. old's s_3	FALL 2 yr. old's s_4
1962	.	.	.	0.778
1963	.	0.719	.	0.699
1964	.	0.64	.	0.998
1965	.	1.299	.	1
1966	.	1.693	.	1.334
1967	.	7.596	1.615	1.611
1968	0.136	0.314	0.546	0.622
1969	0.123	0.343	0.814	1.353
1970	0.381	1.688	0.819	0.632
1971	0.207	0.602	1.833	1.295
1972	2.902	7.443	11.644	6.07
1973	0.521	1.749	4.557	0.654
1974	0.446	0.409	0.378	0.421
1975	0.064	0.994	1.922	2.072
1976	1.301	6.148	3.527	3.424
1977	0.028	0.237	0.187	0.255
1978	0.376	1.855	1.359	1.717
1979	0.435	1.619	2.265	0.564
1980	0.039	0.818	1.916	2.25
1981	2.303	3.525	3.395	2.094
1982	0.488	0.875	1.967	1.022
1983	0.329	0.647	0.462	0.101
1984	0.402	2.496	2.633	0.803
1985	0.098	0.22	0.423	0.153
1986	0.871	2.28	1.612	1.353
1987	0.034	0.414	0.684	0.433
1988	0.7	0.903	1.334	1.03
1989	0.38	2.738	0.926	1.534
1990	0.194	0.362	0.499	.
1991	1.038	.	.	.

Table 2.5. Data for the Southern Grand Banks cod stock from the Canadian and Russian spring surveys. Units are in numbers per tow.

cohort	CANADA 1 yr. old's	CANADA 2 yr. old's	CANADA 3 yr. old's	RUSSIA 3 yr. old's
	s_1	s_2	s_3	s_4
1965	.	.	.	.
1966	.	.	.	.
1967	.	.	.	.
1968	.	.	25.79	.
1969	.	2.56	8.82	.
1970	0	1.15	2.38	.
1971	0.01	2.34	4.05	.
1972	0.06	1.13	4.16	.
1973	0.04	2.8	2.74	.
1974	0.41	3.69	9.41	23
1975	0.55	2.28	7.11	18.4
1976	0.01	0.71	2.33	3.8
1977	0.55	0.93	1.35	3.6
1978	3.09	5.29	4.89	6.6
1979	0.01	0.35	1.18	10
1980	0.32	9.35	.	12.4
1981	1.56	.	6.2	33.4
1982	.	3.28	4.47	64.3
1983	0.01	0.41	0.71	12.4
1984	0.01	0.7	2.84	2.1
1985	0.02	2.76	2.2	7.2
1986	0.21	1.66	1.9	1.7
1987	0.01	0.25	0.96	0.4
1988	0.02	0.47	1.24	1.8
1989	0.04	6.3	.	.
1990	0.02	.	.	.

Table 2.6. Data for the Northeast Arctic cod stock from the Norwegian and Russian surveys. Units for Russian surveys in numbers per hour fished. Units not available for Norwegian surveys.

cohort	RUSSIA 1 yr. old's s_1	NORWAY 1 yr. old's s_2	RUSSIA 2 yr. old's s_3	NORWAY 2 yr. old's s_4
1981	.	.	0.6	17.7
1982	3.7	259	8.9	366
1983	5.4	2170	9.2	647
1984	0.9	39	4.9	103
1985	5	562	2.2	387
1986	0.7	25.3	0.2	63.5
1987	0.01	3.8	0.2	12.7
1988	0.1	7.1	0.1	49.9
1989	0.4	122	2.4	213
1990	3.9	357	.	.

Table 2.7. Parameters estimates for the 1 factor model (SE's in parentheses). The data is for 6 cod stocks. N is the minimum length (in numbers of years) of a survey. $AGFI$ is the adjusted goodness of fit index and $p-value$ corresponds to the chi-square goodness of fit test. $P(\boldsymbol{\lambda} \leq \mathbf{1})$ is the joint probability that all $\lambda_i \leq 1$. $\rho_{s_i s_i}$ is the reliability of survey s_i.

	IRISH SEA	NORTH SEA	GULF OF MAINE	GEORGES BANKS	SOUTHERN GRAND BANKS	NORTH-EAST ARCTIC
N	12	15	23	24	15	9
$AGFI$	0.66	0.97	0.70	0.22	0.98	0.58
$p-value$	0.37	0.91	0.18	0.01	0.95	0.40
λ_1	1.00	1.00	1.00	1.00	1.00	1.00
	-	-	-	-	-	-
λ_2	0.75	0.73	1.40	1.01	0.64	1.03
	(0.16)	(0.22)	(0.60)	(0.25)	(0.24)	(0.17)
λ_3	0.56	0.69	1.43	1.04	0.49	0.81
	(0.15)	(0.17)	(0.60)	(0.24)	(0.20)	(0.16)
λ_4	0.63	0.53	0.84	0.85	0.73	0.67
	(0.16)	(0.14)	(0.40)	(0.22)	(0.29)	(0.14)
$\theta_{1,1}$	0.74	1.32	1.48	0.81	1.90	0.59
	(0.51)	(0.49)	(0.47)	(0.25)	(0.84)	(0.36)
$\theta_{2,2}$	0.47	0.45	0.72	0.24	0.31	0.28
	(0.30)	(0.17)	(0.32)	(0.09)	(0.20)	(0.26)
$\theta_{3,3}$	0.60	-0.02	0.29	0.06	0.36	0.45
	(0.29)	(0.05)	(0.25)	(0.06)	(0.17)	(0.26)
$\theta_{4,4}$	0.73	0.11	0.61	0.23	0.81	0.40
	(0.35)	(0.05)	(0.20)	(0.08)	(0.38)	(0.22)
ϕ	3.62	1.51	0.53	0.75	1.80	3.90
	(1.80)	(0.91)	(0.43)	(0.39)	(1.26)	(2.11)
$P(\boldsymbol{\lambda} \leq \mathbf{1})$	0.930	0.882	0.154	0.353	0.793	0.419
$\rho_{s_1 s_1}$	0.83	0.53	0.26	0.48	0.49	0.87
$\rho_{s_2 s_2}$	0.81	0.64	0.59	0.76	0.71	0.94
$\rho_{s_3 s_3}$	0.66	1.02	0.79	0.93	0.54	0.85
$\rho_{s_4 s_4}$	0.67	0.79	0.38	0.70	0.54	0.82

