TOPICS IN ANALYZING LONGITUDINAL DATA

A Dissertation

by

HYUNSU JU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2004

Major Subject: Statistics

TOPICS IN ANALYZING LONGITUDINAL DATA

A Dissertation

by

HYUNSU JU

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Suojin Wang
(Chair of Committee)

P. Fred Dahm
(Member)

Johan Lim
(Member)

Thomas Kiffe
(Member)

Michael T. Longnecker
(Head of Department)

December 2004

Major Subject: Statistics

ABSTRACT

Topics in Analyzing Longitudinal Data. (December 2004)

Hyunsu Ju, B.A., Chung-Ang University;

M.S., Seoul National University

Chair of Advisory Committee: Dr. Suojin Wang

We propose methods for analyzing longitudinal data, obtained in clinical trials and other applications with repeated measures of responses taken over time. Common characteristics of longitudinal studies are correlated responses and observations taken at unequal points in time. The first part of this dissertation examines the justification of a block bootstrap procedure for the repeated measurement designs, which takes into account the dependence structure of the data by resampling blocks of adjacent observations rather than individual data points. In the case of dependent stationary data, under regular conditions, the approximately studentized or standardized block bootstrap possesses a higher order of accuracy. With longitudinal data, the second part of this dissertation shows that the diagonal optimal weights for unbalanced designs can be made to improve the efficiency of the estimators in terms of mean squared error criterion. Simulation study is conducted for each of the longitudinal designs. We will also analyze repeated measurement data set concerning nursing home residents with multiple sclerosis, which is obtained from a large database termed the minimum data set (MDS).

To my Parents,

Sung Mee,

Young Eun, and Da Eun

ACKNOWLEDGEMENTS

As I approach the completion of Ph.D at Texas A&M, I reflect on many people who made my experience here a successful one. First, my greatest appreciation is given to my mother, Ilsun Yang. Without her constant support, I would never have been able to begin my journey.

I would like to express my sincere gratitude to my advisor, Dr. Suojin Wang, for all he has done to help me complete my Ph.D. during the past few years. He introduced me to new fields I have never explored. His broad knowledge of statistics, great expertise in nonparametric resampling methodology, and his great personalities have made my research a wonderful journey. It is not only an honor but also a privilege to work with him.

Thanks are extended to my committee members, Dr P. Fred Dahm, Dr. Johan Lim, and Dr Thomas Kiffe for their willingness to serve on my committee and for providing me with valuable suggestions. A special thanks to Dr. Robert J. Buchanan for his help and cooperation in research work. I gratefully acknowledge many other professors in the department for their kind words of encouragement that kept me going. Special thanks go to my friends, Diana Van, Nam Tae Park, Johan Lim, Sang Han Lee, Jumyon Park, Jin Kim for their continuous support, prayers, and friendship.

I would like to thank my wife, Sung Mee Kim, who has always cherished and supported me. Without her constant support and sacrifice, I would have never come this far. She was always confident in my ability to finish this dissertation and kept me strong in the past graduate years. I also thank my lovely daughters Eunice and Esther for their expressions of love to me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

TABLE Page

CHAPTER I

INTRODUCTION

The analysis of repeated measures is used to analyze data from clinical trials and other applications with repeated measures of responses taken over time. A key strength of these studies, in which repeated measurements are obtained from each subject, is that this is an effective design with which it is possible to obtain information concerning individual patterns of change. This type of design also economizes on subjects. Another advantage is that subjects can serve as their own controls, in that the outcome variable can be measured under both control and experimental conditions for each subject. Common characteristics of longitudinal studies are: (1) correlated responses, (2) observations taken at unequal points in time. In a longitudinal study subjects are followed over time. At one extreme a small number of subjects may be studied over long period of time. At the other extreme some longitudinal studies follow up a relatively large group for a short time.

We propose to study two problems: a) moving block bootstrap methods under a number of repeated observations (m) per person that is large for the small number of subjects, b) the diagonal optimal weighting scheme under the working independence setting if the number of repeated observations (m) per person is small for a large number of individuals.

In the case of dependent stationary data, under regular conditions, the approximate studentized or standardized block bootstrap possesses a higher order of accu-

---

The format and style follow that of *Journal of the American Statistical Association*.

racy. This property is referred to as the second-order correctness of the bootstrap approximation, which is more accurate than the normal approximation, as it captures the second order term. While recent developments in Edgeworth expansion theory, especially for sums of weakly dependent time series, have broadened the horizon for bootstrap methods, most existing studies have been devoted to the non-parametric identically independent (i.i.d.) and block bootstraps. The foremost attraction for using bootstrap methods in testing hypotheses and confidence interval constructions is its capability for achieving considerable improvement over standard procedures based on first-order asymptotics. These improvements are justified by the use of the analytical tools of Edgeworth expansions.

We use moving block bootstrap (MBB) methods to obtain efficiency of regression coefficient estimators by blocking the centered residuals in longitudinal data. One of them is within the moving block bootstrap and the other is a mixed moving block bootstrap method. For one subject with a correlated series over time, we want to compare them to the ordinary bootstrap. When the number of subjects (n) is small and the number of repeated measurements (m) is large, by simulation we conclude that the ordinary bootstrap variance estimator can be inconsistent and the resampling subject method may not work well for such small subject samples.

In the unbalanced longitudinal data under working independent assumptions, we consider a subject weighting scheme to reach a certain optimization criterion. One corresponds to equal weights for each observation, the other corresponds to equal weights for each subject. An ideal choice of wights depends on the correlated structure of the data. However, since the actual correlation structure is unknown in practice, we suggest using the diagonal optimal weight in a simple way with the idea of creating a working independent model in generalized estimating equations (GEE). The diagonal optimal weight outperforms the first two weighting schemes and has robustness for misspecified correlation structures in a simple mixed linear model.

The remainder of this dissertation is organized as follows: Chapter II includes an extensive bibliography of work on repeated measures and block bootstrap methods. Chapter III presents the basic statistical modeling framework in longitudinal data. Chapter IV demonstrates the moving block bootstrap justification in longitudinal data theoretically and empirically. Chapter V describes the diagonal optimal weights in the unbalanced longitudinal data. Chapter VI provides the results of analyzing a longitudinal data extracted from the minimum data set with multiple sclerosis patients. A conclusion and some discussions are given in Chapter VII.

CHAPTER II

LITERATURE REVIEW

The main problem in parameter estimation of linear mixed models under longitudinal data comes from the evaluation of the likelihood of the function implying an inversion of the huge unknown variance covariance matrix. Several methods for obtaining an efficiency of fixed regression parameter estimations have been proposed.

Liang and Zeger (1986) were the first in the field to use working correlation matrices for longitudinal data. The GEE approach is an extension of the quasi-likelihood to longitudinal data analysis. The GEE method yields consistent and asymptotically normal solutions, even with misspecification of time dependence. The estimating equations reduce to score equations for multivariate normal outcomes. The GEE approach relies on independence across subjects to consistently estimate the variance of the regression coefficients. The GEE method is feasible in many situations where the maximum likelihood approaches are not necessary because the full multivariate distribution of the response vector is not required.

Xie and Yang (2003) presented asymptotic results when either the number of independent subjects or the cluster sizes (the number of observations for each subject) or both go to infinity. A set of general conditions, information matrix based, is developed, which leads to weak and strong consistency as well as the asymptotic normality of the regression parameter estimators.

Feng, McLerran, and Grizzle (1996) investigated, by simulation, the properties of a bootstrap method that resample subjects rather than resample observations under the linear model for correlated data with Gaussian error. They showed that for balanced and near balanced data when the number of independent subjects is small

(≤ 10), the bootstrap is superior if analysts do not want to impose strong distribution and covariance assumptions. Huang, Wu, and Zhou (2002) suggested a global smoothing procedure for estimating the parameters of a varying coefficient model with repeated measurements. Inference procedures, based on a resampling subject bootstrap, are proposed to construct a confidence region and to perform hypothesis testing.

Efron (1979) introduced the bootstrap procedure for estimating sampling distributions of statistics based on independent and identically distributed (i.i.d) observations. It is well known, in the i.i.d setup, the bootstrap often offers more accurate approximations than classical large sample approximations. (e.g. Singh (1981), Babu (1986)). However, when the observations are not necessarily independent, the classical bootstrap no longer succeeds, as showed by Singh (1981).

Resampling methods for strictly stationary dependent data are based on blocking arguments, in which the data are divided into blocks and these blocks, rather than individual data values or estimated residuals, are resampled. Carlstein (1986) proposed non-overlapping blocks, where Künsch (1989) and Liu and Singh (1992) independently introduced the moving block method, which employs overlapping blocks. Politis and Romano (1992) considered a block of blocks scheme to obtain valid inference of parameters of the infinite dimensional joint distribution of the process, such as a spectrum. In both Carlstein's and Künsch's bootstrap, blocks of fixed length are resampled so that the newly generated pseudo time series is no longer stationary. To fix this shortcoming, Politis and Romano (1992) suggested the stationary bootstrap, which joins together blocks of random length–having a geometric distribution with mean p–and thus generates bootstrap sample paths that are stationary series themselves. Thus, dependency will be reduced for Carlstein's and Künsch's bootstrap. However, in typical applications the underlying dependence is sufficiently weak. Therefore, the main contributions come from short lags which are well approximated by the blocking methods, ensuring that these methods work. The moving blocks method has essentially been shown to be valid for functions of statistics and smooth functions (see Künsch (1989) and Bühlmann (1994)).

In a time series case, Lahiri (1996) applied a multiple linear regression model

$$y_j = x_j^{'}\beta + \epsilon_j, j = 1, \cdots, m,$$

where x_j's are known $p \times 1$ vectors, β is a $p \times 1$ vector of parameters, and $\epsilon_1, \epsilon_2, \cdots$ are stationary, strongly mixing random variables. If $\hat{\beta}_m$ is an M-estimator of β corresponding to some score function ϕ, under some conditions, a two-term Edgeworth expansion for studentized multivariate M-estimator was observed. Also, it was shown that the block bootstrap has a second order correctness for some suitable bootstrap analogs of studentized $\hat{\beta}_m$.

Lahiri (1999) compared the asymptotic behavior of some common block bootstrap methods based on nonrandom as well as random block lengths. Expansions for the bias, the variance, and the mean-squared error of different block bootstrap variance estimators were obtained. It followed from these expansions that using overlapping blocks is to be preferred over nonoverlapping blocks, and that using random block lengths typically leads to mean-squared errors larger than those for nonrandom block lengths. Conditions for the validity of some block resampling procedures under certain factors, like strength of dependence (weak dependence verses long-range dependence) and the existence of the second moment, have been obtained in the literature (Lahiri (1993; 1995)). It was also shown by Lahiri (2001) that the block bootstrap method is consistent if the block length grows at a rate slower than the sample size. When the growth rate of blocks is comparable to the sample size, the resulting approximations are no longer consistent.

One drawback of this method is that it depends critically on a block length which has to be chosen by the user. In Hall et al. (1995) it is shown that the optimal asymptotic rate of the block size for the moving blocks method depends significantly on context, being equal to $m^{1/3}$, $m^{1/4}$ and $m^{1/5}$ in the cases of variance or bias estimation, estimation of a one-sided distribution function, and estimation of a two-sided distribution function, respectively. The latter two quantities are needed for construction of equal-tailed and symmetric confidence intervals, respectively. Therefore, it seems that the strategy of Bühlmann and Künsch (1995) is suboptimal for constructing confidence intervals. Hall et al. (1995) present a practical rule for selecting the block size empirically. It is based on the fact that the asymptotic formula is $b \sim Cm^{1/k}$, where $k = 3, 4$ or 5 is known, and C is a constant that depends on the underlying process. The rule suggested provides a way for estimating the optimal block for a time series of smaller length than the original.

Paparoditis and Politis (2002) presented a new block bootstrap variation, the tapered block bootstrap, which is applicable in the general time series case of approximately linear statistics. The asymptotic validity and the favorable bias properties of the tapered block bootstrap are shown in two important cases: smooth function of means and M-estimators.

Goncalves and White (2002) found that confidence intervals that rely on bootstrap standard errors tend to perform better than confidence intervals that rely on asymptotic closed form variances in multiple linear regression models with autocorrelated and heteroskedastic error. In particular, the coverage error of symmetric MBB percentile-t confidence intervals based on bootstrap standard error estimates are substantially smaller than the coverage error typically found in other (asymptotic theory-based and bootstrap-based) confidence intervals, especially under strong autocorrelation.

CHAPTER III

STATISTICAL MODELS IN LONGITUDINAL DATA

In this chapter we review some common statistical models in longitudinal data.

3.1 Mixed Effects Linear Models

Mixed effects linear models (Hartley and Rao, 1967) have become a popular tool for analyzing repeated measures data which arise in many fields as diverse as agriculture, biology, economics and geophysics. The increasing popularity of these models is explained by the flexibility they offer in modeling the within-subjects correlation often present in repeated measures data, by the handling of both balanced and unbalanced data, and by the availability or reliable and efficient software for fitting them. The most commonly used mixed-effects linear model for a continuous response was proposed by Laird and Ware (1982) and is expressed below. Let y_i be the $m_i \times 1$ vector of repeated measurements on the i^{th} subject. Then consider a mixed effects model described as

$$y_i = x_i\beta + z_i\gamma_i + \varepsilon_i,\ i = 1, \ldots, n, \tag{3.1}$$

where x_i and z_i are the known matrices of order m_i by p and m_i by q respectively, and β is the fixed p by 1 vector of unknown(nonrandom) parameters. The q by 1 vectors γ_i are the random effects with $E(\gamma_i) = 0$, and $Var(\gamma_i) = \sigma^2 B_1$. Finally ε_i are the m_i by 1 vectors of random errors whose elements are no longer required to be uncorrelated. Let's assume that $E(\varepsilon_i) = 0,\ Var(\varepsilon_i) = \sigma^2 R_i,\ Cov(\gamma_i, \gamma_{i'}) = 0,\ Cov(\varepsilon_i, \varepsilon_{i'}) = 0,\ Cov(\gamma_i, \varepsilon_{i'}) = 0$ for all $i \neq i^{'}$, and $Cov(\gamma_i, \varepsilon_i) = 0$. Such assumptions seem to be reasonable in repeated measurement data where subjects are assumed to be independent, yet the repeated measures data may be correlated. Note here that R_i is the appropriate $m_i \times m_i$ submatrix of a $m \times m$ positive definite matrix, where m is the number of time points in the data set where observations have been made. An appropriate covariance structure can be assigned to the data by an appropriate choice of matrices B_1 and R_i. Note that since y_i is a m_i by 1 vector, $i = 1, \ldots, n$, the model can account for the unbalanced repeated measures data, that is, when data from all the subjects have not been observed at all time points.

The n submodel in Equation (3.1) can be stacked one below the other to give a single model

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \beta + \begin{bmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_n \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

or

$$y_{\sum m_i \times 1} = X_{\sum m_i \times p} \beta_{p \times 1} + Z_{\sum m_i \times nq} \gamma_{nq \times 1} + \epsilon_{\sum m_i \times 1}, \tag{3.2}$$

where the definitions of y, x, z, γ, and ε in terms of the matrices and vectors of submodels are self explanatory. In view of the assumptions made on, we have $E(\gamma) = 0,\ E(\varepsilon) = 0$,

$$Var(\gamma) = \sigma^2 \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_1 \end{bmatrix} = \sigma^2 I_n \otimes B_1 = \sigma^2 B$$

and

$$Var(\varepsilon) = \sigma^2 \begin{bmatrix} R_1 & 0 & \cdots & 0 \\ 0 & R_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_n \end{bmatrix} = \sigma^2 R.$$

The symbol $\otimes$ here stands for the *Kronecker product* (Rao, 1973) defined for two matrices $U_{s\times t} = (u_{ij})$ and $W_{l\times m} = (w_{ij})$ as

$$U \otimes W = \begin{bmatrix} u_{11}W & u_{12}W & \cdots & u_{1t}W \\ u_{21}W & u_{22}W & \cdots & u_{2t}W \\ \vdots & \vdots & \ddots & \vdots \\ u_{s1}W & u_{s2}W & \cdots & u_{st}W \end{bmatrix} = (u_{ij}W).$$

It follows from that

$$Var(y) = zVar(\gamma)z' + Var(\varepsilon) = \sigma^2[zBz' + R] = \sigma^2 V.$$

It may be remarked that in many situations, the variance covariance matrix of y may not be in the above form where the parameter σ^2 has been explicitly factored out. However, with appropriate (but not necessarily unique) modifications in the matrices B and R, some parameter σ^2 (not necessarily unique) can be factored out. There are many books dealing at length with linear mixed model. We recommend a few: Graybill (1976), Seber (1977), Arnold (1981), Hocking (1985), Searle (1997), and Searle et al. (1992).

3.1.1 Estimation of effects when V is known

If B_1 and $R_1, \ldots, R_n$ are assumed to be known, then the Best (minimum mean squared error) Linear Unbiased Estimator (BLUE) using the generalized least squares estimator of β is given by (assuming that it uniquely exists)

$$\begin{aligned} \widehat{\beta} &= [X'(ZBZ' + R)^{-1}X]^{-1}X'(ZBZ' + R)^{-1}y \\ &= \left[\sum_{i=1}^{n} x_i'(z_iB_1z_i' + R_i)^{-1}x_i\right]^{-1}\left[\sum_{i=1}^{n} x_i'(z_iB_1z_i' + R_i)^{-1}y_i\right]. \end{aligned} \tag{3.3}$$

The variance covariance of $\widehat{\beta}$ is

$$\sigma^2[X'(ZBZ' + R)^{-1}X]^{-1} = \sigma^2\left[\sum_{i=1}^{n} x_i'(z_iB_1z_i' + R_i)^{-1}x_i\right]^{-1}.$$

Similarly, the Best Linear Unbiased Predictor (BLUP) of γ is given by $BZ^{'}(ZBZ^{'} + R)^{-1}(y - X\beta)$. Further an unbiased estimator of σ^2 is obtained as

$$\widehat{\sigma}^2 = \frac{1}{v}\widehat{\varepsilon}^{'}V^{-1}\widehat{\varepsilon},$$

where $\widehat{\varepsilon} = y - X(X^{'}V^{-1}X)^{-}X^{'}V^{-1}y$ and $v = N - Rank(X)$ is the error degrees of freedom. If $X^{'}(ZBZ^{'} + R)^{-1}X$ does not admit an inverse, for most estimation problems a generalized inverse would replace the inverse in Equation (3.3), provided estimability of the function under consideration has been ensured.

The BLUE of β and BLUP of γ above can also be obtained by solving the system of *mixed model equations*

$$\begin{bmatrix} X^{'}RX & X^{'}R^{-1}Z \\ Z^{'}R^{-1}X & Z^{'}R^{-1}Z + B \end{bmatrix} \begin{bmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{bmatrix} = \begin{bmatrix} X^{'}R^{-1}y \\ Z^{'}R^{-1}y \end{bmatrix}.$$

In addition, if multivariate normality is assumed for γ_i and $\varepsilon_i,\ i = 1, \ldots n$, then,

$$y \sim N_{\sum m_i}(X\beta, \sigma^2[ZBZ^{'} + R]).$$

In this case $\widehat{\beta}$ and $\widehat{\gamma}$ are also the maximum likelihood estimator and maximum likelihood predictor of β and γ, respectively.

Consider the problem of testing a linear hypothesis of the form $H_0 : L\beta = 0$, where L is a full (row) rank matrix. Then the usual test statistic for testing H_0 is

$$F = \frac{\widehat{\beta}^{'}L^{'}(L(X^{'}V^{-1}X)^{-1}L^{'})^{-1}L\widehat{\beta}}{\widehat{\sigma}^2 Rank(L)},$$

which under the null hypothesis H_0 is distributed as F_{v_1,v_2}, where $v_1 = Rank(L),\ v_2$ is the error degree of freedom, and $V = (ZBZ^{'} + R)$.