Table 2.8. Simulation results to test the robustness of the estimation model. The number of years of data used to estimate the model is n. D(.) denotes distribution of (.): N - Normal, G - Gamma. The minimum and maximum percentage biases for λ and θ are in parentheses. See text for details.

No.	n	$D(\xi)$	$D(\delta_i)$	θ/ϕ	$\theta_{1,1}/\theta$	mean bias $\lambda_2,\ldots,\lambda_4$	mean bias $\theta_1,\ldots,\theta_4$	bias ϕ	% conv. with var> 0
1	100	N	N	1	1	.013 (.001,.015)	-.020 (-.024,-.018)	.010	100
2	100	N	N	.2	1	.005 (.004,.006)	-.002 (-.003,-.001)	-.002	100
3	20	N	N	1	1	.186 (.149,.215)	-.080 (-.092,-.060)	.084	90.1
4	20	N	N	.2	1	.013 (.010,.017)	-.014 (-.015,-.013)	.002	98.3
5	10	N	N	1	1	.137 (.073,.212)	-.129 (-.218,-.053)	.313	59.2
6	10	N	N	.2	1	.022 (.017,.029)	-.021 (-.027,-.013)	.078	79.4
7	20	log(G)	N	1	1	.196 (.164,.247)	-.088 (-.103,-.072)	-.131	82.8
8	20	log(G)	N	.2	1	.013 (.011,.016)	-.012 (-.015,-.008)	-.185	98.3
9	20	N	G[1]	1	1	.145 (.136,.156)	-.076 (-.09,-.058)	.087	88.9
10	20	N	G[1]	.2	1	.017 (.013,.022)	-.011 (-.016,-.008)	.008	87.5
11	20	N	G[2]	1	1	.128 (.121,.140)	-.079 (-.090,-.070)	.104	87.3
12	20	N	G[2]	.2	1	.011 (.008,.015)	-.010 (-.011,-.008)	.027	86.2
13	20	N	N[3]	1	1	.142 (.133,.149)	-.069 (-.088,-.057)	.118	90.3
14	20	N	N[3]	.2	1	.015 (.009,.018)	-.008 (-.012,-.005)	.025	98.2
15	20	N	N	1	4	.216 (.206,.251)	-.127 (-.399,-.087)	.540	76.9
16	20	N	N	.2	4	.064 (.062,.066)	-.019 (-.054,-.004)	.042	95.3
17	20	N	N	1	9	-.141 (-.205,-.089)	-.250 (-.777,-.066)	.979	72.0
18	20	N	N	.2	9	.123 (.144,.132)	-.036 (-.110,-.007)	.144	93.8

note: 1. left skewed. 2. Right skewed. 3. correlated errors.

Table 3.1. Simulated percentage bias and variance ratio of maximum likelihood (^) and jackknife (~) estimates of the λ's and ϕ for the one factor - 3 variables per factor model (M1). The λ's and ϕ are always 1. $\theta_1, \ldots, \theta_3 = \theta$.

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
$\hat{\lambda}_2$	2.7	0.4	0.4	58.6	24.5	4.9
$\tilde{\lambda}_2$	-0.7	-0.0	0.1	7104.8	1222.7	-39.0
$V(\tilde{\lambda}_2)/V(\hat{\lambda}_2)$	1.2	1.0	1.0	31669.6	3818.5	674.3
$\hat{\lambda}_3$	1.6	0.6	0.3	50.6	15.1	4.6
$\tilde{\lambda}_3$	-1.1	0.1	0.0	-1206.6	-983.8	-10.7
$V(\tilde{\lambda}_3)/V(\hat{\lambda}_3)$	1.0	1.0	1.0	1236.5	34475.2	14.2
$\hat{\phi}$	3.3	2.4	0.2	-11.1	2.3	6.1
$\tilde{\phi}$	1.6	2.0	-0.0	-378.5	-609.6	-0.4
$V(\tilde{\phi})/V(\hat{\phi})$	1.0	1.0	1.0	104.5	1164.9	0.9

Table 3.2. Similar to 3.1 but for the θ's.

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
$\hat{\theta}_1$	-9.6	-2.3	-0.4	10.1	-0.5	-5.0
$\tilde{\theta}_1$	7.0	1.7	1.9	377.5	611.4	1.4
$V(\tilde{\theta}_1)/V(\hat{\theta}_1)$	1.4	1.1	1.0	105.9	1189.5	0.9
$\hat{\theta}_2$	-17.2	-3.4	-3.3	-12.7	-8.7	-6.0
$\tilde{\theta}_2$	-3.2	0.6	-1.0	-3905.3	-149.0	18.3
$V(\tilde{\theta}_2)/V(\hat{\theta}_2)$	1.3	1.1	1.0	73353.2	9730.6	195.4
$\hat{\theta}_3$	-12.3	-5.7	-2.6	-50.1	-12.6	-5.2
$\tilde{\theta}_3$	2.2	-1.8	-0.3	564.3	251.8	2.6
$V(\tilde{\theta}_3)/V(\hat{\theta}_3)$	1.3	1.1	1.0	579.7	13958.2	2.8

Table 3.3. Percentage converged and percentage converged with positive variances in simulations of maximum likelihood (1) and jackknife (2) solutions for the one factor - 3 variables per factor model (M1).

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
% (1) conv	100.0	100.0	100.0	100.0	100.0	100.0
% (2) conv	100.0	100.0	100.0	100.0	100.0	100.0
% (1) conv with +var	76.6	98.7	99.7	51.8	89.1	97.0
% (2) conv with +var	78.1	98.8	99.8	53.9	94.8	99.0

Table 3.4. Simulated percentage bias and variance ratio of maximum likelihood (^) and jackknife (˜) estimates of the λ's and ϕ for the one factor - 4 variables per factor model (M2). The λ's and ϕ are always 1. $\theta_1, \ldots, \theta_4 = \theta$.

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
$\hat{\lambda}_2$	1.7	0.4	0.0	11.8	7.1	4.2
$\tilde{\lambda}_2$	-1.8	-0.1	-0.2	-3.4	-9.7	-1.1
$V(\tilde{\lambda}_2)/V(\hat{\lambda}_2)$	1.6	1.0	1.0	25.4	4.1	0.7
$\hat{\lambda}_3$	1.6	0.6	0.1	18.6	6.9	4.4
$\tilde{\lambda}_3$	-2.0	0.1	-0.2	11.9	-8.1	-1.2
$V(\tilde{\lambda}_3)/V(\hat{\lambda}_3)$	1.8	1.0	1.0	12.3	2.1	0.6
$\hat{\lambda}_4$	1.4	0.6	0.3	13.9	6.7	4.9
$\tilde{\lambda}_4$	-1.9	0.1	0.1	-15.5	-11.7	-0.7
$V(\tilde{\lambda}_4)/V(\hat{\lambda}_4)$	1.4	1.0	1.0	15.2	6.2	0.7
$\hat{\phi}$	2.9	1.5	0.0	24.8	6.8	3.5
$\tilde{\phi}$	1.8	1.2	-0.2	-6.7	1.2	0.1
$V(\tilde{\phi})/V(\hat{\phi})$	1.0	1.0	1.0	4.4	1.3	1.0

Table 3.5. Similar to 3.4 but for the θ's.

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
$\hat{\theta}_1$	-9.8	-3.3	-3.6	-19.8	-5.1	-2.1
$\tilde{\theta}_1$	0.7	-0.1	-1.5	4.8	0.4	1.2
$V(\tilde{\theta}_1)/V(\hat{\theta}_1)$	2.6	1.1	1.1	6.8	1.8	1.1
$\hat{\theta}_2$	-9.7	-3.1	-1.7	-14.3	-5.2	-4.2
$\tilde{\theta}_2$	-0.5	0.4	0.3	0.8	2.3	-0.9
$V(\tilde{\theta}_2)/V(\hat{\theta}_2)$	2.3	1.1	1.1	5.1	3.6	1.1
$\hat{\theta}_3$	-8.1	-3.2	-2.3	-16.2	-1.0	-3.6
$\tilde{\theta}_3$	2.0	0.2	-0.3	-12.4	1.7	-0.4
$V(\tilde{\theta}_3)/V(\hat{\theta}_3)$	2.3	1.1	1.0	6.1	1.8	1.1
$\hat{\theta}_4$	-11.7	-5.8	-3.0	-16.2	-5.6	-2.2
$\tilde{\theta}_4$	-2.1	-2.6	-0.9	-3.7	1.9	1.0
$V(\tilde{\theta}_4)/V(\hat{\theta}_4)$	2.3	1.1	1.1	6.2	2.1	1.1

Table 3.6. Percentage converged and percentage converged with positive variances in simulations of maximum likelihood (1) and jackknife (2) solutions for the one factor - 4 variables per factor model (M2).

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
% (1) conv	100.0	100.0	100.0	90.7	100.0	100.0
% (2) conv	100.0	100.0	100.0	90.7	100.0	100.0
% (1) conv with +var	83.4	99.9	100.0	64.1	98.3	99.9
% (2) conv with +var	71.1	99.7	100.0	52.9	96.9	99.7

Table 3.7. Simulated percentage bias and variance ratio of maximum likelihood (^) and jackknife (~) estimates of the λ's and ϕ's for the two factor - 4 variables per factor model (M3). The λ's, $\phi_{1,1}$ and $\phi_{2,2}$ are always 1. $\phi_{1,2} = 0.9$; $\theta_1, \ldots, \theta_8 = \theta$.