3.1.2 Estimation of σ^2 and V

When the matrices B and R (or V) are unknown, estimation of these matrices can be carried out using the standard likelihood based methods under the assumption of joint multivariate normality of γ and ε. In practice, certain structures on either one or both of these matrices is assumed so that V is a function of a few unknown parameters, say $\theta_1, \ldots \theta_s$. The above method is iterative in that first for a fixed value of V, an estimator of β using the form the BLUE is obtained. Then the likelihood function of V is maximized with respect to $\theta_1, \ldots \theta_s$ in order to get an estimate of V. These two steps are employed until a certain user specified convergence criterion is met.

The ML estimator of $\theta_1, \ldots \theta_s$ and hence V (B and R) and of σ^2 are obtained by maximizing the logarithm of the normal likelihood function

$$l(\theta) = -\frac{1}{2}\log|\sigma^2 V| - \frac{1}{2\sigma^2}(y - X\widehat{\beta})'V^{-1}(y - X\widehat{\beta}) - \frac{N}{2}\log(2\pi) \tag{3.4}$$

simultaneously with respect to these parameters. The ML estimator of σ^2 expressed in terms of $\widehat{V}$ will be $\widehat{\sigma}^2_n = \widehat{\varepsilon}'\widehat{V}^{-1}\widehat{\varepsilon}/n$. The ML estimator of $\theta_1, \ldots \theta_s$, generally have to be obtained using the iterative schemes.

Alternatively, estimators of $\theta_1, \ldots \theta_s$, and finally of σ^2 can be obtained by maximizing the function:

$$-\frac{1}{2}\log|V| - \frac{N}{2}\log(y - X\widehat{\beta})'V^{-1}(y - X\widehat{\beta}) - \frac{N}{2}\left[1 + \log\left(\frac{2\pi}{N}\right)\right],$$

which is obtained from the log-likelihood function after factoring and profiling a residual variance $\widehat{\sigma}^2_N$.

Similarly, another set of estimators commonly known as the Restricted Maximum Likelihood (REML) estimators is obtained by maximizing the function (after profiling $\hat{\beta}$)

$$\begin{aligned}&-\frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X| - \frac{N-k}{2}\log(y - X\widehat{\beta})'V^{-1}(y - X\widehat{\beta})\\ &-\frac{N-k}{2}\left[1 + \log\left(\frac{2\pi}{N-k}\right)\right],\end{aligned}$$

where $k = Rank(X)$. The ML and REML estimators are known to be asymptotically equivalent.

Suppose $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_s)'$ is the ML estimate of $\theta = (\theta_1, \ldots \theta_s)'$. Let $h(\theta)$ be a certain, possibly vector valued, function of θ. Then the three asymptotic tests to test $H_0 : h(\theta) = 0$ against the alternative $H_1 : h(\theta) \neq 0$ are given by

$$\begin{aligned} \text{Wald's Statistic:} \quad & T_W = Nh(\widehat{\theta}) \left[H(\widehat{\theta})' I(\widehat{\theta})^{-1} H(\widehat{\theta})\right]^{-1} h(\widehat{\theta}) \\ \text{Likelihood Ratio Test (LRT) Statistic:} \quad & T_L = 2\left[l(\widehat{\theta}) - l(\widehat{\theta}_0)\right] \\ \text{Rao's Statistic:} \quad & T_R = \frac{1}{N} U(\widehat{\theta}_0)' I(\widehat{\theta}_0)^{-1} U(\widehat{\theta}_0), \end{aligned}$$

where $\widehat{\theta}_0$ is the ML estimator of θ under the null hypothesis H_0, $U(\theta) = \frac{\partial l}{\partial \theta}$, $H(\theta) = \frac{\partial h(\theta)}{\partial \theta}$, and $I(\theta)$ is the *Fisher information matrix.*

Under certain regularity conditions each of statistic T_W, T_L, and T_R has an asymptotic χ^2_r distribution under H_0, where $r = Rank(H(\theta))$. Since REML and ML estimates are asymptotically equivalent one may alternatively use the REML estimates in the above expressions.

Since under certain regularity conditions, the ML estimator $\widehat{\theta}$ follows a multivariate normal distribution with the mean vector θ and the variance covariance matrix $I^{-1}(\theta)$, one can perform a test for the hypothesis about any component θ_i of θ using the standard normal distribution. This asymptotic test is also known as Wald's test. Using this asymptotic result, approximate confidence intervals can be constructed as well.

3.1.3 Estimation of Effect When V is Estimated

Suppose $\widehat{B}$ and $\widehat{R}$ are the estimators of B and R respectively, obtained by using one of the above two methods. Then the respective estimator of β and γ are obtained by solving the plug-in version of the *mixed model equation*,

$$\begin{bmatrix} X'\widehat{R}X & X'\widehat{R}^{-1}Z \\ Z'\widehat{R}^{-1}X & Z'\widehat{R}^{-1}Z+\widehat{B} \end{bmatrix} \begin{bmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{bmatrix} = \begin{bmatrix} X'\widehat{R}^{-1}y \\ Z'\widehat{R}^{-1}y \end{bmatrix},$$

where the estimator $\widehat{B}$ and $\widehat{R}$ respectively have been used for B and R in the mixed model equation. Upon solving, we obtain $\widehat{\beta} = (X'\widehat{V}^{-1}X)^{-1}X'\widehat{V}^{-1}y$ and $\widehat{\gamma} = \widehat{B}Z'\widehat{V}^{-1}(y - X\widehat{\beta})$, where $\widehat{V}$ is obtained by substituting $\widehat{B}$ and $\widehat{R}$ for B and R respectively in V. Note that $\widehat{\beta}$ is an estimator of the BLUE $(X'V^{-1}X)^{-}X'V^{-1}y$ of β and $\widehat{\gamma}$ is an estimator of the BLUP $BZ'V^{-1}(y-X(X'V^{-1}X)^{-}X'V^{-1}y)$ of the random effect vector γ.

For simplicity, let us define the estimator of σ^2 by $\widehat{\sigma}^2$, whatever the method may have been used for the estimation. The estimated variance and covariance matrices of these estimators are: $\widehat{Var}(\widehat{\beta}) = \widehat{\sigma}^2 C_{11} = \widehat{\sigma}^2(X'\widehat{V}^{-1}X)^{-}$, $\widehat{Cov}(\widehat{\beta},\widehat{\gamma}) = \widehat{\sigma}^2 C_{21} = -\widehat{\sigma}^2\widehat{B}Z'\widehat{V}^{-1}XC_{11}$, and $\widehat{Var}(\widehat{\gamma}) = \widehat{\sigma}^2 C_{22} = \widehat{\sigma}^2((Z'\widehat{R}^{-1}Z+\widehat{B}^{-1})^{-1} - C_{21}X'\widehat{V}^{-1}ZB)$. It may however be cautioned that

$$\widehat{\sigma}^2 \begin{bmatrix} C_{11} & C_{21}' \\ C_{21} & C_{22} \end{bmatrix}$$

usually underestimates $Var(\widehat{\beta}',\widehat{\gamma}')'$, the true variance covariance matrix of $(\widehat{\beta}',\widehat{\gamma}')'$.

Let us consider a simple case with a random intercept for an unbalanced data. The model is given by

$$\begin{aligned} E(y_i|\gamma_i) &= \mu 1_{m_i} + \gamma_i 1_{m_i} \\ &= x_i\beta + \gamma_i 1_{m_i}, \end{aligned} \tag{3.5}$$

where $x_i = [1_{m_i}]$ and $\beta = \mu$. The likelihood and log-likelihood are

$$\begin{aligned} L &= \Pi_{i=1}^{n}(2\pi)^{\frac{-m_i}{2}}|V_i|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(y_i-\mu 1_{m_i})' V_i^{-1}(y_i-\mu 1_{m_i})\right], \\ \text{and} \quad \log L &= -\frac{1}{2}N\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}\log(\sigma^2+m_i\sigma_\gamma^2) - \frac{1}{2}(N-n)\log\sigma^2 \\ &\quad -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(y_{ij}-\mu)^2 + \frac{\sigma_\gamma^2}{2\sigma^2}\sum_{i=1}^{n}\frac{(y_{i\cdot}-m_i\mu)^2}{\sigma^2+m_i\sigma_\gamma^2}, \end{aligned} \tag{3.6}$$

with $y_i = [y_{i1}, y_{i2}, \cdots, y_{m_i}]'$ and $y_i \sim N(\mu 1_{m_i}, V_i)$, where $V_i = \sigma_\gamma^2 J_{m_i} + \sigma^2 I_{m_i}$. We define $\lambda_i = \sigma^2 + m_i\sigma_\gamma^2$ and $SSE = \sum_{i=1}^{n}\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_{\cdot\cdot})^2$. The log-likelihood can be derived as follows:

$$\begin{aligned} l = \quad & -\frac{1}{2}N\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}\log\lambda_i - \frac{1}{2}(N-n)\log\sigma^2 \\ & -\frac{SSE}{2\sigma^2} - \sum_{i=1}^{n}\frac{m_i(\bar{y}_{i\cdot}-\mu)^2}{2\lambda_i}. \end{aligned} \tag{3.7}$$

The likelihood estimation equations are

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n}\frac{m_i(\bar{y}_{i\cdot}-\mu)}{\lambda_i}, \tag{3.8}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{(N-n)}{2\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\lambda_i} + \frac{SSE}{2\sigma^4} + \sum_{i=1}^{n}\frac{m_i(\bar{y}_{i\cdot}-\mu)^2}{2\lambda_i^2}, \tag{3.9}$$

$$\text{and} \quad \frac{\partial l}{\partial \sigma_\gamma^2} = -\frac{1}{2}\sum_{i=1}^{n}\frac{m_i}{\lambda_i} + \sum_{i=1}^{n}\frac{m_i^2(\bar{y}_{i\cdot}-\mu)^2}{2\lambda_i^2}. \tag{3.10}$$

We have a solution for μ, which is

$$\begin{aligned} \mu = \frac{\sum_{i=1}^{n}\frac{m_i\bar{y}_{i\cdot}}{\lambda_i}}{\sum_{i=1}^{n}\frac{m_i}{\lambda_i}} &= \frac{\sum_{i=1}^{n}\frac{m_i\bar{y}_{i\cdot}}{\sigma^2+m_i\sigma_\gamma^2}}{\sum_{i=1}^{n}\frac{m_i}{\sigma^2+m_i\sigma_\gamma^2}} \\ &= \frac{\sum_{i=1}^{n}\frac{\bar{y}_{i\cdot}}{var(\bar{y}_{i\cdot})}}{\sum_{i=1}^{n}\frac{1}{var(\bar{y}_{i\cdot})}} \end{aligned} \tag{3.11}$$

with $var(\bar{y}_{i\cdot}) = \sigma_\gamma^2 + \frac{\sigma^2}{m_i} = \frac{1}{w_i}$. For σ_γ^2 and σ^2,

$$\begin{aligned} &\frac{SSE}{\sigma^4} - \frac{(N-n)}{\sigma^2} + \sum_{i=1}^{n}\frac{m_i(\bar{y}_{i\cdot}-\mu)^2}{\lambda_i^2} - \sum_{i=1}^{n}\frac{1}{\lambda_i} = 0 \\ &\sum_{i=1}^{n}\frac{m_i^2(\bar{y}_{i\cdot}-\mu)^2}{\lambda_i^2} = \sum_{i=1}^{n}\frac{m_i}{\lambda_i} \end{aligned} \tag{3.12}$$

with $\lambda = \sigma^2 + m_i\sigma_\gamma^2$.

There is no analytic solution for the estimators in general, but there is when the data are balanced (i.e. $m_i = m$ and $\lambda_i = \lambda$ for all i). In balanced case, we have $\mu = \bar{y}_{..}$, $\sigma^2 = MSE$, $\lambda = \frac{SSA}{n} = (1-\frac{1}{n})MSA$, $\sigma_\gamma^2 = \frac{\lambda-\sigma^2}{m} = \frac{(1-\frac{1}{n})MSA-MSE}{m}$, where $MSA = \frac{SSA}{n-1} = \frac{1}{n-1}\sum_{i=1}^n m(\bar{y}_{i.} - \bar{y}_{..})^2$ and $MSE = \frac{SSE}{n(m-1)} = \frac{1}{n(m-1)}\sum_{i=1}^n\sum_{j=1}^n(y_{ij} - \bar{y}_{..})^2$. If we use $R_i = \sigma^2 I_{m_i}$, we obtain $V_i = \sigma_\gamma^2 J_{m_i} + \sigma^2 I_{m_i}$ and $V_i^{-1} = \frac{1}{\sigma^2}(I_{m_i} - \frac{\sigma_\gamma^2}{\sigma^2+m_i\sigma_\gamma^2}J_{m_i})$. Define m_{ij} as 1 if y_{ij} exists and 0 if y_{ij} does not exist; i.e, n_{ij} is the number of data on subject i at time j, either 0 or 1.

1) Estimating the fixed effect.

$$\begin{aligned} \hat{\beta} &= [\frac{1}{\sigma^2}\sum_{i=1}^n(X_i'X_i - \frac{\sigma_\gamma^2}{\sigma^2+m_i\sigma_\gamma^2}X_i'J_{m_i}X_i)]^{-1} \\ &\times \sum_{i=1}^n\frac{1}{\sigma^2}\left\{m_{ij}[y_{ij} - \frac{\sigma_\gamma^2}{\sigma^2+m_i\sigma_\gamma^2}y_{i.}]\right\}_{j=1}^{T=max(m_i)} \end{aligned} \tag{3.13}$$

2) Predicting the random effect

$$\begin{aligned} BZ'V^{-1}(y - x\hat{\beta}) &= \sigma_\gamma^2\left\{\frac{m_i\bar{y}_{i.}}{\sigma^2+m_i\sigma_\gamma^2} - \frac{\sum_{j=1}^{m_i}m_{ij}\hat{\beta}_j}{\sigma^2+m_i\sigma_\gamma^2}\right\}, \\ \gamma_i^{BLUP} &= \frac{m_i\sigma_\gamma^2}{\sigma^2+m_i\sigma_\gamma^2}(\bar{y}_{i.} - \frac{\sum_{j=1}^{m_i}m_{ij}\hat{\beta}_j}{m_i}). \end{aligned} \tag{3.14}$$

3.1.4 Tests for Fixed Effect Parameters

Consider the problem of testing a linear hypothesis of the form $H_0 : L\beta = 0$, where L is a full-rank matrix. A suggested test statistic for H_0 is

$$F = \frac{\widehat{\beta}'L'(LC_{11}L')^{-1}L\widehat{\beta}}{\widehat{\sigma}^2 Rank(L)}.$$

The exact distribution of F is complicated by many facts. For example, $\widehat{\beta}$ is only an approximate version of the BLUE since B_1 and $R_1, \ldots R_n$ are unknown and hence their estimates have been used in their expressions. The matrix $\widehat{\sigma}^2 C_{11}$ is also an estimated version of the variance covariance matrix of $\widehat{\beta}$. Further, the distribution of F also depends on the type of unbalancedness that exists in the data. However, for large samples, the test statistic F will have an approximate F distribution with numerator degrees of freedom $v_1 = rank(L)$ and denominator degrees of freedom v_2 approximately estimated (Searle et al., 1992).

3.1.5 Selection of Appropriate Structure for B and R

Given numerous choices of structures for B and R, one of the problems a practitioner faces is the selection of appropriate structure. Under the model fitting information, Akike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (BIC) are often used.

Akike's Information Criterion (AIC) is defined as

$$\text{AIC} = -2l(\widehat{\theta}) + 2q, \tag{3.15}$$

where $l(\theta)$ is the log-likelihood function (or unrestricted log-likelihood function) and $l(\widehat{\theta})$ is the maximum log-likelihood function (or unrestricted maximum log-likelihood function) and q is the number of the estimated covariance parameters. The structure expressed in terms of θ with the smallest AIC is preferred.

Schwarz's Bayesian Criterion (BIC) is defined as

$$\text{BIC} = -2l(\widehat{\theta}) + q\log(N^*), \tag{3.16}$$

where $N^* = N$ for ML and $(N - k)$ for REML. Similar to AIC interpretation, a model with a small value of BIC is preferred.

Keselman et al. (1998) indicate through extensive simulation studies that the AIC performs better than BIC in trying to identify the true models. The poor performance of BIC might be due to the fact that the penalty criterion is a function of N, the total number of observations rather than the number of subjects.

In the context of selecting a covariance structure for R, LRT on a covariance structure can be performed to decide if the particular covariance structure is deemed adequate. One can use the log-likelihood ratio test statistic or the chi-square statistic associated with that. The degree of freedom of the chi-square distributions are determined by taking the difference between the number of parameters in the full model and that in the reduced (under the null hypothesis) model.

3.2 Generalized Estimating Equations (GEE)

The generalized estimating equations (GEE) are the marginalization modeling methods for analyzing repeated measurement data. The GEE approach is an extension of quasi-likelihood to longitudinal data analysis. The method is semiparametric in that the estimating equation is derived without fully specifying the joint distribution of a subject's observations. It is only required that the likelihood for the marginal distribution and a working covariance matrix for the vector of repeated measurements be obtained for each subject. The GEE estimators are consistent and asymptotic normally distributed even with misspecifying covariance structure. The estimating equations reduce to the score equations for multivariate normal outcomes. The method avoids the need for multivariate distributions by assuming a functional form of the marginal distribution at each point. The covariance structure is considered as a nuisance. The GEE approach relies on independence across subjects to estimate consistently the variance of the regression coefficient, even when the assumed correlation is incorrect. There are many books dealing with GEE in detail. We recommend some books: Hand and Croweder (1996), Diggle et al. (1994), Lindsey (1999), Dobson (2002), Davis (2002).

3.2.1 Assumptions of the Method

The GEE model requires the first and second moment conditions. The marginal response

$$\mu_{ij} = E(y_{ij}) \tag{3.17}$$

has linked to a linear combination of the covariates,

$$g(\mu_{ij}) = x_{ij}\beta, \tag{3.18}$$

where y_{ij} is the response of subject i at time j, $x_{ij} = (x_{ij1}, \ldots x_{ijp})$ is the corresponding $1 \times p$ vector of covariates, and $\beta = (\beta_1, \ldots \beta_p)'$ is a $p \times 1$ vector of unknown parameters. $g(\cdot)$ is the link function.

The second moment condition is that the variance of y_{ij} as a function of the mean,

$$Var(y_{ij}) = \phi V(\mu_{ij}), \tag{3.19}$$

where $V(\cdot)$ is the variance function and ϕ is a possible unknown scale parameter. For normally distributed responses, a natural choice is

$$g(\mu_{ij}) = \mu_{ij}, \qquad V(\mu_{ij}) = 1, \qquad Var(y_{ij}) = \phi.$$

If the response variable is binary, the choice is

$$g(\mu_{ij}) = \log(\frac{\mu_{ij}}{1-\mu_{ij}}), \qquad V(\mu_{ij}) = \mu_{ij}(1-\mu_{ij}), \qquad \phi = 1.$$

If the response variable is Poisson count,

$$g(\mu_{ij}) = \log(\mu_{ij}), \qquad V(\mu_{ij}) = \mu_{ij}, \qquad \phi = 1,$$

are often used.