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
$\hat{\lambda}_2$	1.4	0.4	0.5	9.2	5.1	3.0
$\tilde{\lambda}_2$	-1.1	-0.0	0.2	-0.5	-4.8	-1.1
$V(\tilde{\lambda}_2)/V(\hat{\lambda}_2)$	4.9	1.0	1.0	5.8	1.0	0.8
$\hat{\lambda}_3$	1.9	0.5	0.2	8.5	5.9	3.2
$\tilde{\lambda}_3$	-1.4	0.1	-0.1	1.6	-5.1	-0.8
$V(\tilde{\lambda}_3)/V(\hat{\lambda}_3)$	1.4	1.0	1.0	6.2	0.8	0.8
$\hat{\lambda}_4$	1.6	0.0	0.2	11.7	6.0	2.6
$\tilde{\lambda}_4$	-1.3	-0.4	-0.0	10.6	-4.4	-1.4
$V(\tilde{\lambda}_4)/V(\hat{\lambda}_4)$	1.7	1.0	1.0	6.1	1.0	0.8
$\hat{\lambda}_5$	-2.2	-0.6	-0.4	-16.3	-7.6	-4.4
$\tilde{\lambda}_5$	-2.0	0.2	0.0	-4.9	2.0	0.3
$V(\tilde{\lambda}_5)/V(\hat{\lambda}_5)$	15.6	6.3	4.2	31.0	6.6	1.6
$\hat{\lambda}_6$	0.2	0.1	0.2	-17.3	-3.8	-2.3
$\tilde{\lambda}_6$	-0.8	0.5	0.4	-16.3	1.2	-0.3
$V(\tilde{\lambda}_6)/V(\hat{\lambda}_6)$	3.0	1.1	1.0	20.0	2.0	1.1
$\hat{\lambda}_7$	-0.5	-0.2	0.1	-13.8	-3.3	-2.7
$\tilde{\lambda}_7$	-1.6	0.2	0.3	0.1	1.0	-0.6
$V(\tilde{\lambda}_7)/V(\hat{\lambda}_7)$	2.5	1.1	1.0	16.3	1.9	1.1
$\hat{\lambda}_8$	0.4	0.3	0.2	-14.2	-2.8	-2.0
$\tilde{\lambda}_8$	-0.5	0.7	0.4	-2.2	2.7	0.1
$V(\tilde{\lambda}_8)/V(\hat{\lambda}_8)$	2.8	1.1	1.0	15.1	1.9	1.1
$\hat{\phi}_{1,1}$	1.2	-1.0	-0.4	24.0	3.6	4.5
$\tilde{\phi}_{1,1}$	0.5	-1.3	-0.7	-2.0	-0.3	2.2
$V(\tilde{\phi}_{1,1})/V(\hat{\phi}_{1,1})$	1.1	1.0	1.0	3.2	1.1	1.0
$\hat{\phi}_{1,2}$	1.5	-0.6	-0.4	24.2	6.1	7.1
$\tilde{\phi}_{1,2}$	3.0	-1.2	-0.8	11.3	0.2	3.2
$V(\tilde{\phi}_{1,2})/V(\hat{\phi}_{1,2})$	1.1	1.0	1.0	4.1	1.2	1.0
$\hat{\phi}_{2,2}$	4.7	0.7	0.0	64.0	20.5	14.4
$\tilde{\phi}_{2,2}$	3.7	-1.1	-1.0	18.9	-2.2	2.0
$V(\tilde{\phi}_{2,2})/V(\hat{\phi}_{2,2})$	1.3	1.0	1.0	9.5	1.3	1.1

Table 3.8. Similar to 3.7 but for the θ's.

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
$\hat{\theta}_1$	-5.1	-2.7	-2.4	-20.0	-5.0	-2.5
$\tilde{\theta}_1$	1.2	0.8	-0.4	-5.2	-1.2	-0.0
$V(\tilde{\theta}_1)/V(\hat{\theta}_1)$	4.0	1.1	1.0	5.3	1.2	1.1
$\hat{\theta}_2$	-6.2	-3.5	-2.4	-8.7	-5.4	-3.1
$\tilde{\theta}_2$	-5.5	-0.4	-0.4	-3.6	-1.2	-0.6
$V(\tilde{\theta}_2)/V(\hat{\theta}_2)$	2.9	1.1	1.0	5.9	1.3	1.1
$\hat{\theta}_3$	-11.4	-4.1	-2.2	-7.5	-5.2	-2.1
$\tilde{\theta}_3$	-16.8	-0.7	-0.1	-6.0	-0.3	0.5
$V(\tilde{\theta}_3)/V(\hat{\theta}_3)$	2.8	1.1	1.0	4.6	1.2	1.1
$\hat{\theta}_4$	-11.3	-3.0	-0.3	-12.1	-6.4	-3.4
$\tilde{\theta}_4$	-11.7	0.3	1.8	-11.6	-2.7	-1.0
$V(\tilde{\theta}_4)/V(\hat{\theta}_4)$	3.1	1.1	7.5	1.4	1.0	
$\hat{\theta}_5$	-7.7	-3.1	-2.4	-10.6	-3.7	-2.3
$\tilde{\theta}_5$	-3.4	0.3	-0.3	4.8	0.4	0.3
$V(\tilde{\theta}_5)/V(\hat{\theta}_5)$	2.9	1.1	1.0	10.8	1.4	1.1
$\hat{\theta}_6$	-7.4	-3.0	-1.8	-7.2	-4.3	-3.3
$\tilde{\theta}_6$	-7.7	0.3	0.3	4.1	0.6	-0.7
$V(\tilde{\theta}_6)/V(\hat{\theta}_6)$	3.3	1.1	1.0	5.1	2.7	1.1
$\hat{\theta}_7$	-10.9	-3.1	-2.7	-9.4	-6.7	-1.0
$\tilde{\theta}_7$	-7.4	0.1	-0.6	-8.5	-2.5	1.4
$V(\tilde{\theta}_7)/V(\hat{\theta}_7)$	3.0	1.1	1.1	5.9	1.5	1.1
$\hat{\theta}_8$	-5.5	-3.3	-1.9	-10.7	-5.0	-2.7
$\tilde{\theta}_8$	-6.6	-0.0	0.0	-6.4	0.3	-0.2
$V(\tilde{\theta}_8)/V(\hat{\theta}_8)$	3.2	1.1	1.0	9.0	2.1	1.1

Table 3.9. Percentage converged and percentage converged with positive variances in simulations of maximum likelihood (1) and jackknife (2) solutions for the two factor - 4 variables per factor model (M3).

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
% (1) conv	99.9	100.0	100.0	83.8	100.0	100.0
% (2) conv	99.9	100.0	100.0	83.8	100.0	100.0
% (1) conv with +var	74.3	100.0	100.0	63.4	99.0	100.0
% (2) conv with +var	40.4	99.8	100.0	16.3	95.4	99.7

Table 4.1. Comparison of the simulated bias (top) and the Taylor's series approximation of bias (bottom) of maximum likelihood estimates for a one factor - 4 variables per factor model (M2). Numbers are multiplied by 100.

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
λ_2	1.65	0.40	0.02	11.80	7.09	4.25
	1.17	0.36	0.21	16.67	5.17	3.06
λ_3	1.63	0.58	0.06	18.64	6.87	4.38
	1.17	0.36	0.21	16.67	5.17	3.06
λ_4	1.43	0.55	0.33	13.87	6.69	4.87
	1.17	0.36	0.21	16.67	5.17	3.06
ϕ	2.87	1.49	0.00	24.81	6.77	3.45
	1.17	0.36	0.21	16.67	5.17	3.06
$\theta_{1,1}$	-0.98	-0.33	-0.36	-19.83	-5.09	-2.12
	-1.17	-0.36	-0.21	-16.67	-5.17	-3.06
$\theta_{2,2}$	-0.97	-0.31	-0.17	-14.28	-5.22	-4.16
	-1.17	-0.36	-0.21	-16.67	-5.17	-3.06
$\theta_{3,3}$	-0.81	-0.32	-0.23	-16.20	-4.04	-3.65
	-1.17	-0.36	-0.21	-16.67	-5.17	-3.06
$\theta_{4,4}$	-1.17	-0.58	-0.30	-16.15	-5.64	-2.19
	-1.17	-0.36	-0.21	-16.67	-5.17	-3.06

Table 4.2. Simulated percentage bias and variance ratio (compared to maximum
likelihood (^) of the Taylor's series biased reduced estimates
(˘) of the parameters for the one factor - 4 variables per factor
model (M2). The λ's and ϕ are always 1. $\theta_1, ..., \theta_4 = \theta$. The
percentage of solutions converged with positive variances is
given in the last row.

	θ					
	0.1			1.0		
	n			n		
	10	30	50	10	30	50
$\breve{\lambda}_2$	-0.0	0.0	-0.2	-221.0	-2.7	-0.1
$V(\breve{\lambda}_2)/V(\hat{\lambda}_2)$	0.9	1.0	1.0	760.8	0.5	0.7
$\breve{\lambda}_3$	-0.1	0.2	-0.2	-169.7	-2.5	-0.1
$V(\breve{\lambda}_3)/V(\hat{\lambda}_3)$	0.8	1.0	1.0	195.8	0.5	0.7
$\breve{\lambda}_4$	-0.3	0.2	0.1	-124.8	-2.8	0.4
$V(\breve{\lambda}_4)/V(\hat{\lambda}_4)$	0.8	1.0	1.0	233.5	0.5	0.7
$\breve{\phi}$	1.8	1.1	-0.2	-45.1	1.6	0.4
$V(\breve{\phi})/V(\hat{\phi})$	1.0	1.0	1.0	186.1	1.0	1.0
$\breve{\theta}_1$	0.6	0.2	-1.5	50.1	0.1	0.9
$V(\breve{\theta}_1)/V(\hat{\theta}_1)$	1.2	1.1	1.0	301.4	1.0	1.0
$\breve{\theta}_2$	0.7	0.4	0.4	12.9	0.0	-1.2
$V(\breve{\theta}_2)/V(\hat{\theta}_2)$	1.2	1.1	1.0	31.1	1.0	1.0
$\breve{\theta}_3$	2.4	0.3	-0.3	24.9	1.0	-0.6
$V(\breve{\theta}_3)/V(\hat{\theta}_3)$	1.2	1.1	1.0	27.7	1.0	1.0
$\breve{\theta}_4$	-1.5	-2.4	-0.9	17.9	-0.5	0.9
$V(\breve{\theta}_4)/V(\hat{\theta}_4)$	1.2	1.1	1.0	47.6	1.0	1.0
% conv with +var	83.8	99.9	100.0	69.8	99.3	99.9