3.2.2 Working Correlation Matrix

It is required for the GEE models to choose the form of a $m_i \times m_i$ working correlation matrix $R_i(\alpha)$ for each $y_i = (y_{i1}, \dots y_{im_i})'$. The (j, j') element of $R_i(\alpha)$ is the known, hypothesized, or estimated correlation between y_{ij} and $y_{ij'}$. This working correlation matrix may depend on a vector of unknown parameter α, which is same for all subjects. Although this correlation matrix can differ from subject to subject, we can commonly use a working correlation matrix $R = R(\alpha)$ that approximates the average dependence among repeated observations of subjects. We should choose the form of R to be consistent with the empirical correlations. The R is called a working correlation matrix because with nonnormal response, the actual correlation among subjects' outcomes may depend on the mean values, and hence on $x_{ij}\beta$. The commonly used specific choices of the form of the working correlation matrix are

- Independence:

  $R = I$ - the GEE reduce to the independence estimating equation.

- Exchangeable:

  $R_{jj'} = \alpha \quad$ for $j \neq j'$ - same structure as in random-intercepts model.

- AR(1):

  $R_{jj'} = \alpha^{|j-j'|}$.

- m-dependent:

  $$R_{jj'} = \begin{cases} \alpha^{|t_j - t_{j'}|} & \text{if } |t_j - t_{j'}| \quad \leq m \\ 0 & \text{if } |t_j - t_{j'}| \quad > m \end{cases}.$$

- Unstructured:

$R_{jj'} = \alpha_{jj'}$ - $m(m+1)/2$ parameters to be estimated. It is most efficient in some cases, but useful only when there are relatively few observations. The missing data make a more complicated estimation of R, and the estimate obtained using nonmissing data is not guaranteed to be positive definite

$$y'Ry > 0 \quad \text{for all} \quad y \neq 0, \tag{3.20}$$

which is a problematic inversion of R.

The GEE method yields consistent estimates of the regression coefficient and their variance, even with misspecfication of the structure of the covariance matrix.

For ith subject, let A_i be the $m_i \times m_i$ diagonal matrix with marginal variance of y_{ij}, i.e. $A_i = diag\{v(\mu_{i1}), \ldots v(\mu_{im_i})\}$. Also, let $R_i(\alpha)$ be the $m_i \times m_i$ invertible *working* correlation matrix for the ith subject. The working covariance matrix for $y_i = (y_{i1}, \ldots y_{im_i})'$ is

$$V_i(\alpha) = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2}. \tag{3.21}$$

The working correlation matrix is not usually known and must be estimated. It is estimated in the iterative fitting process using the current value of the parameter vector β to compute appropriate functions of the standardized Pearson residuals

$$r_{ij} = \frac{y_{ij} - \widehat{\mu}_{ij}}{\sqrt{[V_i]_{jj}}}. \tag{3.22}$$

It is noted that, in the normal case, the denominator is $\sqrt{[V_i]_{jj}} = 1$, namely $r_{ij} = y_{ij} - \widehat{\mu}_{ij}$. The $Var(r_{it}) = \phi$ and

$$\widehat{\phi} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{m_i} r_{it}^2}{\sum_{i=1}^{n} m_i - p}, \tag{3.23}$$

where p is the number of regression parameters. As an example of estimating α, suppose we assume an exchangeable correlation structure. Then

$$corr(y_{it}, y_{it'}) \approx corr(r_{it}, r_{it'})\phi^{-1} \approx E(r_{it}, r_{it'}), \tag{3.24}$$

so that the method of moment estimator is

$$\widehat{\alpha} = \widehat{\phi}^{-1} \frac{\sum_{i=1}^{n} \sum_{t \neq t'}^{m_i} r_{it} r_{it'}}{\sum_{i=1}^{n} m_i(m_i - 1) - p}. \tag{3.25}$$

3.2.3 Solving the GEE

The GEE estimate of β is the solution of

$$U(\beta) = \sum_{i=1}^{n} \left(\frac{\partial \mu(x_i \beta)}{\partial \beta} \right)' [V_i(\widehat{\alpha})]^{-1} (y_i - \mu(x_i \beta)) = 0_p, \tag{3.26}$$

where $\widehat{\alpha}$ is a consistent estimate of α and 0_p is the $p \times 1$ vector $(0, \ldots 0)'$. The iterative procedure begins with starting value β_0 and calculate updated value β_{s+1} from β_s by

$$\beta_{s+1} = \beta_s - \left[\sum_{i=1}^{n} \frac{\partial \mu}{\partial \beta}' V_i^{-1} \frac{\partial \mu}{\partial \beta} \right]^{-1} \left[\sum_{i=1}^{n} \frac{\partial \mu}{\partial \beta}' V_i^{-1} (y_i - \mu_i) \right]. \tag{3.27}$$

The estimator $\widehat{\beta}$ can be obtained by Fisher scoring, which can be viewed as iteratively weighted least square estimates.

For the normal case,

$$\begin{aligned} \mu_i &= x_i \beta, \\ \frac{\partial \mu_i}{\partial \beta} &= x_i, \\ V_i(\hat{\alpha}) &= R_i(\hat{\alpha}). \end{aligned} \tag{3.28}$$

Thus,

$$\sum_{i=1}^{n} x_i' \big[R_i(\hat{\alpha}) \big]^{-1} (y_i - x_i \beta) = 0, \tag{3.29}$$

and solving for β yields

$$\hat{\beta} = \Big\{ \sum_{i=1}^{n} x_i' [R_i(\hat{\alpha})]^{-1} x_i \Big\}^{-1} \Big\{ \sum_{i=1}^{n} x_i' [R_i(\hat{\alpha})]^{-1} y_i \Big\}, \tag{3.30}$$

which is solved as an iteratively weighted least square estimate.

The GEE method has some desirable properties that make it an attractive method for dealing with correlated data. It reduced to the independence estimating equation when $m_i = 1$, and is the maximum score equation for multivariate Gaussian data. Also, it is shown by that

$$\sqrt{n}(\widehat{\beta} - \beta) \rightarrow N(0, I),$$

if the mean model is correct even if a working covariance matrix in Equation (3.21) V_i is incorrectly specified, where

$$\begin{aligned} I &= M_0^{-1} M_1 M_0^{-1}, & (3.31) \\ M_0 &= \sum_{i=1}^{n} \frac{\partial \mu}{\partial \beta}' V_i^{-1} \frac{\partial \mu}{\partial \beta}, \quad \text{and} & (3.32) \\ M_1 &= \sum_{i=1}^{n} \frac{\partial \mu}{\partial \beta}' V_i^{-1} Cov(y_i) V_i^{-1} \frac{\partial \mu}{\partial \beta}. & (3.33) \end{aligned}$$

The property listed means that we don't have to specify the working correlation matrix correctly in order to have a consistent estimator of the regression parameters. Choosing the working correlation closer to the true correlation increases the statistical efficiency of the regression parameter estimator, so we should specify the working correlation as accurately as possible based on knowledge of the measurement process.

3.2.4 Robust Variance Estimate

The model based estimator of $Var(\widehat{\beta})$ is given by

$$\widehat{Var}(\widehat{\beta}) = \hat{M}_0^{-1}, \tag{3.34}$$

where $\hat{M}_0 = \sum_{i=1}^{n} \left(\frac{\partial \widehat{\mu}_i}{\partial \beta}\right)' \hat{V}_i^{-1} \left(\frac{\partial \widehat{\mu}_i}{\partial \beta}\right)$ and $\hat{V}_i = V_i(\widehat{\alpha})$. This is the GEE equivalent for the inverse of the Fisher information matrix that is often used in generalized linear models as an estimator of the covariance estimate of the maximum likelihood estimator of β. It is a consistent estimator of the covariance matrix of $\widehat{\beta}$, if the mean model and working correlation matrix are correctly specified.

Liang and Zeger (1986) recommended that the variance-covariance of $\widehat{\beta}$ be estimated by

$$\widehat{Var}(\widehat{\beta}) = \hat{M}_0^{-1}\hat{M}_1\hat{M}_0^{-1}, \tag{3.35}$$

where

$$\hat{M}_1 = \sum_{i=1}^{n} \frac{\partial \mu'}{\partial \beta} \hat{V}_i^{-1}(y_i - \widehat{\mu_i})(y_i - \widehat{\mu_i})' \hat{V}_i^{-1} \frac{\partial \mu}{\partial \beta}. \tag{3.36}$$

This estimator was defined by Royall (1986) and is known as *robust* or *information sandwich* estimator. It has the property of being a consistent estimator of the covariance matrix of $\widehat{\beta}$, even if the working correlation matrix is misspecified. Note that if the true correlation structure is correctly modeled, then $Var(y_i) = V_i$ and it simplifies from Equations (3.31) and (3.33) to

$$Var(\widehat{\beta}) = M_0^{-1}M_1M_0^{-1} = M_0^{-1}M_0M_0^{-1} = M_0^{-1}, \tag{3.37}$$

which can be estimated by a model-based estimator, $\widehat{Var}(\hat{\beta}) = \hat{M}_0^{-1}$.

3.2.5 Hypothesis Testing

After estimating the vector of regression coefficient $\widehat{\beta}$, it may be of interest to test the hypothesis concerning the elements of β. Consider a hypothesis of the form

$$H_0 : C\beta = d,$$

where C is a $c \times p$ matrix of constants with imposing c linearly independent constraints on the element of β and d is a $p \times 1$ vector of constants. Because $\widehat{\beta}$ is asymptotically normal, the Wald statistic

$$Q_C = (C\widehat{\beta} - d)' \left[C\widehat{Var}(\widehat{\beta})C' \right]^{-1} (C\widehat{\beta} - d) \tag{3.38}$$

has an asymptotic χ_c^2 distribution if H_0 is true.

CHAPTER IV

MOVING BLOCK BOOTSTRAP

4.1 Introduction

Many longitudinal designs are the case that the number of subjects n is large and the number of replications m_i is bounded. Liang and Zeger (1986) proved that the GEE estimator is consistent and asymptotically normal with misspecification of covariance parameters. That asymptotic property is considered when the number of subjects n goes to infinity and m_i is bounded. Xie and Yang (2003) proved the almost sure existence and strong consistency of GEE estimators, which are focused on three large sample settings:

- $n \to \infty$ and $m = m(n) = \max_{1 \le i \le n}(m_i)$ is bounded above, for all n;
- n is bounded but $m \to \infty$;
- $m \to \infty$ as $n \to \infty$.

We will consider the case of the longitudinal design in which the number of subjects n is bounded and the number of replications m, the same number for all subjects, is large. The model on which we focus is given by

$$y_i = x_i\beta + e_i, \tag{4.1}$$

where $i = 1, \cdots, n$, x_i is $m \times p$ design matrix, β is $p \times 1$ vector of unknown parameters, $y_i = (y_{i1}, \cdots, y_{im})'$, and $e_i = (y_{i1}, \cdots, e_{im})'$. Note that Equation (3.1) is rewritten as Equation (4.1) with $e_i = z_i\gamma_i + \varepsilon_i$ using the marginal model representation. The repeated observations are correlated for each subject. Since bootstrap methods that resample small subjects or resample observations independently may not work well, we will investigate the moving block bootstrap method developed for time series correlated data.

4.2 Inadequacy of the Standard Bootstrap for Dependent Data

We refer to the nonparametric resampling scheme of Efron (1979), introduced in the context of iid (identical independent distributed), as the standard bootstrap. There are also some alternative terms such as "naive" and "ordinary" bootstrap in the literature for Efron's (1979). For notational simplicity in this section, we consider the one subject model which is given by

$$y_j = \theta + e_j, \tag{4.2}$$

where $y_j = (y_1, \cdots, y_m)'$.

Definition 4.2.1 *A sequence of random variables $\{y_j\}_{j\in\mathbf{Z}}$ is called stationary if for every $j_1 < j_2 < \cdots < j_p$, $p \in \mathbb{N}$ and for every $m \in \mathbb{Z}$, the distributions of $(y_{j_1}, \ldots y_{j_p})'$ and $(y_{j_1+m}, \ldots y_{j_p+m})'$ are the same.*

Definition 4.2.2 *A sequence of random variables $\{y_j\}$ is said to be k-dependent if for $s - r > k$, the two subsequence $\{, \ldots, y_{r-2}, y_{r-1}, y_r\}$ and $\{y_s, y_{s+1}, y_{s+2}, \ldots\}$ are independent.*

Suppose that $y_1, y_2, \ldots$ is a stationary sequence. Let $\theta = Ey_j$, $\sigma^2 = Var(y_j)$, and $\gamma_j = Cov(y_1, y_{1+j})$. The variance of sample mean $\overline{y}_m$ for stationary distribution is

$$Var\{\sqrt{m}(\overline{y}_m - \theta)\} = \sigma^2 + \frac{2}{m}\sum_{j=1}^{m-1}(m-j)\gamma_j. \tag{4.3}$$

Suppose that

$$\frac{2}{m}\sum_{j=1}^{m-1}(m-j)\gamma_j \to \gamma \tag{4.4}$$

as $m \to \infty$. It seems reasonable to get

$$\sqrt{m}(\overline{y}_m - \theta) \to^d N(0, \sigma^2 + \gamma). \tag{4.5}$$

Let us consider k-dependent sequence. For an k-dependent sequence, we have $\gamma_p = 0$ for all $p > k$, and therefore

$$\frac{2}{m}\sum_{j=1}^{k}(m-j)\gamma_j \to 2\sum_{j=1}^{k}\gamma_j. \tag{4.6}$$

This leads to the cental limit theorem for k-dependence.

- (CLT for k-dependent sequences): If for some $k \geq 0$, $y_1, y_2, \ldots$ is a stationary k-dependent sequence with $Ey_j = \theta$ and $Var(y_j) = \sigma^2 < \infty$, then

  $$\sqrt{m}(\overline{y}_m - \theta) \to^d N(0, \tau^2), \tag{4.7}$$

  where

  $$\tau^2 = \sigma^2 + 2\sum_{j=1}^{k} \quad Cov(y_1, y_{1+j}). \tag{4.8}$$

We sketch the idea of the proof as follows.

Let $M = M(m)$ be integer that goes to ∞ as $m \to \infty$, but at a slower rate than m so that $m/M \to \infty$. Then $\sum_{j=1}^{m} y_j$ may be broken into two parts as follows:

$$\begin{aligned}
\sum_{j=1}^{m} y_j &= (y_1 + \cdots + y_M) + (y_{M+1} + \cdots y_{M+k}) \\
&\quad + (y_{M+k+1} + \cdots y_{2M+k}) + (y_{2M+k+1} + \cdots y_{2M+2k}) + \cdots \\
&= A_1 + B_1 + A_2 + B_2 + \cdots,
\end{aligned}$$

where the A_j each consists of M term and the B_j each consist of k terms. Thus we obtain for $m = r(M+k)$

$$\sqrt{m}(\overline{y}_m - \theta) = \frac{1}{\sqrt{m}}\sum_{j=1}^{r}(A_j - \theta) + \frac{1}{\sqrt{m}}\sum_{j=1}^{r}(B_j - \theta). \tag{4.9}$$

The key is that the A_j are iid, so we can use the CLT. On the other hand, the B_j have an asymptotically negligible contribution to the sum.

As one example, the first order moving average process can be modeled as follows:

$$y_j = \theta + \varepsilon_j + \phi\varepsilon_{j-1}, \tag{4.10}$$

where ε_j are iid $N(0, \sigma^2)$. The mean, variance and covariance of y_j are

$$\begin{aligned} E(y_j) &= \theta, \quad Var(y_j) = (1+\phi^2)\sigma^2 \quad \text{and} \\ \gamma_s &= E[y_j y_{j+s}] - E[y_j]E[y_{j+s}] = \begin{cases} (1+\phi^2)\sigma^2, & s = 0 \\ \phi\sigma^2, & s = 1 \\ 0, & s > 1 \end{cases} \end{aligned} \tag{4.11}$$

So the $\sqrt{m}(\overline{y}_m - \theta) \to N(0, (1+2\phi+\phi^2)\sigma^2)$ from the k-dependent CLT results. The ordinary central limit theorem gives us $\sqrt{m}(\overline{y}_m - \theta) \to N(0, (1+\phi^2)\sigma^2)$, since the stationarity implies $Var(y_j) = (1+\phi^2)\sigma^2$ for all j.

Definition 4.2.3 *Let $\{y_j\}_{j\in\mathbb{Z}}$ be a sequence of random vectors. Then the strong mixing or α-mixing coefficient of $\{y_j\}_{j\in\mathbb{Z}}$ is defined as*

$$\begin{aligned} \alpha(m) = \sup\{&|P(A \cap B) - P(A)P(B)| : A \in \sigma(\{y_j : j \le p\}), \\ &B \in \sigma(\{y_j : j \ge p+m+1\}, \quad p \in \mathbb{Z})\}, \quad m \in \mathbb{N}. \end{aligned}$$

The sequence $\{y_j\}_{j\in\mathbb{Z}}$ is called strongly mixing (or α-mixing) if $\alpha(m) \to 0$ as $m \to \infty$.

- (CLT for strongly mixing sequences): Let $y_1, y_2, \ldots$ be a sequence of stationary of random variables with strong mixing coefficient $\alpha(\cdot)$.

  (i) Suppose that $\sum_{m=1}^{\infty} \alpha(m) < \infty$. Then

$$0 \le \sigma_\infty^2 \equiv \sum_{j=-\infty}^{\infty} Cov(y_1, y_{1+j}) < \infty. \tag{4.12}$$

If, in addition, $\sigma^2_\infty > 0$, then

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m}(y_j - Ey_1) \to N(0,\sigma^2_\infty). \tag{4.13}$$

(ii) Suppose that for some $\delta \in (0,\infty)$, $E|y_1|^{2+\delta} < \infty$ and $\sum_{m=1}^{\infty}[\alpha(m)]^{\delta/2+\delta} < \infty$. Then (4.12) holds. If in addition, $\sigma^2_\infty > 0$, then (4.13) holds.

As another example, we consider the first-order stationary autoregressive process. Suppose $\varepsilon_1,\varepsilon_2,\ldots$ are iid $N(0,\sigma^2)$. For $j \geq 1$,

$$y_{j+1} = \theta + \rho(y_j - \theta) + \varepsilon_{j+1}, \tag{4.14}$$

for some ρ with $|\rho| < 1$. Let $y_1 \sim N(\theta, \sigma^2_y)$. From the stationarity assumption, it can be shown that $Var(y_{j+1}) = \rho^2\sigma^2_y + \sigma^2$, and $Var(y_{j+1}) = \sigma^2_y = \frac{\sigma^2}{1-\rho^2}$ for all j. We may write

$$y_{r+1} - \theta = \rho^r(y_1 - \theta) + \rho^{r-1}\varepsilon_2 + \cdots + \rho\varepsilon_r + \varepsilon_{r+1}. \tag{4.15}$$

It is easy to see that $Cov(y_1, y_{1+r}) = \rho^r\frac{\sigma^2}{1-\rho^2}$. Therefore,

$$\begin{aligned} Var[\sqrt{m}(\overline{y}_m - \theta)] &= \sigma^2_y + \frac{2}{m}\sum_{j=1}^{m-1}(m-j)\rho^j\sigma^2_y \\ &= \frac{\sigma^2}{1-\rho^2}\left[1 + 2\sum_{j=1}^{m-1}\rho^j - \frac{2}{m}\sum_{j=1}^{m-1}j\rho^j\right]. \end{aligned}$$

Since

$$\sum_{j=1}^{m-1}j\rho^j = \frac{\rho\left\{1 + (m-1)\rho^m - m\rho^{m-1}\right\}}{(1-\rho)^2}, \tag{4.16}$$

we see that $\frac{2}{m}\sum_{j=1}^{m-1}j\rho^j \to 0$. Thus

$$Var\left[\sqrt{m}(\overline{y}_m - \theta)\right] \to \frac{\sigma^2}{1-\rho^2}(1 + \frac{2\rho}{1-\rho}) = \frac{\sigma^2}{1-\rho^2}\left(\frac{1+\rho}{1-\rho}\right). \tag{4.17}$$

Since $\overline{y}_m$ is normal, this implies that

$$\sqrt{m}(\overline{y}_m - \theta) \to^d N(0,\sigma^2_\infty), \tag{4.18}$$

where $\sigma_\infty^2 = \frac{\sigma^2}{1-\rho^2}\left(\frac{1+\rho}{1-\rho}\right)$. The ordinary central limit theorem gives us $\sqrt{m}(\overline{y}_m - \theta) \rightarrow N(0, \sigma^2/(1-\rho^2))$ since the stationarity implies $Var(y_j) = \sigma^2/(1-\rho^2)$ for all j.

If we want to estimate the sampling distribution of the random variable $T_m = \sqrt{m}(\overline{y}_m - \theta)$ using the standard bootstrap, then the bootstrap version T_m^* of T_m from $\chi_m = (y_{1,} \ldots y_m)$, equal number of bootstrap variables $y_{1,}^* \ldots y_m^*$, is given by

$$T_m^* = \sqrt{m}(\overline{y}_m^* - \overline{y}_m), \tag{4.19}$$

where $\overline{y}_m^* = \frac{1}{m}\sum_{j=1}^m y_j^*$. The conditional distribution of T_m^* under the standard bootstrap method still converges to a normal distribution, but with a wrong variance. This is justified as follows. First,

$$\sup_x |P^*(T_m^* \leq x) - \Phi(x/\sigma)| = o(1) \qquad \text{as} \quad m \rightarrow \infty, \quad \text{a.s.} \tag{4.20}$$

with $Ey_1 = \theta$, and $\sigma^2 = Var(y_1) \in (0, \infty)$. If the covariance $\sum_{j=1}^k Cov(y_1, y_{1+j}) \neq 0$ and $\sigma_\infty^2 \neq 0$, defined in (4.12), then for any $x \neq 0$,

$$\lim_{m\rightarrow\infty}[P^*(T_m^* \leq x) - P(T_m \leq x)] = [\Phi(x/\sigma) - \Phi(x/\sigma_\infty)] \neq 0 \quad \text{a.s.} \tag{4.21}$$

As a result in the previous examples, the standard bootstrap method fails drastically for dependent data. It ignores the dependence structure and fails to account for the lag-covariance terms in the asymptotic variance (Singh, 1981).

Let $\hat{\theta}$ be an estimator of a level-1 parameter θ and $T_m = \sqrt{m}(\hat{\theta}_m - \theta)/s_m$ be a scaled version of $\hat{\theta}_m$ such that $T_m \rightarrow^d N(0,1)$. If we set s_m to be an asymptotic standard deviation of $\sqrt{m}(\hat{\theta}_m - \theta)$, then T_m is called a *normalized* or *standardized* version of $\hat{\theta}_m$. If s_m is an estimator of the asymptotic standard deviation of $\sqrt{m}(\hat{\theta}_m - \theta)$, then T_m is called a *studentized* version of $\hat{\theta}_m$. Hall (1992) showed that it could be possible to expand the distribution function of T_m in a series of the form

$$P(T_m \leq x) = \Phi(x) + m^{-1/2}p_1(x;\gamma)\phi(x) + o(m^{-1/2}) \tag{4.22}$$

uniformly in $x \in \mathbb{R}$, where Φ and ϕ, respectively, denote the distribution function and the density (with respect to the Lebesque measure) of the standard normal distribution on $\mathbb{R}$ and where $p_1(\cdot;\gamma)$ is a polynomial such that its coefficients are smooth functions of some population parameters γ. The right side of (4.22) is called a first-order Edgeworth expansion for the distribution of T_m.

Next, let T_m^* denote the bootstrap version of T_m based on the bootstrap method. The expansion of T_m^* in an Edgeworth expansion of the form

$$P^*(T_m^* \leq x) = \Phi(x) + m^{-1/2} p_1(x;\hat{\gamma}_m)\phi(x) + o_p(m^{-1/2}) \tag{4.23}$$

uniformly in $x \in \mathbb{R}$, where $p_1(\cdot;\cdot)$ is the same function of (4.22) and where $\hat{\gamma}_m$ is a data-based version of the population parameter γ, generated by the resampling method. Relation (4.22) and (4.23) may be readily combined to assess the rate of approximation of the bootstrap distribution function estimator $P^*(T_m^* \leq x)$. It follows that

$$\begin{aligned} \sup_{x\in\mathbb{R}} & \quad |P^*(T_m^* \leq x) - P(T_m \leq x)| \\ &= m^{-1/2} \sup_{x\in\mathbb{R}} |p_1(x;\hat{\gamma}_m)\phi(x) - p_1(x;\gamma)\phi(x)| + o_p(m^{-1/2}) \\ &= o_p(m^{-1/2}), \end{aligned} \tag{4.24}$$

provided $\hat{\gamma}_m$ is a consistent estimator of γ and the coefficients of the polynomial $p_1(\cdot, t)$ is continuous in a second argument t. In this case, the bootstrap approximation $P^*(T_m^* \leq x)$ to $P(T_m \leq x)$ has a smaller order of error than the normal approximation to $P(T_m \leq x)$, which is only of the order $O(m^{-1/2})$. This property is referred to as the *second-order correctness* of the bootstrap approximation, as it captures the second-order term asymptotically.

Singh (1981) established the second-order correctness of the standard bootstrap method of Efron (1979) for the normalized sample mean of iid random variables, and provided the first theoretical confirmation of the superiority of the bootstrap approximation over the classical normal approximation.

The standard iid bootstrap estimation under independent assumption has

$$\sup_{x\in\mathbb{R}}\left|P^*\left\{\sqrt{m}\frac{\bar{y}_m^* - E^*(\bar{y}_m^*)}{\sqrt{Var^*(\sqrt{m}\bar{y}_m^*)}} \le x\right\} - P\left\{\sqrt{m}\frac{\bar{y}_m - \theta}{\sigma} \le x\right\}\right| = o_p(m^{-1/2}). \quad (4.25)$$

As a result, the standard bootstrap method fails for dependent data like below,

$$\sup_{x\in\mathbb{R}}\left|P^*\left\{\sqrt{m}\frac{\bar{y}_m^* - E^*(\bar{y}_m^*)}{\sqrt{Var^*(\sqrt{m}\bar{y}_m^*)}} \le x\right\} - P\left\{\sqrt{m}\frac{\bar{y}_m - \theta}{\sigma_\infty} \le x\right\}\right| \ne o_p(m^{-1/2}). \quad (4.26)$$

Proof.

$$\begin{aligned}
& \left|P^*\left\{\sqrt{m}\frac{\bar{y}_m^* - E^*(\bar{y}_m^*)}{\sqrt{Var^*(\sqrt{m}\bar{y}_m^*)}} \le x\right\} - P\left\{\sqrt{m}\frac{\bar{y}_m - \theta}{\sigma_\infty} \le x\right\}\right| \\
= & \left|P^*\left\{\sqrt{m}\frac{\bar{y}_m^* - E^*(\bar{y}_m^*)}{\sqrt{Var^*(\sqrt{m}\bar{y}_m^*)}} \le x\right\} - P\left\{\sqrt{m}\frac{\bar{y}_m - \theta}{\sigma} \le x\right\}\right. \\
+ & \left. P\left\{\sqrt{m}\frac{\bar{y}_m - \theta}{\sigma} \le x\right\} - P\left\{\sqrt{m}\frac{\bar{y}_m - \theta}{\sigma_\infty} \le x\right\}\right| \\
= & \; o_p(m^{-1/2}) + O_p(1) \\
= & \; O_p(1). \qquad (4.27)
\end{aligned}$$

The standard bootstrap cannot be taken into account for the dependent structure. Note that in (4.22) $\hat{\gamma}_m$ is not a consistent estimator of γ. Recall that $\sigma^2 = Var(y_1)$ and $\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} Cov(y_1, y_{1+k})$. Figure 1 is the Lynx data collected by Brockwell and Davis (1991) which contain data about Canadian Lynx Trappings in 1821-1934. Figure 2 shows the results for a single replicate using block simulation, and tells us that the standard bootstrap method fails to reproduce the original dependent structure. Figure 3 presents several series with autocorrelated data. In the standard bootstrap methods, the dependence structure could not be preserved. The moving block bootstrap should be used for dependent data.

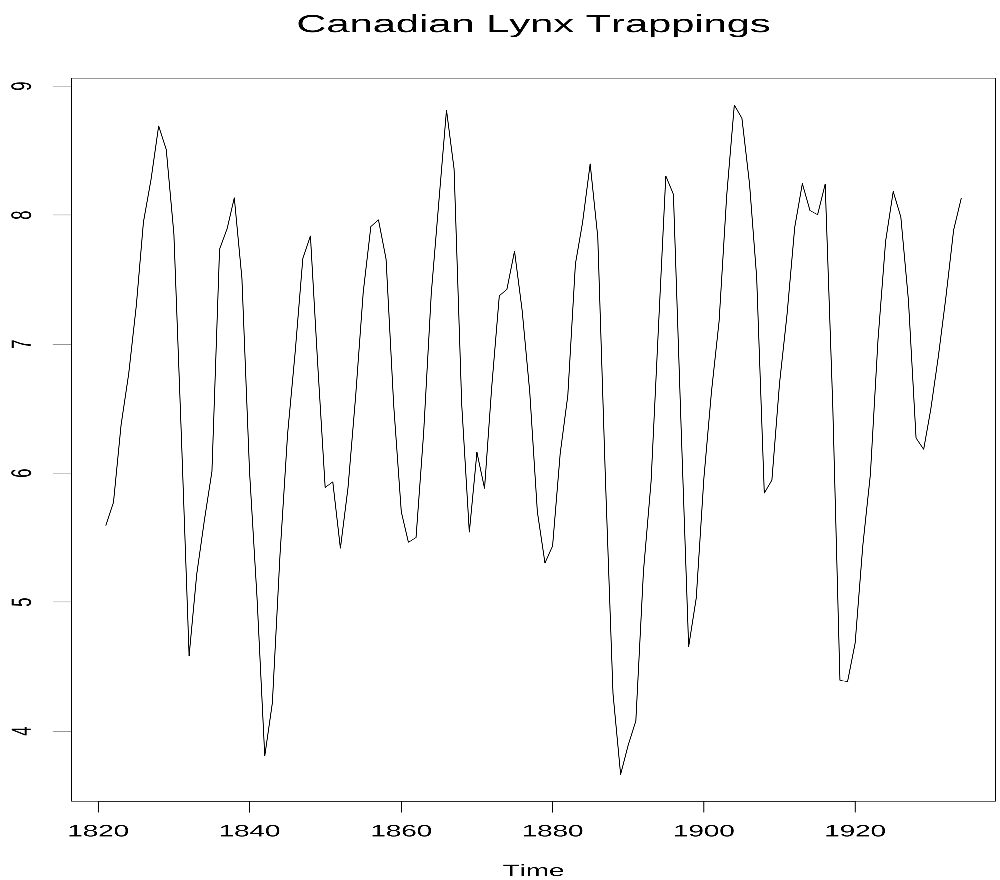

Figure 1: Canadian lynx trappings in 1821-1934

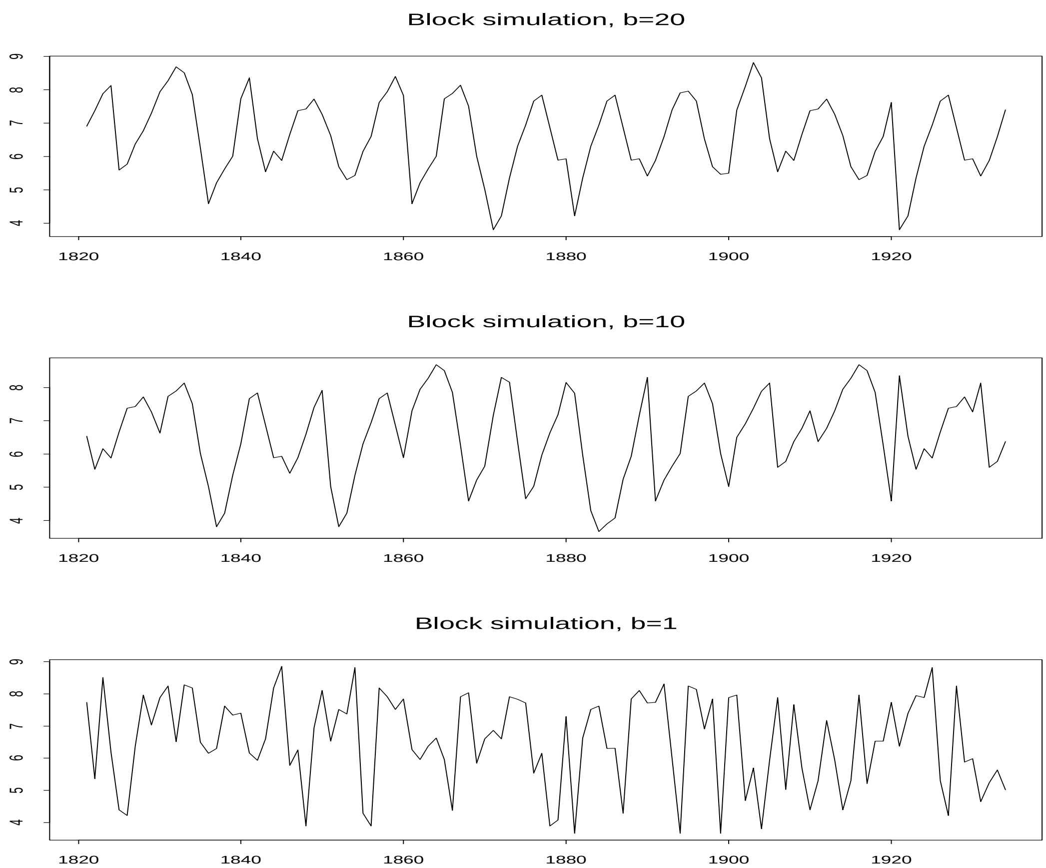

Figure 2: The block simulation for Canadian lynx trapping data

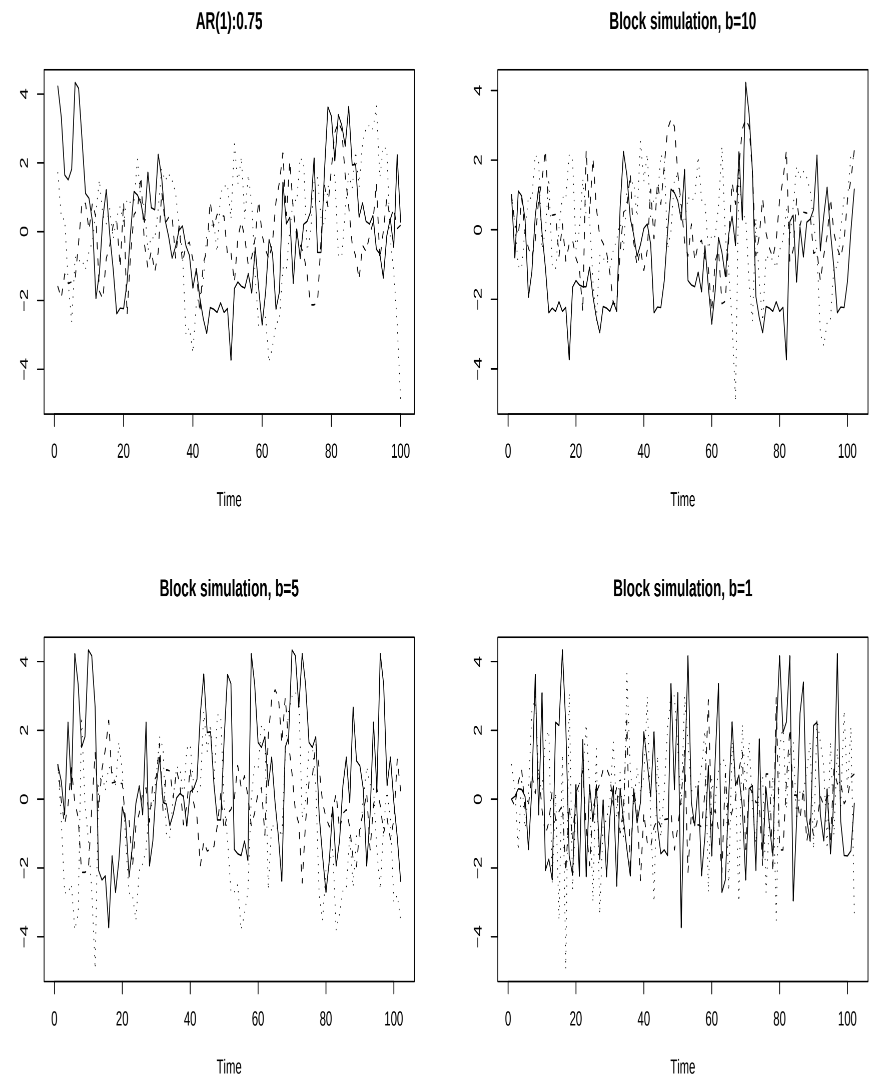

Figure 3: The block simulation for three AR(1) series

4.2.1 Short and Long Memory Process

In this section, we consider the correlation structure. We assume that the process follows stationarity condition for each subject. The series $\{y_{ij}, j \in \mathbb{Z}\}$, with index set $\mathbb{Z} = \{0, \pm1, \pm2, \cdots\}$, $1 \leq i \leq n_0$, is said to be stationary if

$$i) \qquad \mathrm{E}|y_{ij}|^2 < \infty \quad \text{for all } j \in \mathbb{Z},$$

$$ii) \qquad \mathrm{E}y_{ij} = 0 \quad \text{for all } j \in \mathbb{Z},$$

and

$$iii) \qquad \gamma_{y_i}(r, s) = \gamma_{y_i}(r + j, s + j) \quad \text{for all } r, s, j \in \mathbb{Z}.$$

If $\{y_{ij}, j \in \mathbb{Z}\}$ is stationary then $\gamma_{y_i}(r, s) = \gamma_{y_i}(r - s, 0)$ for all $r, s \in \mathbb{Z}$. It is therefore convenient to redefine the auto-covariance function of a stationary process as the function of just one variable,

$$\gamma_{y_i}(k) \equiv \gamma_{y_i}(k, 0) = \mathrm{cov}(y_{i(j+k)}, y_{ij}) \quad \text{for all } j, k \in \mathbb{Z}. \tag{4.28}$$

The function $\gamma_{y_i}(\cdot)$ will be referred to as the the auto-covariance function of $\{y_{ij}\}$ and $\gamma_{y_i}(k)$ as its value at lag k. The autocorrelation function (ACF) of $\{y_{ij}\}$ is defined analogously as the function whose value set lag k is

$$\rho_{y_i}(k) \equiv \frac{\gamma_{y_i}(k)}{\gamma_{y_i}(0)} = \mathrm{Corr}(y_{i(j+k)}, y_{ij}) \quad \text{for all } j, k \in \mathbb{Z}. \tag{4.29}$$

Definition 4.2.4 *(Short Memory and Long Memory). The covariance between y_{i1} and $y_{i(1+k)}$ decrease rapidly as $k \to \infty$. The autocorrelation function is geometrically bounded, i.e. ,*

$$|\rho(k)| \leq Cr^{-k}, \quad k = 1, 2, \cdots, \tag{4.30}$$

where $C > 0$ *and* $0 < r < 1$ *which is called a "short memory process". "A long memory process" is a stationary process for which*

$$\rho(k) \sim Ck^{2d-1} \quad as \quad k \to \infty, \tag{4.31}$$

where $C > 0$ *and* $d < 1/2$. *[Some authors make a distinction between "intermediate memory" process for which* $d < 0$ *and hence* $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$, *and "long memory" process for which* $0 < d < 1/2$ *and* $\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$ *].*

All stationary invertible *autoregressive moving average* (ARMA) processes are short memory. In the case $d > 0$, the autocorrelations decay to zero so slowly that they are not summable, i.e $\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$ and $Var(y_{ij})$ decays to zero more slowly than $1/m$. If $d < 0$, then the autocorrelation are summable, $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$, but they still decay to zero more slowly than the exponential rate achieved by the stationary invertible ARMA process. There is a non-negligible correlation even between distant past and distant future. We will use the terminology of Brockwell and Davis (1991) for long memory whenever $d \neq 0$.