Chapter 6

Appendix B Derivatives and expectations

6.1 Loglikelihood

The distributional assumption for the data is:

$$\mathbf{x}_1, ..., \mathbf{x}_N \quad iid \quad \mathbf{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ and MVN denotes the multivariate normal distribution. The analysis is in terms of deviations from the mean so there are only $n = N - 1$ independent deviates defined as follows. Let $\mathbf{z}_i = \mathbf{x}_i - \bar{\mathbf{x}}$, then

$$\mathbf{z}_1, ..., \mathbf{z}_n \quad iid \quad \mathbf{MVN}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Note that $n\mathbf{S} = \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i'$. The loglikelihood of $\mathbf{z}_1, ..., \mathbf{z}_n$ is

$$L_1 = -\frac{n}{2} log \mid \boldsymbol{\Sigma} \mid -\frac{n}{2} tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) + k_1,$$

where k_1 is a constant. The analysis is in terms of covariance structures so the loglikelihood can be expressed in terms of $\mathbf{S}$ - the sample covariance matrix. If $s_{i,j}$ is the (i,j)th element of $\mathbf{S}$ then the loglikelihood of $s_{1,1}, s_{1,2}, ..., s_{p,p}$ (from the Wishart distribution) is

$$L_2 = -\frac{n}{2} log \mid \boldsymbol{\Sigma} \mid + \frac{n-p-1}{2} log \mid \mathbf{S} \mid -\frac{n}{2} tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) + k_2,$$

where k_2 is a constant. Both loglikelihoods can be written as

$$L = \frac{n}{2} log \mid \mathbf{S}\boldsymbol{\Sigma}^{-1} \mid -\frac{n}{2} tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) + k.$$

where k is a constant. The constants k_1 and k_2 change for L_1 and L_2 but the derivatives with respect to parameters will be the same as for L.

6.2 Matrix derivatives

The chain rule is used for differentiating a function f of a matrix (Graham 1981). Let $a_{i,j}$ be the (i,j)th element of an $n \times n$ matrix $\mathbf{A}$ and $a_{i,j}$ is a function of b_s say. Then

$$\frac{\partial f(\mathbf{A})}{\partial b_s} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f(\mathbf{A})}{\partial a_{i,j}} \frac{\partial a_{i,j}}{\partial b_s}.$$

Other results required are (Graham 1981):

$$\begin{aligned} \frac{\partial tr(\mathbf{A})}{\partial \mathbf{A}} &= \mathbf{I}, \\ \frac{\partial log|\mathbf{A}|}{\partial \mathbf{A}} &= (\mathbf{A}^{-1})', \\ \frac{\partial \mathbf{A}^{-1}}{\partial a_{i,j}} &= -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial a_{i,j}} \mathbf{A}^{-1}. \end{aligned}$$

where $\mathbf{A}$ is symmetric matrix functional.

6.3 Other matrix results

1. If $\mathbf{A}$ is $m \times n$ and $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_n]$ then define the vec operator (Graham 1981) as

$$\text{vec}(\mathbf{A})_{mn \times 1} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ . \\ . \\ . \\ \mathbf{a}_n \end{bmatrix}.$$

2. If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $r \times s$ matrix then the Kronecker product (direct product or tensor product - Graham 1981) of $\mathbf{A}$ and $\mathbf{B}$ is

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m,1}\mathbf{B} & a_{m,2}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{bmatrix}.$$

$\mathbf{A} \otimes \mathbf{B}$ is a matrix of order $(mr \times ns)$.

3. $\mathrm{tr}(\mathbf{AB}) = \mathrm{vec}(\mathbf{A}')'\mathrm{vec}(\mathbf{B})$.

4. $\mathrm{vec}(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A})\mathrm{vec}(\mathbf{X})$.

6.4 First order derivatives

To compute the first order derivatives of $L = L(\boldsymbol{\Sigma}(\boldsymbol{\theta}))$ with respect to θ_i ($\frac{\partial L}{\partial \theta_i}$, i=1,...,p) make the substitution $\mathbf{U} = \mathbf{S}\boldsymbol{\Sigma}^{-1}$ in L. Using the standard matrix derivative results given in the previous section,

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{U}} &= -\frac{n}{2}(\mathbf{I} - (\mathbf{U}^{-1})') \quad \text{and} \\ \frac{\partial \mathbf{U}}{\partial \theta_i} &= -\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}. \end{aligned}$$

Using the chain rule gives

$$\frac{\partial L}{\partial \theta_i} = \frac{n}{2} tr\left\{\boldsymbol{\Sigma}^{-1}(\mathbf{S} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\right\}.$$

An alternative version of this formula is

$$\frac{\partial L}{\partial \theta_i} = \frac{n}{2}\frac{\partial \mathrm{vec}(\boldsymbol{\Sigma})'}{\partial \theta_i}(\boldsymbol{\gamma}_S - \boldsymbol{\gamma}),$$

where $\boldsymbol{\gamma} = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})$ and $\boldsymbol{\gamma}_S = \mathrm{vec}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1})$. Let $\boldsymbol{\Delta}_{p^2 \times p} = \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\theta}'}$, then

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = \frac{n}{2}\boldsymbol{\Delta}'(\boldsymbol{\gamma}_S - \boldsymbol{\gamma}).$$

$\frac{\partial L}{\partial \boldsymbol{\theta}}$ is the multivariate score function and is used to compute maximum likelihood estimates.