When the sample ACF of a time series decays slowly, there is a need to difference the series until it seems stationary. Lone memory time series were considered in Hosking (1981) and Granger and Joyeux (1980) as intermediate compromises between the short memory ARMA models and the fully integrated nonstationary processes in the Box-Jenkins sense. Figure 4 shows the correlation decays exponentially with the difference in time for the stationary and short memory case.

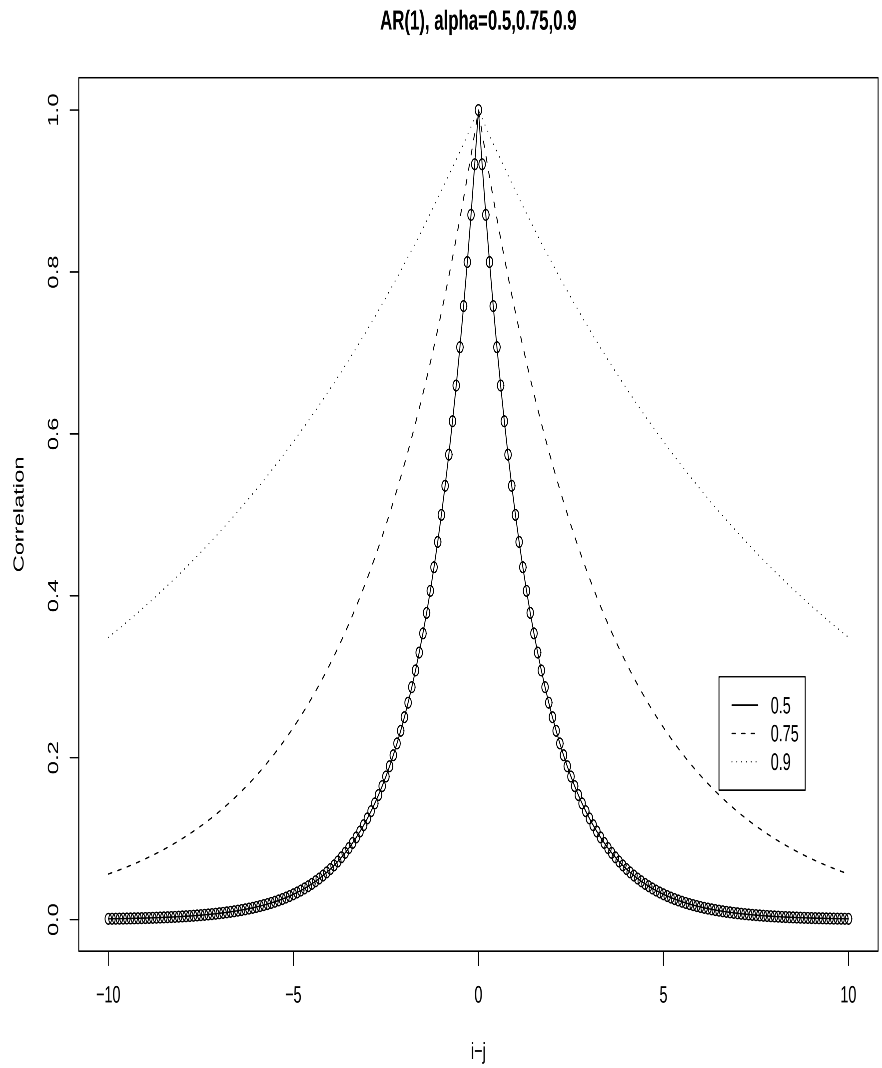

Figure 4: The correlation decays exponentially with the difference in time

4.3 Moving Block Bootstrap Methods for Longitudinal Data

The major drawback with model-based resampling is that in practice not only the parameters of a model, but also its structure, must be identified from the data. If the chosen structure is incorrect, the resampled series will be generated for the wrong model, and they will not have the same statistical properties as the original data. The model-based approach is inconsistent if the model used for resampling is misspecified.

The moving bootstrap involves resampling possibly overlapping blocks. The MBB does not force one to select a model and the only parameter required is the block length. If the block is long enough the original dependence will be reasonably preserved in the resampled series. This approximation is better if the dependence is weak and the blocks are as long as possible, thus preserving the dependence more faithfully. But the distinct values of the statistics must be as numerous as possible to provide a good estimate of the distribution of the statistics and this points toward short blocks.

Unless the length of the series is considerable to accommodate longer and more number of blocks the preservation of the dependence structure may be difficult, especially for complex, long range dependent structures. In such cases, the block resampling scheme tends to generate resampled series that are less dependent than the original ones. Furthermore, the resampled series often exhibits artifacts which are caused by joining randomly selected blocks. Then, the asymptotic variance-covariance matrices of the estimators based on the original series and those based on the bootstrap series are different and a modification of the original scheme is needed. This suggests a strategy intermediate between model-based and block resampling. The idea comes from pre-whitening the series by fitting a model is intended to remove much of the dependence between the original observations. A resampling series is generated by block resampling of residuals from the simple fitted model, and the innovation series is then post-blackened by applying the simple estimated model to the resampled innovations. The post-blackened version works more consistently in practice (Davison and Hinkley, 1997).

Bühlmann (1997) suggested the sieve bootstrap which is model based. The AR(p) model is just used to filter the residual series. If the model used in the sieve bootstrap is not appropriate, the resulting residuals cannot be treated as iid. A hybrid approach between the model based method and moving block bootstrap, named post-blacken bootstrap, was suggested by Davison and Hinkley (1997). The procedure is similar to the sieve bootstrap, but the residuals from AR(p) model are not resampled in an iid manner but by using the MBB bootstrap. If some residual dependent structure is present in the AR residuals, this is kept from the blockwise bootstrap. The simple linear model is used to pre-whiten the series by fitting the model that is intended to remove much of the dependence present the observations. A series of innovations is then generated by block resampling of residuals obtained from the fitted model, and the innovation series is then post-blackened by applying the estimated model to the resampled innovations.

The Block Bootstrap Algorithm in Longitudinal Model

We continue to assume (4.1) as our longitudinal model under consideration.

1) Let $\hat{e}_{ij},\ i = 1, \cdots, n_0,\ j = 1, \cdots, m$ be the residuals from the model fit.

$$\hat{e}_{ij} = y_{ij} - x_{ij}\hat{\beta},$$

where $\hat{\beta}$ is the ordinary least square estimate.

2) Now assuming that $m = bk$ with b and k integers: Let $B_1^*, \cdots, B_k^*$ denotes k uniform draws with replacement from the integers $\{0, 1, \cdots, m-b\}$. These represent

the starting point for each block of length b. A block bootstrap resample of residuals, $(\hat{e}^*_{i1}, \cdots, \hat{e}^*_{im})$, is defined by:

$$\hat{e}^*_{i,(j-1)b+s} = \hat{e}_{i,B^*_j+s}, \quad (1 \le j \le k, 1 \le s \le b) \text{ for each } \quad i.$$

3) The bootstrapped response, y^*_{ij}, are then generated from the estimated model with residuals $\hat{e}^*_{ij}$ and the original covariates:

$$y^*_{ij} = x_{ij}\hat{\beta} + \hat{e}^*_{ij}. \tag{4.32}$$

4) From the resampled responses y^*_{ij}, and original covariates we fit the model and obtain new parameter estimates.

5) Repeating steps 2) through 4) a large number, R, of times one obtains R bootstrap replicates from which features of the distribution of the parameter estimates can be estimated. In particular, the bootstrap variance estimates are simply variance of the B computed values for each parameter.

We consider the six different kinds of block bootstrap methods in a balanced longitudinal design in which the number of subjects is small and the number of replications is large.

1) MBB1: Within block bootstrap

For each i subject, we construct overlapping blocks with $m - b + 1$ blocks and block size b, i.e $B_1, \cdots, B_{m-b+1}$. Let us define $m/b = k$ which is assumed to be an integer for simplicity, in general $k = [m/b]$. We can add the k blocks with replacement among $B_1, \cdots, B_{m-b+1}$. We get the $B^*_1, \cdots, B^*_k$ with $kb = m$, and create $\hat{e}^*_{i1}, \cdots, \hat{e}^*_{im}$ from $\hat{e}_{i1}, \cdots, \hat{e}_{im}$, where $\hat{e}_{ij} = y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 x_{ij}$. We can add up to n_0 individuals and plug this into the model and the results is a pseudo sample series $y^*_{11}, \cdots, y^*_{n_0 m}$. From the model $y^*_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{e}^*_{ij}$, we fit model and produce new parameters $\hat{\beta}^*_0$ and $\hat{\beta}^*_1$.

2) MBB2: Mixed block bootstrap

We have $m-b+1$ blocks and block size b, i.e $B_1, \cdots, B_{m-b+1}$ and add up to n_0 subjects. We sample n_0k blocks with replacements among $B_1, \cdots, B_{n_0(m-b+1)}$. We construct $B_1^*, \cdots, B_{n_0k}^*$ with $kb = m$, and plug this into the model and obtain a pseudo series $y_{11}^*, \cdots, y_{n_0m}^*$. Similarly, from the model $y_{ij}^* = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{e}_{ij}^*$, we fit the model and produce new parameters $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$.

3) One-line moving block bootstrap

One can make up to one long series and perform the moving block bootstrap using a time series without splitting the different individual consecutive data.

4) Standard bootstrap

This is a special case of $b = 1$ in MBB2.

5) Resampling subject bootstrap

This is a special case of $b = m$ in MBB2.

6) Stratified standard bootstrap

This is a special case of $b = 1$ in MBB1.

4.4 Justification of Moving Block Bootstrap in Longitudinal Data

We consider the justification of moving block bootstrap in longitudinal data. We focus on the within block bootstrap method (MBB1) in the six different kinds of synario in previous section. Let's consider the relationship between the GEE and M-estimators. The robust approach can be extended to the regression setup to analyze a predictor-outcome relationship. Suppose we have model (4.1) with $n = n_0$. The estimator $\widehat{\beta}$ is called a robust regression estimator or an M-estimator if it solves

$$\sum_{i=1}^{n_0} \sum_{j=1}^{m} x_{ij}' \psi(y_{ij} - x_{ij}\beta) = 0, \tag{4.33}$$

for some choice of function $\psi(\cdot)$.

4.4.1 Expansion for M-estimator

It is known that

$$(\sum_{i=1}^{n_0}\sum_{j=1}^{m} x_{ij}^{'}x_{ij})^{1/2}(\hat{\beta}_{n_0m}-\beta) \sim N_p\Big(0, \frac{\mathrm{E}\psi^2(e_{11})}{(\mathrm{E}\psi^{'}(e_{11}))^2} I_p\Big), \tag{4.34}$$

where I_p denotes the identity matrix of order p.

Let $\hat{e}_{ij} = y_{ij} - x_{ij}\hat{\beta}_{n_0m}$ denote residuals. Define

$$\begin{aligned}
\sigma_i(k) &= \mathrm{E}\psi(e_{i1})\psi(e_{i(1+k)}), \qquad k \geq 0; \qquad \tau = \tau_i = \mathrm{E}\psi^{'}(e_{i1}) \quad \text{for each} \quad i \\
\hat{\sigma}_{im}(k) &= (m-k)^{-1}\sum_{j=1}^{m-k}\psi(\hat{e}_{ij})\psi(\hat{e}_{i(j+k)}), \qquad 0 \leq k \leq m-1, \qquad \hat{\tau}_{im} = m^{-1}\sum_{j=1}^{m}\psi^{'}(\hat{e}_{ij}).
\end{aligned}$$

Also, let $\sigma_i(k) = \sigma(k)$ and $\hat{\sigma}_{im}(k) = \hat{\sigma}_m(k)$.

Assumption 4.4.1 *(A.1) (i) ψ is twice differentiable, and $\psi^{''}$ satisfies a Lipschitz condition of order $\delta_1 > 0$,*

(ii) $\psi, \psi^{'}, \psi^{''}$ are bounded.

(A.2) (i) for each i $\mathrm{E}\psi(e_{i1}) = 0$, $\tau \equiv \mathrm{E}\psi^{'}(e_{i1}) \neq 0$,

(ii) $\sigma_\infty \equiv \sigma(0) - 2\sum_{k=1}^{\infty}|\sigma(k)| > 0$.

(A.3) There exists $\rho > 0$ such that

(i) $\sup\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^{r}, \quad B \in \mathcal{F}_{r+k}^{\infty}, r \geq 1\} \leq \rho^{-1}\exp(-\rho k)$ for all $k \geq 1$,

(ii) for all $r \geq 1$, and all $k \geq \rho^{-1}$, there exists a $\mathcal{F}_{r-k}^{r+k}$-measurable random variable $\tilde{e}_{ir,k}$ such that $\mathrm{E}|e_{ir} - \tilde{e}_{ir,k}| \leq \rho^{-1}\exp(-\rho k)$,

(iii) for all $r, k, q \geq \rho^{-1}$ and $A \in \mathcal{F}_{r-q}^{r+q}, \mathrm{E}|P(A|\mathcal{F}_j : j \neq r) -$ $P(A|\mathcal{F}_j : 0 < |j-r| \leq q+k)| \leq \rho^{-1}\exp(-\rho k)$, and

(iv) for all $r \geq \rho^{-1}, k \leq r$ and all $t_{r-k}, \ldots, t_{r+k} \in \mathbb{R}$ with $|t_r| > \rho$, $\mathrm{E}|\mathrm{E}(\exp(\sqrt{-1}\sum_{i=1}^{n_0}\sum_{j=r-k}^{r+k} t_j\psi(e_{ij}))|\mathcal{F}_j : j \neq r| < \exp(-\rho)$.

(A.4) $\max\{\|x_{ij}\| : 1 \leq j \leq m\} = O(1)$ and $\lim inf_{m\to\infty} m^{-1}\lambda_m \equiv \lambda > 0$, where λ_m denotes the smallest eigenvalue of $(\sum_{j=1}^{m} x_{ij}^{'}x_{ij})$.

Let $D_{n_0m} = (\sum_{i=1}^{n_0}\sum_{j=1}^{m} x_{ij}^{'}x_{ij})^{-1/2}$ and $d_{ij} = D_{n_0m}x_{ij}^{'}, \quad 1 \le j \le m$. When e_{ij} are weakly dependent for each i, the asymptotic covariance of $D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta)$ matrix is given by

$$Cov_{n_0m} \equiv (\mathrm{E}\psi^{'}(e_{11}))^{-2} \times \sum_{i=1}^{n_0}\sum_{k=0}^{m} L_{ikm}\mathrm{E}\psi(e_{i1})\psi(e_{i(1+k)}), \tag{4.35}$$

where $L_{i0m} = I_p$ and $L_{ikm} = \sum_{j=1}^{m-k}(d_{ij}d_{i(j+k)}^{'} + d_{i(j+k)}d_{ij}^{'}), 1 \le k \le m-1$.

To define the studentized version of $\hat{\beta}_{n_0m}$, note that the asymptotic matrix $D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta)$ is given by

$$\Sigma_{n_0m} \equiv Cov\Big(\sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}\psi(e_{ij})\Big) = \sum_{i=1}^{n_0}\sum_{k=0}^{m-1} L_{ikm}\sigma(k).$$

Therefore, a natural estimator of Σ_{n_0m} is

$$\hat{\Sigma}_{n_0m} = \sum_{i=1}^{n_0}\sum_{k=0}^{l} L_{ikm}\hat{\sigma}_m(k),$$

where $1 \le l \equiv l_m \le m-1$ is an integer. If $l \to \infty$ slowly with m, then $\|\hat{\Sigma}_{n_0m} - \Sigma_{n_0m}\| = o_p(1)$. $\hat{\Sigma}_{n_0m}$ is non singular with high probability for m large, and can be inverted to define the studentized statistic,

$$T_{n_0m} = \hat{\Sigma}_{n_0m}^{-1/2}D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta).$$

Next, we extend Lahiri's (1996) results for longitudinal case:

assume that (A.1),(A.2),(A.3)(i),(ii), and (A.4) hold. Then, there exists a sequence of statistics $\{\hat{\beta}_m\}$ such that

$$P(\hat{\beta}_{n_0m} \text{ satisfies } (4.33) \text{ and } \|D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta)\|^2 \le C\log m) = 1 - o(m^{-1/2}).$$

If we have a unique solution $\hat{\beta}_{n_0m}$, then $\|D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta)\| = O_P((\log m)^{1/2})$. When (4.33) has a unique solution, one can obtain the strong consistency of $\hat{\beta}_{n_0m}$ as in Lahiri (1992). The next result gives a first order Edgeworth expansion for the studentized M-estimator.

Theorem 4.4.1 *Assume that Assumptions (A.1)-(A.4) hold and that* $\{\hat{\beta}_{n_0 m}\}$ *is a sequence of measurable solutions of (4.33). Then, there exist a polynomial* $p_m(\cdot)$ *on* $\mathbb{R}^p$ *such that*

$$\sup_{B\in\mathcal{B}}\left|P(T_{n_0 m}\in B)-\int_B(1+p_{n_0 m}(x))d\Phi(x)\right|=o(m^{-1/2}) \tag{4.36}$$

for every class $\mathcal{B}$ *of Borel subsets of* $\mathbb{R}^p$ *satisfying*

$$\sup_{B\in\mathcal{B}}\Phi((\partial B)^{\epsilon})=O(\epsilon)\quad as\quad \eta\downarrow 0. \tag{4.37}$$

Here $\|p_{n_0 m}\phi\|_\infty = O(m^{-1/2})$ *with sup norm* $\|\|_\infty$, Φ *denotes the standard normal distribution on* $\mathbb{R}^p$ $(p\geq 1)$, *and the coefficient of* $p_{n_0 m}(\cdot)$ *are continuous functions of cross-product moments of* $\psi(e_{ij}), \psi^{'}(e_{ij})$, *and* $\psi^{''}(e_{ij})$. *Here* ∂B *denote the boundary of a set* $B\subseteq\mathbb{R}^p$ *and* $(\partial B)^\epsilon = \{x : \|x-y\| < \epsilon$ *for some* $y\in\partial B\}$.

Proof.

We follow Lahiri (1996) notation and definitions. For a smooth function $h:\mathbb{R}^p\to\mathbb{R}$, let us $D_j h$ denote the partial derivative of $h(x)$ with respect to the jth coordinate of $x, 1\leq j\leq p$. For $p\times 1$ vectors $\nu=(\nu_1,\ldots,\nu_p)'\in\mathbb{Z}_+^p$ and $w=(w_1,\ldots,w_p)'\in\mathbb{R}^p$, let $|\nu|=\nu_1+\cdots+\nu_p,\ \nu!=\nu_1!\cdots\nu_p!,\ w^\nu=\Pi_{i=1}^p(w_i)^{\nu_i}$, and $\|w\|=(w_1^2+\cdots+w_p^2)^{1/2}$. Let D^ν denote the differential operator $D_1^{\nu_1}\cdots D_p^{\nu_p}$, namely $D^\nu=\Pi_{j=1}^p\left(\frac{\partial}{\partial t^{(j)}}\right)^{\nu(j)}$. For $\nu\in\mathbb{Z}_+^p$ with $1\leq|\nu|\leq s$, let χ_ν denote the νth cumulant and μ_ν is the νth moment of w. Note that $\sqrt{-1}^{|\nu|}\mu_\nu=D^\nu\hat{\Phi}(0),\ D^\alpha\hat{\mu}(t)=(\sqrt{-1})^{|\alpha|}\int w^\alpha e^{\sqrt{-1}t'w}\mu(dw),\ t\in\mathbb{R}^p$, and $\sqrt{-1}^{|\nu|}\chi_\nu=(D^\nu\log\hat{\Phi})(0)$.

We consider w is a $\mathbb{R}^p$-valued random vector with $\mathrm{E}w=0$ and $\mathrm{E}\|w\|^s<\infty$ for some integer $s\geq 3$.

Let $m_3=[\log m\log\log(3+m)], v_{1m}=m^{-1/2}(\log m)^{1/2}, v_m=m^{-1/2}$,

$v_{2m} = v_m(\log m)^{-1}$, and $v_{3m} = v_m(\log m)^{-3/2}$. Furthermore, define

$$\begin{aligned}
G_{n_0m} &= D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta), \quad G_{1n_0m} = \sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}\psi(e_{ij}) \quad d_{ij} = (\sum_{i=1}^{n_0}\sum_{j=1}^{m} x_{ij}^{'}x_{ij})^{-1/2}x_{ij},\\
D_{n_0m} &= (\sum_{i=1}^{n_0}\sum_{j=1}^{m} x_{ij}^{'}x_{ij})^{-1/2}, \quad \tilde{\psi} = \psi(\cdot) - \hat{\mu}_{n_0m}, \quad A_{n_0m} = \sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}d_{ij}^{'}\psi^{'}(e_{ij}),\\
A &= \tau I_p = \mathrm{E}\psi^{'}(e_{ij})I_p, \quad \tau = \tau_i = E\psi^{'}(e_{i1}) \quad \text{for each} \quad i,\\
W_{i1j} &= \psi(e_{i1}), \quad W_{i2j} = \psi^{'}(e_{ij}) - \tau, W_{i3j} = \psi^{''}(e_{ij}) - \mathrm{E}\psi^{''}(e_{i1}), \quad \text{and}\\
W_{i4j}(k) &= \psi(e_{ij})\psi(e_{i(j+k)}) - \sigma_i(k).
\end{aligned}$$

Also, write $\chi(U) = (-1)^{p/2}D_1 \cdots D_p\mathrm{E}\exp(\sqrt{-1}t^{'}U)|_{t=0}$ for a random vector U in $\mathbb{R}^p$.

Let $\Delta = D_{n_0m}^{-1}(t-\beta), t \in \mathbb{R}^p$. Then, by Taylor's expansion, one can get

$$\begin{aligned}
\Big[\sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}d_{ij}^{'}\psi^{'}(e_{ij})\Big]\Delta = \quad & \sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}\psi(e_{ij}) + \frac{1}{2}\sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}(d_{ij}^{'}\Delta)^2\psi^{''}(e_{ij})\\
& + R_{n_0m}(t), \qquad (4.38)
\end{aligned}$$

where $\|R_{n_0m}(t)\| \le C\sum_{i=1}^{n_0}\sum_{j=1}^{m}\|d_{ij}\|^{3+\delta_1}\|\Delta\|^{2+\delta}, t \in \mathbb{R}^p$.

Following Lahiri (1992), we obtain that

$$\begin{aligned}
G_{n_0m} = \quad & (A^{-1} + \tau^{-2}(A - A_{n_0m}))G_{1n_0m} + (2\tau^3)^{-1}\sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}(d_{ij}^{'}\theta_{1n_0m})^2\mathrm{E}\psi^{''}(e_{ij})\\
& + R_{n_0m}^{'}, \qquad (4.39)
\end{aligned}$$

where $P(\|R_{n_0m}^{'}\| > C(\sigma_\infty)v_{2m}) = o(v_m)$,

$$\begin{aligned}
T_{n_0m} = \quad & \Sigma_{n_0m}^{-1/2}\Big[G_{1n_0m} + \tau^{-1}G_{1n_0m}\Big(m^{-1}\sum_{j=1}^{n_0}\sum_{j=1}^{m} W_{i2j}\\
& -(m\tau)^{-1}\sum_{i=1}^{n_0}\sum_{j=1}^{m}(d_{ij}^{'}G_{1n_0m})\mathrm{E}\psi^{''}(e_{i1})\Big)\\
& + \ \tau^{-1}(A - A_{n_0m})G_{1n_0m} + (2\tau^2)^{-1}\sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}(d_{ij}^{'}G_{1n_0m})^2\mathrm{E}\psi^{''}(e_{i1})\Big]\\
& + \sum_{|\beta|=1}(\hat{\Sigma}_{n_0m})^{\beta}D^{\beta}(\Sigma_{n_0m}^{-1/2})G_{1n_0m} + R_{n_0m}^{\dagger}. \qquad (4.40)
\end{aligned}$$

If we have $T_{n_0m} = T_{1n_0m} + R_{n_0ms}$, where R_{n_0ms} is the remainder term that under the moment condition $E\|y_{11}\|^s < \infty$ satisfies $P(\|R_{n_0ms}\| > \delta_{m,s}) = \delta_{m,s}$ for some sequence $\delta_{m,s} = o(m^{-(s-2)/2})$, then the random variable T_{1n_0m} is called a $(s-2)$th order stochastic approximation to T_{n_0m}. Note that the $(s-2)$th order Edgeworth expansions for T_{n_0m} and T_{1n_0m} coincide. The reason for T_{1n_0m} is that the first term is the same as T_{n_0m}, but the remaining terms consist of all independent variables for deriving a more simple expansion. The stochastic approximation T_{1n_0m} can be expressed in the form

$$\begin{aligned} T_{1n_0m} = \quad & \Sigma_{n_0m}^{-1/2} G_{1n_0m} + \sum_{r=1}^{p} G_{1n_0m}^{'} \Lambda_{n_0rm} G_{1n_0m} q_r + \sum_{r=1}^{p} \tilde{W}_{2n_0m}^{'} \Lambda_{1n_0rm} G_{1n_0m} q_r \\ & + \sum_{|\nu|=1} (\hat{\Sigma}_{1n_0m})^{\nu} \Lambda_{\nu m} G_{1n_0m}, \end{aligned} \tag{4.41}$$

where $q_1 = (1, 0, \ldots, 0), \cdots, q_p = (0, 0, \ldots, 1)$ are the standard basis of $\mathbb{R}^p$, $\tilde{W}_{2n_0m} = ((A - A_{n_0m})^{'} : \sum_{i=1}^{n_0} m^{-1} \sum_{j=1}^{m} W_{i2j})^{'}$ and $\Lambda_{n_0rm}, \Lambda_{1n_0rm}, \Lambda_{\nu m}$ are nonrandom matrices satisfying $\max\{m^{1/2}\|\Lambda_{n_0rm}\| + \|\Lambda_{1n_0rm}\| + \|\Lambda_{\nu m}\| : 1 \leq r \leq p, |\nu| = 1\} = o(1)$. In the following C, $C(\cdot)$ dentes pure constants which depend on each arguments, and the dependance of $C(\cdots)$ on p, α, and the finite moments of $\psi(e_{ij}), \psi^{'}(e_{ij})$, and $\psi^{''}(e_{ij})$ will be suppressed for notational simplicity. Using Lahiri's (1992,1996) arguments, we can show that

$$P\Big(\|\sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}\psi(e_{ij})\| > C(\log m)^{1/2}\Big) \;=\; o(m^{-1/2}),$$

$$P\Big(\|A_{n_0m} - A\| > Cm^{-1/4}(\log m)^{-2}\Big) = o(m^{-1/2}), \tag{4.42}$$

$$P\Big(|\sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{iju}d_{ijl}d_{ijz}| > Cm^{-5/8}\Big) \;\leq\; Cm^{-3/4} \tag{4.43}$$

for all $1 \leq u, l, z \leq p$, where d_{iju} denote the uth component of d_{ij}, we have $T_{n_0m} = T_{1n_0m} + R_{n_0m}$, where $P(\|R_{n_0m}\| > Cv_{2m}) = o(v_m)$. We know that the first order Edgeworth expansions for T_{1n_0m} and T_{n_0m} coincide.

Let $G_{2n_0m} = \sum_{i=1}^{n_0}\sum_{j=a}^{m} d_{ij}W_{i1j}$ with $a = [m^{(1-2\delta_0)/2}]$,

$A_{2n_0m} = ((\sum_{i=1}^{n_0}\sum_{j=a}^{m} d_{ijk}d_{ijl}W_{i2j}))_{p\times p}$,

$\hat{W}_{2n_0m} = ((A'_{2n_0m} - \mathrm{E}A'_{2n_0m}) : \sum_{i=1}^{n_0} m^{-1}\sum_{j=1}^{m} W_{i2j})'$,

$\hat{\Sigma}_{2n_0m} = \sum_{i=1}^{n_0}\sum_{k=0}^{l}(m-k)^{-1}L_{ikm}\sum_{j=a}^{m-k}[W_{i4j} - G'_{2n_0m}\gamma_{ijk}]$, and

$$\begin{aligned} T_{2n_0m} = \quad & \Sigma_{n_0m}^{-1/2}G_{1n_0m} + \sum_{r=1}^{p} G'_{2n_0m}\Lambda_{n_0rm}G_{2n_0m}q_r + \sum_{r=1}^{p}\hat{W}'_{2n_0m}\Lambda_{1n_0rm}G_{2n_0m}q_r \\ & + \sum_{|\nu|=1}(\hat{\Sigma}_{2n_0m})^{\nu}\Lambda_{\nu m}G_{2n_0m}. \end{aligned}$$

The reason for T_{2n_0m} is that the first term is the same as T_{1n_0m}, but the remaining terms consist of truncated independent variables for obtaining the simplified forms of expansion. Using an Edgeworth expansion under dependence for T_{1n_0m} (Lahiri (1994; 1996)), we have

$$P(\|T_{1n_0m} - T_{2n_0m}\| > Cv_m) = o(v_m). \tag{4.44}$$

Let $Q_{n_0m}(t) = E\exp(\sqrt{-1}t'T_{2n_0m}), t \in \mathbb{R}^p$. We have the reduction to truncated statistics T_{2n_0m}, and obtain the following result for the Fourier transform of the Edgeworth expansion for the density of T_{2n_0m}

$$\max_{|\alpha|\le p+1}\int_{\Gamma_m}|D^{\alpha}(Q_{n_0m}(t) - \Psi_{n_0m}(t))|dt = o(v_m), \tag{4.45}$$

where $\Gamma_m = \{t \in \mathbb{R}^p : \|t\| < v_{3m}^{-1}\}$, and Ψ_{n_0m} is a Fourier transform which can be defined as in (4.47) (cf. Lahiri (1994) p.216). Next we write $t_m = t'\Sigma_{n_0m}^{-1/2}$. Also, using the results of Lahiri (1994), we have

$$|D^{\alpha}[\mathrm{E}(1 + \sqrt{-1}t'\Delta_{n_0m})\exp(\sqrt{-1}t'_mG_{1n_0m}) - \Psi_{n_0m}(t)]| \le C(\alpha)m^{-1/2}m^{-\delta} \tag{4.46}$$

for some constant $C(\alpha)$ and $\delta > 0$, where $\Delta_{n_0m} = T_{2n_0m} - \Sigma_{n_0m}^{-1/2}G_{1n_0m}$ and

$$\exp(\|t\|^2/2)\Psi_{n_0m}(t) = 1 + \mathrm{E}(\sqrt{-1}t'_mG_{1n_0m})^3/3! + R_{\Psi}, \tag{4.47}$$

where

$$
\begin{aligned}
R_\Psi &= \sqrt{-1}\sum_{h=1}^{n_0}\sum_{j=1}^{m}\sum_{k=1}^{m}\mathrm{E}\Big[v_{h1jk}W_{h1j}W_{hjk} \\
&+ \hat{W}'_{2n_0m,j}\Big(\sum_{r=1}^{p}(t'_m q_r)\Lambda_{1n_0rm}d_{hj}\Big)W_{j1k}\Big] \\
&\times [1 - t'_m G_{3n_0m}(1,\{i\})G_{3n_0m}(1,\{j\})' t_m] \\
&- \sqrt{-1}\sum_{h=1}^{n_0}\sum_{j=1}^{m}\sum_{k=0}^{l}\sum_{i=1}^{m}\Big(\sum_{|\nu|=1}(L_{n_0km})^\nu (m-k)^{-1}t'\Lambda_{\nu m}d_{hj}\Big) \\
&\times \mathrm{E}\{W_{4n_0i}(k)W_{1n_0j}(1 \\
&- t'_m G_{3n_0m}(1,\{i,k\})t'_{n_0m}G_{3n_0m}(1,\{j\}))\}.
\end{aligned} \tag{4.48}
$$

Therefore, by Taylor expansion for $m_3 < \|t\| \le m_3(m/a)^{1/2}$ and weak dependence of Δ_{n_0m} (Lahiri (1994) p.216)

$$
\int_{\Gamma_{1m}} |D^\alpha Q_{n_0m}(t)|dt = o(v_m) \tag{4.49}
$$

for all $\|\alpha\| \le p+1$, where $\Gamma_{1m} = \{t \in \Gamma_m : \|t\| > m_3\}$. Combining the results in (4.45)-(4.49) and using the results of Lahiri (1994; 1996), we obtain

$$
\begin{aligned}
&\sup_{B\in\mathcal{B}}\Big|P(T_{2n_0m}\in B) - \int_B (1+p_{n_0m}(x))d\Phi(x)\Big| \le \\
&C\max_{\|\alpha\|\le p+1}\int |D^\alpha(Q_{n_0m}(t) - \Psi_{n_0m}(t))|dt + C\int_{\Gamma_{1m}}|D^\alpha Q_{n_0m}(t)|dt, \\
&= o(m^{-1/2})
\end{aligned} \tag{4.50}
$$

where $C > 0$ is a constant. We have the same first order Edgeworth expansions for T_{2n_0m}, T_{1n_0m}, and T_{n_0m}, namely, three statistics are close to each other. We obtain that

$$
\sup_{B\in\mathcal{B}}\Big|P(T_{n_0m}\in B) \;-\; \int_B (1+p_{n_0m}(x))d\Phi(x)\Big| = o(m^{-1/2}), \tag{4.51}
$$

with $\|p_{n_0m}\phi\|_\infty = O(m^{-1/2})$. The proof is then complete. $\square$

4.4.2 Expansion for bootstrap M-estimator

Define the bootstrap M-estimator $\hat{\beta}^*_{n_0m}$ as a solution of the equation in $t \in \mathbb{R}^p$

$$\sum_{i=1}^{n_0}\Big(\sum_{j=1}^{m} x'_{ij}(\psi(y^*_{ij} - x_{ij}t)) - \hat{\mu}_m\Big) = 0, \tag{4.52}$$

where $\hat{\mu}_m = \frac{1}{b}\mathrm{E}^*_{n_0m}\{\psi(\hat{e}^*_{11}) + \cdots + \psi(\hat{e}^*_{1b})\}$ and y^*_{ij} is given in (4.32). The $\Sigma^*_{n_0m}$ is the conditional covariance matrix of $\sum_{i=1}^{n_0}\sum_{j=1}^{m} d_{ij}\psi(\hat{e}^*_{ij})$ which is given by

$$\sum_{i=1}^{n_0}\sum_{u=1}^{k} Cov_m(\sum_{j=1}^{b} d_{i,(u-1)b+j}\psi(\hat{e}^*_{ij})). \tag{4.53}$$

The natural estimator of $\Sigma^*_{n_0m}$ is

$$\hat{\Sigma}^*_{n_0m} = \sum_{i=1}^{n_0}\sum_{j=0}^{b-1}\sum_{u=1}^{k}\sum_{l=1}^{b-j} D^*_{i,luj}\hat{\sigma}^*_{n_0m}(j), \tag{4.54}$$

where $D^*_{luj} = (1-2^{-1}I(j=0))(\tilde{D}^*_{luj}+\tilde{D}^{*'}_{luj})$, $\tilde{D}^*_{luj} = d_{(u-1)b+l}d'_{(u-1)b+l+j}$. The bootstrap version $T^*_{n_0m}$ of T_{n_0m} is given by

$$T^*_{n_0m} = (\hat{\Sigma}^*_{n_0m})^{-1/2}D^{-1}_{n_0m}(\hat{\beta}^*_{n_0m} - \hat{\beta}_{n_0m}). \tag{4.55}$$

By assumptions, there exists a sequence of statistics $\{\hat{\beta}^*_{n_0m}\}$ such that

$$P(\hat{\beta}^*_{n_0m} \text{ satisfies (4.52) and } \|D^{-1}_{n_0m}(\hat{\beta}^*_{n_0m} - \hat{\beta}_{n_0m})\|^2 \le C\log m) = 1 - o_p(m^{-1/2}). \tag{4.56}$$

Theorem 4.4.2 *Assume that the conditions in Theorem 4.4.1. hold. Suppose that $T^*_{n_0m}$ is defined for some measurable sequence $\{\hat{\beta}^*_{n_0m}\}$ satisfying (4.56) and also suppose that $m^\delta b^{-1} = O(1)$ and $b = O(m^{(1-\kappa)/4})$ for some $\delta > 0$, and $\kappa > \max\{p+3, 5\}\delta_0$. Then*

$$\sup_{B\in\mathcal{B}}\Big|P^*(T^*_{n_0m} \in B) - P(T_{n_0m} \in B)\Big| = o_p(m^{-1/2}) \tag{4.57}$$

for any class $\mathcal{B}$ of Borel subset of $\mathbb{R}^p$.

Proof.