6.5 Second order derivatives

$\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ is computed as follows. Let $\dot{\boldsymbol{\Delta}}_i = \frac{\partial \text{vec}(\boldsymbol{\Sigma})}{\partial \theta_i \partial \boldsymbol{\theta}'}$ and $\mathbf{W} = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}$. First compute:

$$\begin{aligned} \frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\theta}'} &= -\mathbf{W}\boldsymbol{\Delta}, \\ \frac{\partial \boldsymbol{\gamma}_S}{\partial \boldsymbol{\theta}'} &= -\mathbf{W}[(\mathbf{I} \otimes \mathbf{S}\boldsymbol{\Sigma}^{-1}) + (\mathbf{S}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})]\boldsymbol{\Delta}. \end{aligned}$$

Then

$$\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \frac{n}{2}\left\{[\dot{\boldsymbol{\Delta}}_1'(\boldsymbol{\gamma}_S - \boldsymbol{\gamma}), ..., \dot{\boldsymbol{\Delta}}_p'(\boldsymbol{\gamma}_S - \boldsymbol{\gamma})] + \boldsymbol{\Delta}'\mathbf{W}(\mathbf{I}^* - [(\mathbf{I} \otimes \mathbf{S}\boldsymbol{\Sigma}^{-1}) - (\mathbf{S}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})]\boldsymbol{\Delta})\right\},$$

where $\mathbf{I}^*$ is a $p^2 \times p^2$ identity matrix Note that $\text{E}(\boldsymbol{\gamma}_S) = \boldsymbol{\gamma}$ so that

$$\mathcal{I} = -\text{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right) = \frac{n}{2}\boldsymbol{\Delta}'\mathbf{W}\boldsymbol{\Delta}.$$

The matrix $\mathcal{I}$ is known as Fisher's information matrix, the inverse of this matrix is a large sample approximation of the covariance matrix of the maximum likelihood estimates of $\boldsymbol{\theta}$.

6.6 Third order derivatives

To compute $\frac{\partial^3 L}{\partial \theta_k \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ rewrite $\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ as

$$\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \frac{n}{2}\left\{[\dot{\boldsymbol{\Delta}}_1'(\boldsymbol{\gamma}_S - \boldsymbol{\gamma}), ... \dot{\boldsymbol{\Delta}}_p'(\boldsymbol{\gamma}_S - \boldsymbol{\gamma})] - \boldsymbol{\Delta}'\mathbf{W}\mathbf{V}\boldsymbol{\Delta}\right\},$$

where $\mathbf{V} = (\mathbf{I} \otimes \mathbf{S}\boldsymbol{\Sigma}^{-1}) + (\mathbf{S}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}) - \mathbf{I}^*$. First compute

$$\frac{\partial \mathbf{V}}{\partial \theta_k} = -(\mathbf{I} \otimes \mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1}) - (\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}).$$

Let $\boldsymbol{\delta}_k$ be the kth column of $\boldsymbol{\Delta}$. Then

$$\begin{aligned}\frac{\partial^3 L}{\partial\theta_k\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= \frac{n}{2}\left\{\frac{\partial}{\partial\theta_k}[\dot{\boldsymbol{\Delta}}_1'(\boldsymbol{\gamma}_S-\boldsymbol{\gamma}),\ldots,\dot{\boldsymbol{\Delta}}_p'(\boldsymbol{\gamma}_S-\boldsymbol{\gamma})]-[\dot{\boldsymbol{\Delta}}_1',\ldots,\dot{\boldsymbol{\Delta}}_p']\mathbf{W}\mathbf{V}\boldsymbol{\delta}_k\right.\\ &\quad -\dot{\boldsymbol{\Delta}}_k'\mathbf{W}\mathbf{V}\boldsymbol{\Delta}+\boldsymbol{\Delta}'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{V}\boldsymbol{\Delta}\\ &\quad \left.-\boldsymbol{\Delta}'\mathbf{W}\frac{\partial\mathbf{V}}{\partial\theta_k}\boldsymbol{\Delta}-\boldsymbol{\Delta}'\mathbf{W}\mathbf{V}\dot{\boldsymbol{\Delta}}_k\right\}.\end{aligned}$$

Taking expectations yields

$$\begin{aligned}\mathrm{E}\left(\frac{\partial^3 L}{\partial\theta_k\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) &= \frac{n}{2}\left\{-[\dot{\boldsymbol{\Delta}}_1',\ldots,\dot{\boldsymbol{\Delta}}_p']\mathbf{W}\boldsymbol{\delta}_k-\boldsymbol{\Delta}'\mathbf{W}\dot{\boldsymbol{\Delta}}_k-\dot{\boldsymbol{\Delta}}_k'\mathbf{W}\boldsymbol{\Delta}\right.\\ &\quad +\boldsymbol{\Delta}'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{\Delta}\\ &\quad \left.+\boldsymbol{\Delta}'\mathbf{W}(\mathbf{I}\otimes\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}+\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\boldsymbol{\Delta}\right\},\\ &= \frac{n}{2}\left\{4\boldsymbol{\Delta}'\mathbf{W}(\mathbf{I}\otimes\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Delta}-\boldsymbol{\Delta}'\mathbf{W}\dot{\boldsymbol{\Delta}}_k-\dot{\boldsymbol{\Delta}}_k'\mathbf{W}\boldsymbol{\Delta}\right.\\ &\quad \left.-[\dot{\boldsymbol{\Delta}}_1',\ldots.\dot{\boldsymbol{\Delta}}_p']\mathbf{W}\boldsymbol{\delta}_k\right\}.\end{aligned}$$