Let $G^*_{n_0m} = D^{-1}_{n_0m}(\hat{\beta}^*_{n0m} - \hat{\beta}_{n_0m})$, $\quad G^*_{1n_0m} = \sum_{i=1}^{n_0}\sum_{u=1}^{k} W^*_{i1u}$,

$A^*_{n_0m} = \sum_{u=1}^{k} W^*_{i2u}$, $\quad \hat{A}_{n_0m} = E_{n_0m}A^*_{n_0m}$, $\tau^*_{1n_0m} = m^{-1}\sum_{i=1}^{n_0}\sum_{j=1}^{m}\psi^{'}(\hat{e}^*_{ij})$,

$\hat{\tau}_{1n_0m} = E^*_{n_0m}(\tau^*_{1n_0m})$, $\xi^*_{uj} = j$th component of B^*_u, for $u = 1, \cdots, k$,

$$W^*_{i1u} = \sum_{j=1}^{b} d_{i((u-1)b+j)}\tilde{\psi}(\xi^*_{uj}), \quad W^*_{i2u} = \sum_{j=1}^{b} d_{i((u-1)b+j)}d^{'}_{i((u-1)b+j)}\psi^{'}(\xi^*_{uj}), \quad 1 \le k \le m.$$

As in the proof of Theorem 4.5.1 and in Lahiri (1996)'s result, we have

$$T^*_{n_0m} = T^*_{1n_0m} + R^{\dagger *}_{n_0m}, \tag{4.58}$$

where $P_{n_0m}(\|R^{\dagger *}_{n_0m}\| > Cv_{2m}) = O_p(v_{3m})$. We also use $T^*_{1n_0m}$ and $T^*_{1n_0m}$ which is the same definition of T_{1n_0m} and T_{2n_0m} in Theorem 4.5.1. The stochastic approximation $T^*_{1n_0m}$ is given by

$$\begin{aligned} T^*_{1n_0m} = \quad & (\hat{\Sigma}^*_{n_0m})^{-1/2}G^*_{1n_0m} + \sum_{r=1}^{p} G^{*'}_{1n_0m}\hat{\Lambda}_{n_0rm}G^*_{1n_0m}q_r \\ & + \sum_{r=1}^{p} \tilde{W}^{*'}_{2n_0m}\hat{\Lambda}_{1n_0rm}G^*_{1n_0m}q_r + \sum_{|\nu|=1}(\hat{\Sigma}^*_{1n_0m})^{\nu}\hat{\Lambda}_{\nu m}G^*_{1n_0m}, \end{aligned} \tag{4.59}$$

where $\tilde{W}^*_{2n_0m} = ((\hat{A}_{n_0m} - A^*_{n_0m})^{'} : (\tau^*_{1n_0m} - \hat{\tau}_{1n_0m}))^{'}$ and $\hat{\Lambda}_{n_0rm}$, $\hat{\Lambda}_{1n_0rm}$, and $\hat{\Lambda}_{\nu m}$ are random matrices which satisfy

$$\begin{aligned} \max \quad & \{\sqrt{m}\|\hat{\Lambda}_{n_0rm} - \Lambda_{1n_0rm}\| + \|\hat{\Lambda}_{1n_0rm} - \Lambda_{1n_0rm}\| \\ & + \|\hat{\Lambda}_{\nu m} - \Lambda_{\nu m}\| : 1 \le r \le p, |\nu| = 1\} = o_p(1). \end{aligned} \tag{4.60}$$

The characteristic function of the Edgeworth expansion for density of $T^*_{n_0m}$, $\Psi^*_{n_0m}(t)$, is shown to satisfy

$$\Psi^*_{n_0m}(t)\exp(\|t\|^2/2) = 1 + \sum_{i=1}^{n_0}\sum_{u=1}^{k} E_m(\sqrt{-1}t^{*'}_m W^*_{i1u})^3/3! + R^*_{\Psi^*}, \tag{4.61}$$

where $t^*_m = t^{'}(\Sigma^*_{n_0 m})^{-1/2}$,

$$\begin{aligned} R^*_{\Psi^*} &= -\sqrt{-1}\sum_{i=1}^{n_0}\sum_{j=1}^{k}\sum_{u=1}^{k} E_m V^*_{ju}(t)\cdot(t^*_m Z^*_{1j})\cdot(t^{*'}_m Z^*_{1u}) \\ &\quad +\sqrt{-1}\sum_{i=1}^{n_0}\sum_{u=1}^{k} E_m V^*_{uu}(t), \end{aligned} \tag{4.62}$$

, and

$$\begin{aligned} V^*_{jk} &= \sum_{r=1}^{p}(t^{'}q_r)[W^*_{1j}\hat{\Lambda}_{n_0 rm}W^*_{1k} + \hat{W}^*_{2m,j}\hat{\Lambda}_{1n_0 rm}W^*_{1k}] - W^*_{1j}\bar{L}^*_{1m}(t)W^*_{1k} \\ &\quad +k^{-1}\sum_{|\nu|=1}[\sum_{j=0}^{b}(L^*_{jm})^{\nu}(W_{4k}(j) - E_m W_{4k}(j))](t^{'}\hat{\Lambda}_{\nu m})W^*_{1k}, \end{aligned} \tag{4.63}$$

with $\bar{L}^*_{1m}(t) = \sum_{j=0}^{b}\sum_{|\nu|=1}(L^*_{jm})^{\nu}\gamma^*_{jn}(t^{'}\Lambda_{\nu m})$. Now, using the results of Bhattachararya and Ranga Rao (1986) and Lahiri (1996), we have

$$\int_{\Gamma_{2m}} |D^{\alpha}(Q^*_{n_0 m}(t) - \Psi^*_{n_0 m}(t))|dt = O_p(v_{3m}), \tag{4.64}$$

where $\Gamma_{2m} = \{t : \|t\| < mb^{-1}(\log m)^{-10}\}$ and $Q^*_{n_0 m}(t) = E^*_{n_0 m}\exp(\sqrt{-1}t^{'}T^*_{n_0 2m})$. Finally, using Lahiri (1994)'s results, it is shown that

$$\max_{|\alpha|\le p+1}\int_{\Gamma_{3m}} |D^{\alpha}(Q^*_{n_0 m}(t))| = O_p(v_{3m}), \tag{4.65}$$

where $\Gamma_{3m} = \Gamma_m/\Gamma_{2m}$. Similar to (4.50), we have

$$\begin{aligned} &\sup_{B\in\mathcal{B}}\left|P^*(T^*_{2n_0 m}\in B) - \int_B (1 + p^*_{n_0 m}(x))d\Phi^*(x)\right| \le \\ &C\max_{\|\alpha\|\le p+1}\int |D^{\alpha}(Q^*_{n_0 m}(t) - \Psi^*_{n_0 m}(t))|dt + C\int |D^{\alpha}Q^*_{n_0 m}(t)|dt, \end{aligned} \tag{4.66}$$

where $C > 0$ is a constant and $p^*_{n_0 m}$ is obtained from $p_{n_0 m}$ on replacing population moments by sample moments in coefficients. If we have $b = m^{1/4}$, we obtain

$$\sup_{B\in\mathcal{B}}\left|P^*(T^*_{2n_0 m}\in B) - \int_B (1 + p^*_{n_0 m}(x))\phi(x)dx\right| = o_p(v_m),\ x\in\mathbb{R}^p. \tag{4.67}$$

We have the empirical Edgeworth expansion (Bhattacharya and Qumsiyeh (1988), Lahiri (1994)) for T_{2n_0m}

$$\sup_{B\in\mathcal{B}}\left|P(T_{2n_0m}\in B) \quad - \quad \int_B(1+p^*_{n_0m}(x))\phi(x)dx\right| = o_p(v_m). \tag{4.68}$$

From (4.64)-(4.68), assuming $p^*_{n_0m}-p_{n_0m}=o(v_m)$, we have

$$\sup_{B\in\mathcal{B}}\left|P^*(T^*_{2n_0m}\in B)-P(T_{2n_0m}\in B)\right| = o_p(v_m). \tag{4.69}$$

We have the same first order Edgeworth expansion forms for $T^*_{2n_0m}$, $T^*_{1n_0m}$, and $T^*_{n_0m}$, since those are close to each other. We obtain

$$\sup_{B\in\mathcal{B}}\left|P^*(T^*_{n_0m}\in B)-P(T_{n_0m}\in B)\right| = o_p(v_m). \tag{4.70}$$

□

Theorem 4.4.2 shows that the MBB indeed provides more accurate approximation for studentized multivariate M-estimator of the regression parameter vector β than normal approximation. Consequently, Theorem 4.4.2 is useful for constructing second-order correct multivariate inference procedures for β under multiple regression model. The studentized moving block bootstrap statistics obtain the second order accuracy for the bounded $n=n_0$ case.

4.5 Simulation Work

The block bootstrap captures the dependence in the series of residuals without the need to know the correlation structure. It can be simple and account for correlations in a regression model with correlated error.

To define the bootstrap version of $\hat{\beta}_{n_0m}$, first form the observed blocks of residual length b as $\xi_{ih}=(\hat{e}_{ij},\cdots,\hat{e}_{i(h+b-1)}), 1\leq h\leq q$, where $q=m-b+1$ and

$\hat{e}_{ij} = y_{ij} - x_{ij}\hat{\beta}_{n_0m}, 1 \le j \le m, 1 \le i \le n$. Next draw $\xi^*_{i1}, \cdots, \xi^*_{ik}$ randomly, with replacements from $\xi_{i1}, \cdots, \xi_{iq}$, where $m/b = k$ is assumed to be an integer for simplicity. Note that each ξ^*_{ik} has b components. Denote the lth component of $\xi^*_{ik}, 1 \le l \le b$ by ξ^*_{ikl}. Also, set $\hat{e}^*_{i((b-1)k+l)} = \xi^*_{ikl}, 1 \le l \le b$, and we have the bootstrap pseudo-observations

$$y^*_{ij} = x_{ij}\hat{\beta}_{nm} + \hat{e}^*_{ij}, \quad 1 \le i \le n_0, \quad 1 \le j \le m. \tag{4.71}$$

Adapting Shorack's approach, we obtain the bootstrapped estimator $\hat{\beta}^*_{n_0m}$ as a solution of the equation $t \in R^p$,

$$g'_{n_0m} = \sum_{i=1}^{n_0}\sum_{j=1}^{m} x'_{ij}\left((y^*_{ij} - x_{ij}\beta) - \hat{\mu}_{nm}\right) = 0, \tag{4.72}$$

where $\hat{\mu}_{n_0m} = b^{-1}E_{n_0m}\{e^*_{11}+\cdots+e^*_{n_0b}\}$, and E_{n_0m}denotes the conditional expectations under the MBB resampling scheme, given $\hat{e}_{11}, \cdots, \hat{e}_{nm}$. Centering the above equation by $\hat{\mu}_{n_0m}$ makes the estimating equation conditionally unbiased at $\beta = \hat{\beta}_{n_0m}$ and ensures the bootstrap analog. The bootstrap estimator is as follows:

$$\hat{\beta}^*_{n_0m} = (\sum_{i=1}^{n_0}\sum_{j=1}^{m} x'_{ij}x_{ij})^{-1}\sum_{i=1}^{n_0}\sum_{j=1}^{m} x'_{ij}(y^*_{ij} - \hat{\mu}_{n_0m}). \tag{4.73}$$

Consider the following specific model in simulation work:

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 x_{ij} + \gamma_i + e_{ij},\ i = 1, \cdots, n_0\ , j = 1, \cdots, m, \\ & e_{ij} = \phi e_{i(j-1)} + u_{ij}, \\ & \gamma_i \sim N(0, \sigma^2_\gamma),\ \text{and}\ \ u_{ij} \sim N(0, \sigma^2_u). \end{aligned}$$

In particular, let $n_0 = 5$, $m = 20$, and $x_{ij} = (1, \cdots, 20)'$.

$$\beta_0 = 10, \beta_1 = 1, b = 4, k = 5,$$

$$\sigma^2_\gamma = 1, \sigma^2_u = 1,\ \text{and}\ \ \phi = 0.75.$$

4.5.1 Bootstrapping the distribution of statistics

Let R be the number of bootstrap simulations $(r = 1, \cdots, R)$, and $\hat{\beta}^*$ be the bootstrap estimate of β for the r samples. The important result is that the distribution of $\hat{\beta}^*$, estimated by the empirical distribution function of the $\hat{\beta}^{*,r}$, $(r = 1, \cdots, R)$, approximates the distribution of $\hat{\beta}$. Now we define the studentized statistics as follows

$$\hat{T}^* = \frac{\hat{\beta}^* - \hat{\beta}}{\hat{S}^*_{\hat{\beta}^*}}. \tag{4.74}$$

The difference between the distribution functions of $\hat{T}$ and $\hat{T}^*$ tends to 0, when the number of observations is large; thus we can use the quartiles of $\hat{T}^*$ instead of $\hat{T}$ to construct intervals or tests. Let $(\hat{T}^{*,r}, r = 1, \cdots, R)$ be the r-th sample of $\hat{T}^*$, where $\hat{T}^*$ is calculated in the same way as $\hat{T}$, replacing y_{ij} with y^*_{ij}. Let $\hat{q}_\alpha$ be the percentile of the $\hat{T}^{*,r}$. It can be shown that $P(\hat{T} \leq \hat{q}_\alpha)$ tends to α, when m tends to infinity. This gives a bootstrap confidence interval for β

$$\hat{I}_R = \left[\hat{\beta}^* - \hat{q}_{1-\frac{\alpha}{2}} \hat{S}^*_{\hat{\beta}^*} \ \ , \ \ \hat{\beta}^* - \hat{q}_{\frac{\alpha}{2}} \hat{S}^*_{\hat{\beta}^*}\right]. \tag{4.75}$$

For large m and R, the coverage probability of $\hat{I}_R$ is close to $1 - \alpha$. The bootstrap estimation of the variance is calculated using the empirical variance of the R sample $(\hat{\beta}^{*,r}, r = 1, \cdots, R)$:

$$\hat{S}^{*2} = \frac{1}{R-1} \sum_{r=1}^{R} (\hat{\beta}^{*,r} - \overline{\hat{\beta}^*})^2, \tag{4.76}$$

where $\overline{\hat{\beta}^*}$ is the sample mean $\overline{\hat{\beta}^*} = \sum_{r=1}^{R} \hat{\beta}^{*,r}/R$.

Coverage accuracy, where coverage is the probability that a confidence interval includes β, is the important property for a confidence interval procedure. Bootstrap confidence interval methods differ in their asymptotic properties. Our simulation results are given in Table 1. MBB1 and MBB2 are similar to each other. Those two block bootstrap methods obtained correct coverage probability at the nominal level

Table 1: Coverage probability and length of CI ($\hat{\beta}_1^*$): 500 replications; $\phi = 0.75$, SOB is stratified ordinary bootstrap, SB is a standard bootstrap estimation, and $\hat{\beta}_1$ is a robust estimation with unknown covariance structure.

Methods	CI($\hat{\beta}_1^*$)	Probability	Length
MBB1	(0.849,1.168)	0.949	0.318
MBB2	(0.844,1.163)	0.952	0.320
SOB	(0.805,1.091)	0.768	0.286
SB	(0.804,1.086)	0.747	0.282
$\hat{\beta}_1$	(0.840,1.169)	0.950	0.329

of 95%. The standard bootstrap or stratified ordinary bootstrap did not perform well in highly correlated longitudinal data with low coverage probabilities.

CHAPTER V

DIAGONAL OPTIMAL WEIGHT FOR UNBALANCED LONGITUDINAL DATA

5.1 Introduction

In this chapter, we investigate four different kinds of weight schemes in an unbalanced longitudinal data. We focus on the longitudinal design in which the number of subjects is large and the number of replications is small. The observations for each subject would be take at unequal points in time. One weight scheme corresponds to equal weight for subjects, and the other weight scheme corresponds to equal weight for observations. We introduce the diagonal optimal weight in GEE with working independent correlation matrix in minimizing the variance of the regression parameter over all choices.

5.2 Diagonal Optimal Weight

In this section, we can see the optimal weight scheme in a working independent setting in a case when $max(m_i)$ is bounded and n is large. The model is

$$y_{ij} = \beta_0 + x_{ij}\beta_1 + e_{ij}, \quad i = 1, \cdots, n, j = 1, \cdots, m_i. \tag{5.1}$$

Let $V_i = \text{cov}(e_i)/\sigma^2$, $V = \text{diag}(V_i)$ and $N = \sum_{i=1}^{n} m_i$ with $e_i = (e_{i1}, \cdots, e_{im_i})'$. Suppose we wish to estimate $Q'\beta$, the estimator $\hat{\beta} = (\sum_{i=1}^{n} x_i' V_i^{-1} x_i)^{-1} \sum_{i=1}^{n} x_i' V_i^{-1} y_i$ is the best linear unbiased estimator (BLUE) of β. The conventional GEE is given by

$$g_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} x_{ij}' w_i^{-1} (y_{ij} - x_{ij}\beta) = 0, \tag{5.2}$$

where w_i is a weight for subject i. We represent the above equation

$$g_n(\beta) = X^{'}W(Y - X\beta), \tag{5.3}$$

where $W = diag(I_{m_i}w_i)$. Let $\tilde{\beta} = (X^{'}WX)^{-1}X^{'}WY$ then $Q^{'}\tilde{\beta}$ is unbiased for $Q^{'}\beta$, but it is not a BLUE, unless $W = V^{-1}$. We can make a choice to weight the estimating function. For example, 1) $w_i = \frac{1}{N}$ and 2) $w_i = \frac{1}{nm_i}$ (Huang et al., 2002). We suggest using the optimal weight w_i in working independence setting for minimizing var$(Q^{'}\tilde{\beta})$ over all choices of w_i. The variance of var$(Q^{'}\tilde{\beta})$ is written as var$(Q^{'}\tilde{\beta}) = Q^{'}A^{-1}BA^{-1}Q$. After we apply the Lagrange multiple technique, we get the optimal weight form with a condition $\sum_{i=1}^{n} m_i w_i = N$

$$w_i = \left(L^{'}x_i{'}x_i A^{-1}BL - \frac{1}{2}\lambda m_i\right) / \left(L^{'}x_i^{'}V_i x_i L\right) \tag{5.4}$$

$$\lambda = 2\frac{\left[\sum_{k=1}^{n}(m_k L^{'}x_k^{'}x_k A^{-1}BL/L^{'}x_k^{'}V_k x_k L) - N\right]}{\sum_{k=1}^{n}\left(m_k^2/L^{'}x_k^{'}V_k x_k L\right)}, \tag{5.5}$$

where $L^{'} = Q^{'}(\sum_{i=1}^{n} w_i x_i^{'}x_i)^{-1}$, $A = \sum_{i=1}^{n} w_i x_i^{'}x_i$, and $B = \sum_{i=1}^{n} w_i^2 x_i^{'}V_i x_i$. We define $w_i^{(1)} = \frac{1}{N}$, $w_i^{(2)} = \frac{1}{nm_i}$, $w_i^{(3)} = D$-optimal, and $W_i^{(4)} = V_i^{-1}$ (true optimal), which is a non-diagonal $m_i \times m_i$ matrix. We show that for only intercept term regression, the diagonal optimal weight $w_i^{(3)}$ and the true optimal weight $W_i^{(4)}$ have exactly the same optimality.

Lemma 5.2.1 *For only an intercept term, the diagonal optimal weight* $w_i^{(3)}$ *and the true optimal weight* $W_i^{(4)}$ *give the same optimality.*

Proof.

$$x_i^{'}V_i^{-1}x_i = x_i^{'}W_i x_i \tag{5.6}$$

For general m_i, we have

$$x_i^{'}V_i^{-1}x_i = \mathbf{1}^{'}\frac{1}{\sigma_\varepsilon^2}\left(I_{m_i} - \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2 + m_i\sigma_\gamma^2}J_{m_i}\right)\mathbf{1} \tag{5.7}$$

$$= \frac{1}{\sigma_\varepsilon^2}\mathbf{1}^{'} \begin{pmatrix} 1-\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & \cdots & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} \\ -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & 1-\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & \cdots & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} \\ -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & 1-\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & \cdots & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & -\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} & \cdots & 1-\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2} \end{pmatrix} \mathbf{1}$$

$$= \mathbf{1}^{'} \begin{pmatrix} \frac{1}{\sigma_\varepsilon^2}(1-\frac{m_i\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2}) & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_\varepsilon^2}(1-\frac{m_i\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2}) & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_\varepsilon^2}(1-\frac{m_i\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_\varepsilon^2}(1-\frac{m_i\sigma_\gamma^2}{\sigma_\varepsilon^2+m_i\sigma_\gamma^2}) \end{pmatrix} \mathbf{1}$$

$$= x_i^{'} W_i x_i,$$

where

$$w_i = \frac{1}{\sigma_\varepsilon^2}\left(1 - \frac{m_i\sigma_\gamma^2}{\sigma_\varepsilon^2 + m_i\sigma_\gamma^2}\right), \tag{5.8}$$

which leads to the optimal estimating equation, completing the proof.

5.3 Simulation Results

We generate the simplest case example. Assume that the total number of subjects is 40 and the total number of observations is 160. The first group consists of 10 replications for 10 subjects, and the second group is composed of 2 replications for 30 subjects. The correlation structure is compound symmetric. The fixed covariates are intercept and time effect, and one random intercept model is considered. The notation is total data $N = \sum_{i=1}^{40} m_i = 160$ and subject $n = 40$. The first group has $m_i = 10$ replications and the second group has $m_i = 2$.