6.7 Some additional expectations

It may be necessary to compute $\mathrm{E}\left(\frac{\partial L}{\partial\theta_k}\frac{\partial^2 L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)$. This will be done elementwise. In doing so the following result is required

$$\mathrm{E}\left(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{Q})\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{R})\right)=\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Q})\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{R})+\frac{2}{n}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Q}\boldsymbol{\Sigma}^{-1}\mathbf{R}),$$

where $\mathbf{R}$ is a $p\times p$ symmetric matrix and $\mathbf{Q}$ is a $p\times p$ matrix. This expectation is computed as follows. Write

$$\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{Q})\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{R})=n^{-2}\mathrm{tr}(\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'\mathbf{Q}^*)\mathbf{tr}(\sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{z}_i\mathbf{z}_i'\mathbf{R}^*)$$

where $\mathbf{z}_1,\ldots,\mathbf{z}_n$ *iid* $\mathbf{MVN(0,I)}$, $\mathbf{Q}^*=\boldsymbol{\Sigma}^{-1/2}\mathbf{Q}\boldsymbol{\Sigma}^{-1/2}\mathbf{and}\mathbf{R}^*=\boldsymbol{\Sigma}^{-1/2}\mathbf{R}\boldsymbol{\Sigma}^{-1/2}$. Hence

$$\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{Q})\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{R}) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{z}_i'\mathbf{Q}^*\mathbf{z}_i)(\mathbf{z}_\mathbf{j}'\mathbf{R}^*\mathbf{z}_\mathbf{j}).$$

$$= n^{-2}\left\{\sum_{i=1}^{n}\sum_{j\neq 1}^{n}(\mathbf{z}_i'\mathbf{Q}^*\mathbf{z}_i)(\mathbf{z}_j'\mathbf{R}^*\mathbf{z}_j) + \sum_{i=1}^{n}(\mathbf{z}_i'\mathbf{Q}^*\mathbf{z}_i)(\mathbf{z}_i'\mathbf{R}^*\mathbf{z}_i)\right\}.$$

Taking the expectation of $\{.\}$ yields

$$n(n-1)\mathrm{tr}(\mathbf{Q}^*)\mathrm{tr}(\mathbf{R}^*) + n\mathrm{E}\left((\mathbf{z}'\mathbf{Q}^*\mathbf{z})(\mathbf{z}'\mathbf{R}^*\mathbf{z})\right),$$

where $\mathbf{z} \sim \mathbf{MVN}(\mathbf{0}, \mathbf{I})$. It is not difficult to show that the last expectation is $\mathrm{tr}(\mathbf{Q}^*)\mathrm{tr}(\mathbf{R}^*) + 2\mathrm{tr}(\mathbf{Q}^*\mathbf{R}^*)$ where $\mathbf{R}^*$ is symmetric. Using the last two expectations yields the required result

$$\begin{aligned}\mathrm{E}\left(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{Q})\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{R})\right) &= \mathrm{tr}(\mathbf{Q}^*)\mathrm{tr}(\mathbf{R}^*) + \frac{2}{n}\mathrm{tr}(\mathbf{Q}^*\mathbf{R}^*). \\ &= \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Q})\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{R}) + \frac{2}{n}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Q}\boldsymbol{\Sigma}^{-1}\mathbf{R}).\end{aligned}$$

Returning to the computation of $\mathrm{E}\left(\frac{\partial L}{\partial\theta_k}\frac{\partial^2 L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)$, write

$$\begin{aligned}\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} &= \frac{n}{2}\left\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i})\right. \\ &\quad - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}) + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}) \\ &\quad \left. - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j})\right\}.\end{aligned}$$

Using $\frac{\partial L}{\partial\theta_k}$ from the section **First order derivatives** gives

$$\begin{aligned}\frac{\partial L}{\partial\theta_k}\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} &= -\frac{n}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k})\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} + \frac{n^2}{4}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k})\left[\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i})\right. \\ &\quad \left. - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j})\right] + \frac{n^2}{4}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k})\left[\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j})\right. \\ &\quad \left. - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i})\right].\end{aligned}$$

Using the expectation results derived previously gives

$$\begin{aligned}\mathrm{E}(\frac{\partial L}{\partial\theta_k}\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}) &= -\frac{n}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j}) \\ &\quad -\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}).\end{aligned}$$

Hence

$$\mathrm{E}\left(\frac{\partial L}{\partial \theta_k}\frac{\partial^2 L}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) = \frac{n}{2}[\dot{\boldsymbol{\Delta}}_1', \ldots, \dot{\boldsymbol{\Delta}}_p']\mathbf{W}\boldsymbol{\delta}_k - n\boldsymbol{\Delta}'\mathbf{W}(\mathbf{I} \otimes \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Delta}.$$