As we can see in the Lemma (5.2.1), the diagonal optimal weight $w_i^{(3)}$ is the same as the true optimal weight in case of a single intercept term with known variance, as Table 2 shows. Table 3 presents the results of a slope estimator with known variance. Table 4 includes the results of a slope estimator with unknown variance and known structure using REML. In Table 5, we see the results of a slope with unknown variance and unknown structure (misspecified variance) using REML. Table 6 presents the diagonal optimal weights for two unbalanced groups for the linear combination of regression parameters. Table 7 shows the results for the linear combination of regression parameters with known variance. In Table 8, we investigate that the performance of the diagonal wight with unknown variance and known structure. Table 9 is the results in the case of unknown variance and unknown structure (misspecified variance). In summary, we conclude that for only intercept term the diagonal optimal weight is the exactly the same as the true optimal and for intercept and slope regression estimators the diagonal optimal weight performs better rather than the observational weight and individual weight. We also observe that when the correlation is high the diagonal optimal weight reaches the individual weight, and when the data has a low correlation the diagonal optimal weight arrives at the observational weight.

Table 2: Empirical bias, Empirical standard error and Average standard error for an intercept estimator $\hat{\beta}_0$ with known V_i: 500 replications.

ρ	Methods	$EB(\hat{\beta}_0)$	$ESE(\hat{\beta}_0)$	$Avg(SE(\hat{\beta}_0))$
0.909	$w_i^{(1)}$	0.041	0.498	0.471
	$w_i^{(2)}$	0.059	0.388	0.361
	$w_i^{(3)} = (0.412, 1.980)$	0.015	0.366	0.361
	$W^{(4)}$	0.015	0.366	0.361
0.667	$w_i^{(1)}$	-0.021	0.214	0.217
	$w_i^{(2)}$	0.022	0.172	0.173
	$w_i^{(3)} = (0.459, 1.929)$	-0.016	0.164	0.173
	$W^{(4)}$	-0.016	0.164	0.173
0.333	$w_i^{(1)}$	-0.008	0.178	0.168
	$w_i^{(2)}$	0.030	0.165	0.150
	$w_i^{(3)} = (0.575, 1.727)$	-0.008	0.155	0.146
	$W^{(4)}$	-0.008	0.155	0.146
0.010	$w_i^{(1)}$	0.002	0.085	0.100
	$w_i^{(2)}$	0.036	0.106	0.123
	$w_i^{(3)} = (0.973, 1.049)$	0.002	0.085	0.100
	$W^{(4)}$	0.002	0.085	0.100

Table 3: Empirical bias, Empirical standard error and Average standard error for a slope estimator $\hat{\beta}_1$ with known V_i: 500 replications.

ρ	Methods	$EB(\hat{\beta}_1)$	$ESE(\hat{\beta}_1)$	$Avg(SE(\hat{\beta}_1))$
0.909	$w_i^{(1)}$	0.000	0.087	0.087
	$w_i^{(2)}$	0.197	0.139	0.118
	$w_i^{(3)} = (1.485, 0.192)$	0.000	0.033	0.033
	$W^{(4)}$	0.001	0.024	0.024
0.667	$w_i^{(1)}$	-0.001	0.042	0.042
	$w_i^{(2)}$	0.197	0.064	0.056
	$w_i^{(3)} = (1.584, 0.027)$	-0.001	0.025	0.024
	$W^{(4)}$	-0.001	0.024	0.024
0.333	$w_i^{(1)}$	-0.002	0.039	0.038
	$w_i^{(2)}$	0.196	0.056	0.047
	$w_i^{(3)} = (1.426, 0.291)$	-0.001	0.031	0.032
	$W^{(4)}$	-0.001	0.031	0.031
0.010	$w_i^{(1)}$	-0.001	0.027	0.027
	$w_i^{(2)}$	0.197	0.033	0.029
	$w_i^{(3)} = (1.027, 0.958)$	-0.001	0.027	0.027
	$W^{(4)}$	-0.001	0.027	0.027

Table 4: Empirical bias, Empirical standard error and Average standard error for a slope estimator $\hat{\beta}_1$ with unknown V_i, we assume the correlation structure is known and use the restricted maximum likelihood estimator (REML) $\hat{V}_i$: 500 replications.

$Avg(\hat{\rho})$	Methods	$EB(\hat{\beta}_1)$	$ESE(\hat{\beta}_1)$	$Avg(SE(\hat{\beta}_1))$
0.905	$w_i^{(1)}$	0.000	0.093	0.087
	$w_i^{(2)}$	0.197	0.146	0.118
	$w_i^{(3)}$	0.000	0.039	0.033
	$W^{(4)}$	0.000	0.026	0.024
0.658	$w_i^{(1)}$	0.000	0.042	0.042
	$w_i^{(2)}$	0.200	0.068	0.056
	$w_i^{(3)}$	0.002	0.025	0.024
	$W^{(4)}$	0.001	0.024	0.024
0.329	$w_i^{(1)}$	0.000	0.037	0.038
	$w_i^{(2)}$	0.198	0.051	0.047
	$w_i^{(3)}$	0.001	0.033	0.033
	$W^{(4)}$	0.001	0.033	0.031
0.027	$w_i^{(1)}$	0.001	0.026	0.027
	$w_i^{(2)}$	0.200	0.030	0.030
	$w_i^{(3)}$	0.002	0.026	0.027
	$W^{(4)}$	0.002	0.026	0.027

Table 5: Empirical bias, Empirical standard error and Average standard error for a slope estimator $\hat{\beta}_1$ with unknown V_i, we assume the correlation structure is unknown (Misspecified) and use the restricted maximum likelihood estimator (REML) $\hat{V}_i$: 500 replications.

$Avg(\hat{\rho})$	Methods	$EB(\hat{\beta}_1)$	$ESE(\hat{\beta}_1)$	$Avg(SE(\hat{\beta}_1))$
0.906	$w_i^{(1)}$	-0.001	0.088	0.096
	$w_i^{(2)}$	0.198	0.142	0.115
	$w_i^{(3)}$	-0.001	0.058	0.086
	$W^{(4)}$	0.000	0.049	0.027
0.658	$w_i^{(1)}$	0.000	0.042	0.048
	$w_i^{(2)}$	0.198	0.067	0.051
	$w_i^{(3)}$	0.000	0.041	0.048
	$W^{(4)}$	0.000	0.039	0.045
0.326	$w_i^{(1)}$	0.000	0.038	0.040
	$w_i^{(2)}$	0.198	0.056	0.042
	$w_i^{(3)}$	0.000	0.038	0.040
	$W^{(4)}$	0.000	0.038	0.039
0.026	$w_i^{(1)}$	0.001	0.026	0.027
	$w_i^{(2)}$	0.200	0.032	0.029
	$w_i^{(3)}$	0.000	0.026	0.027
	$W^{(4)}$	0.001	0.026	0.027

Table 6: Differences in weights for $\hat{\beta}_0$ and $\hat{\beta}_1$ in two group cases and for $\hat{\beta}_0 + \hat{\beta}_1 x_q$, where $x_q =$ is q-th percentile. Diagonal optimal weights for two unbalanced groups for the linear combination $Q'\beta$ of regression quantiles

ρ	$Q^{'}$	x_q	w_{i1}	w_{i2}
0.909	$Q^{'} = (1, 1)$	10-th	0.902	1.163
	$Q^{'} = (1, 1.5)$	25-th	0.792	1.347
	$Q^{'} = (1, 2.5)$	50-th	0.427	1.955
	$Q^{'} = (1, 6.5)$	75-th	0.048	2.587
0.667	$Q^{'} = (1, 1)$	10-th	0.915	1.142
	$Q^{'} = (1, 1.5)$	25-th	0.820	1.300
	$Q^{'} = (1, 2.5)$	50-th	0.514	1.810
	$Q^{'} = (1, 6.5)$	75-th	0.168	2.387
0.333	$Q^{'} = (1, 1)$	10-th	0.942	1.097
	$Q^{'} = (1, 1.5)$	25-th	0.880	1.200
	$Q^{'} = (1, 2.5)$	50-th	0.694	1.511
	$Q^{'} = (1, 6.5)$	75-th	0.364	2.059
0.010	$Q^{'} = (1, 1)$	10-th	0.997	1.004
	$Q^{'} = (1, 1.5)$	25-th	0.995	1.009
	$Q^{'} = (1, 2.5)$	50-th	0.988	1.020
	$Q^{'} = (1, 6.5)$	75-th	0.911	1.148

Table 7: Empirical bias, Empirical standard error and Average standard error for the linear combination of regression parameters with known V_i: 500 replications.

ρ	Methods	$EB(\hat{\beta}_0 + 2.5\hat{\beta}_1)$	$ESE(\hat{\beta}_0 + 2.5\hat{\beta}_1)$	$Avg(SE(\hat{\beta}_0 + 2.5\hat{\beta}_1))$
0.909	$w_i^{(1)}$	-0.009	0.371	0.401
	$w_i^{(2)}$	0.234	0.365	0.361
	$w_i^{(3)}$	-0.013	0.342	0.361
	$W^{(4)}$	-0.013	0.342	0.361
0.667	$w_i^{(1)}$	-0.007	0.189	0.188
	$w_i^{(2)}$	0.238	0.182	0.173
	$w_i^{(3)}$	-0.006	0.172	0.173
	$W^{(4)}$	-0.006	0.172	0.173
0.333	$w_i^{(1)}$	-0.013	0.155	0.153
	$w_i^{(2)}$	0.236	0.159	0.150
	$w_i^{(3)}$	-0.010	0.150	0.147
	$W^{(4)}$	-0.010	0.150	0.147
0.010	$w_i^{(1)}$	0.001	0.087	0.090
	$w_i^{(2)}$	0.242	0.102	0.101
	$w_i^{(3)}$	0.001	0.087	0.090
	$W^{(4)}$	0.001	0.087	0.090

Table 8: Empirical bias, Empirical standard error and Average standard error for the linear combination of regression parameters with unknown V_i, we assume the correlation structure is known and use the restricted maximum likelihood estimator (REML) $\hat{V}_i$: 500 replications.

$Avg(\hat{\rho})$	Methods	$EB(\hat{\beta}_0 + 2.5\hat{\beta}_1)$	$ESE(\hat{\beta}_0 + 2.5\hat{\beta}_1)$	$Avg(SE(\hat{\beta}_0 + 2.5\hat{\beta}_1))$
0.904	$w_i^{(1)}$	0.007	0.419	0.397
	$w_i^{(2)}$	0.249	0.402	0.357
	$w_i^{(3)}$	0.007	0.376	0.357
	$W^{(4)}$	0.007	0.376	0.357
0.662	$w_i^{(1)}$	0.001	0.187	0.187
	$w_i^{(2)}$	0.243	0.185	0.173
	$w_i^{(3)}$	0.001	0.174	0.172
	$W^{(4)}$	0.001	0.174	0.172
0.329	$w_i^{(1)}$	0.007	0.153	0.153
	$w_i^{(2)}$	0.250	0.158	0.150
	$w_i^{(3)}$	0.007	0.149	0.146
	$W^{(4)}$	0.007	0.149	0.146
0.025	$w_i^{(1)}$	0.000	0.088	0.092
	$w_i^{(2)}$	0.244	0.105	0.102
	$w_i^{(3)}$	0.000	0.089	0.092
	$W^{(4)}$	0.000	0.089	0.092

Table 9: Empirical bias, Empirical standard error and Average standard error for the linear combination of regression parameters with unknown V_i, we assume the correlation structure is unknown (Misspecified) and use the restricted maximum likelihood estimator (REML) $\hat{V}_i$: 500 replications.

$Avg(\hat{\rho})$	Methods	$EB(\hat{\beta}_0 + 2.5\hat{\beta}_1)$	$ESE(\hat{\beta}_0 + 2.5\hat{\beta}_1)$	$Avg(SE(\hat{\beta}_0 + 2.5\hat{\beta}_1))$
0.906	$w_i^{(1)}$	0.014	0.392	0.382
	$w_i^{(2)}$	0.252	0.373	0.352
	$w_i^{(3)}$	0.008	0.352	0.350
	$W^{(4)}$	0.007	0.354	0.347
0.657	$w_i^{(1)}$	-0.004	0.191	0.165
	$w_i^{(2)}$	0.238	0.184	0.164
	$w_i^{(3)}$	-0.005	0.176	0.159
	$W^{(4)}$	-0.005	0.176	0.157
0.326	$w_i^{(1)}$	0.004	0.150	0.133
	$w_i^{(2)}$	0.748	0.152	0.143
	$w_i^{(3)}$	0.004	0.146	0.132
	$W^{(4)}$	0.004	0.146	0.131
0.024	$w_i^{(1)}$	0.001	0.091	0.090
	$w_i^{(2)}$	0.275	0.104	0.101
	$w_i^{(3)}$	0.001	0.091	0.090
	$W^{(4)}$	0.001	0.091	0.090

CHAPTER VI

ANALYZING NURSING HOME RESIDENTS WITH MULTIPLE SCLEROSIS USING MINIMUM DATA SET

6.1 Introduction

In this chapter, we consider the model selection/comparison for repeated measurement data. The appropriate covariance structure for the response vector of repeated measurement for each subject need to be specified by using the information criterions. We analyze the information on the nursing home residents with multiple sclerosis in a minimum data set.

6.2 Minimum Data Set with Multiple Sclerosis

The Minimum Data Set (MDS) is a federally-mandated assessment instrument that includes all nursing home residents (regardless of payment source) in all Medicare- and Medicaid-certified nursing facilities. Trained clinical professionals (such as nurses, social workers, or therapists) assess residents by direct observation over all shifts prior to the MDS assessment. Each MDS item is defined, with guidance on how to ask questions, what to observe, and whom to contact for information. Each resident's preadmission, admission, or transfer notes are reviewed, as well as the current plan of care and recent physician notes or orders for the resident's immediate care.

MDS assessments are required for each resident at admission, upon significant changes in status, and at least annually. In addition, residents are assessed quarterly on a subset of MDS items. The MDS contains comprehensive assessments of nursing home residents, including gender, birth, date, marital status, race/ethnicity, place of residence, and payment source. In addition, the data set include information behavior, psychological well being, cognitive patterns, ability to communicate, a range of physical functioning variables, disease and infections, medications, and treatments. Federal requirements for the status and nursing homes to encode and transmit the MDS began on June 22, 1998.

Multiple sclerosis (MS) is the most common neurologic disease among younger adults, with as many as 350,000 Americans diagnosed with this disease by a physician. MS is a demyelinating disease of the central nervous system that may lead to the manifestation of a range of symptoms, including spasticity, movement disorders, fatigue, bladder and bowel dysfunctions, pain, depression, visual disorders, numbness, cognitive difficulties, speech disorder, and dysphagia. The clinical course of MS usually follows a variable pattern over time, but typically is characterized by either episodic acute periods of worsening condition (relapses, exacerbations, bouts, attacks), gradual progressive deterioration of neurologic function, or combinations both. MS is characterized by episodes of neurological symptoms that are often followed by fixed neurologic deficits, increasing disability, and medical, socioeconomic, and physical decline over 30-40 years. Females are about twice as likely as males to be diagnosed with multiple sclerosis. Females also tend to develop symptoms of MS at an earlier age than males, while males tend to have a more progressive and severe from of MS.

6.3 Statistical Analysis for MS Residents

6.3.1 Porell's model

Porell et al. (1988) used Quarterly Management Minutes Questionnaire survey data for Medicaid case-mix reimbursement of nursing homes in Massachusetts from 1991 to 1994 for specification of outcomes and resident attributes. The state-dependence regression models are considered for the activities of daily living (ADL) functional status, incontinence status, and mental status outcomes from longitudinal residence histories of Medicaid residents spanning 3 to 36 months in length. Outcomes are specified to be a function of resident demographic and diagnostic attributes and facility-level operating and nurse staffing attributes. They concluded that the absence of uniform associations between facility attributes and the various long-term care health outcomes studied suggests that strong facility performance on one health outcome may coexist with much weaker performance on each outcome, and this has implications for the aggregation of individual facility performance measure on multiple outcomes and the development of overall outcome performance measures.

We employ the Porell et al. (1988) model for our MS longitudinal analysis. The state-dependence specification for the longitudinal modeling of functional and health outcomes is appropriate for two reasons. First, nursing home residents of long-term age are inclined to display modest but often irreversible deterioration in functional status over the long run, making current functional status a good predictor of subsequent functional status. Second, better adjustments should truly result from using a resident's own experience through a lagged outcome measure, rather than from generic demographic or diagnostic variables alone. State-dependence, through the specification of the lagged ADL long scale, accounts for much of the high explanatory power of the model.

Given the MDS format in our analysis, part of the nontrivial work is data management. To analyze the data, we sort them by residents, and exclude the residents who have a weight above 500 pounds or below 30 pounds and a height above 100 inches or below 30 inches. After the data is cleaned up, there are 12,858 residents and a total of 51,505 observations with a diagnosis of MS recorded in the MDS between June 23, 1998 and December 31, 2000. The response is the ADL long scale (0-28 scores) and the average is 20.28. The mean of the time difference between the dates is 170.8 days. Following Porell et al. (1988) study of nursing home outcomes, we apply a multiple state-dependence model:

$$y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 \text{White}_t + \beta_3 \text{Male}_t + \beta_4 \text{Age}_t + \beta_5 \text{BMI}_t + \beta_6 \text{Cog}_t + \beta_7 \text{Disease}_t + \varepsilon_t, \tag{6.1}$$

where:

$$\begin{aligned}
y_{t+1} &= \text{ADL long score at time } \quad t+1 \quad \text{and} \\
y_t &= \text{Lagged ADL long score at time } \quad t; \\
\text{White} &= \text{White residents and other;} \\
\text{Male} &= \text{Male and Female;} \\
\text{Age} &= \text{Ages in years;} \\
\text{BMI} &= \text{Body Mass Index;} \\
\text{Cog} &= \text{Cognitive scale;} \\
\text{Disease} &= \text{Resident risk factor (0 or 1);} \\
\varepsilon_t &= \text{A random disturbance term.}
\end{aligned}$$

The R^2 and adjusted R^2 are both about 0.86. The overall fit reflected seems very effective. Table 10 shows that the past ADL long score is highly significant, which implies that the past ADL long score is a good projection of the future ADL long score. The estimates of age, BMI, and cogscale are significant. Some disease factors are significant. Among them are congestive heart failure, deep vein thrombosis, hip fracture, missing limb, osteoporosis, paraplegia, quadriplegia, manic depressive, diabetic retinopathy, glaucoma. Many other diseases are not significant. Among infections, only antibiotic resistant infections are significant. The White and Male variables are not significant.

Table 10: Model fitting estimates of fixed effects parameters for Porell's model fitted to the MDS with MS patient data.

Effects	Estimates	Standard Error	z-value	p-value
Intercept	4.244	0.149	28.58	$<$ .0001
Long	0.823	0.003	280.30	$<$ .0001
White	-0.037	0.063	-0.59	0.5571
Male	0.006	0.043	0.15	0.8837
Age	-0.008	0.001	-5.39	$<$ .0001
BMI	-0.007	0.003	-2.35	0.0186
Cogscale	0.098	0.012	8.34	$<$ .0001

6.3.2 Longitudinal data analysis for MS patients

In the MDS data for nursing home residents with MS during 1999-2000, the number of observations was 51,969, the number of subjects was 12961, the maximum number of observations per subject was 19, and the minimum number of observation per subject was 2. We do not include 24 people who received 464 observations because their weight was below 30 pounds or above 500 pounds and their height was less than 30 inches. The resulting data set is 51,505 observations and 12,937 subjects. The linear model assume that the relationship between the mean of the dependent variable y and the fixed and random effects can be modeled as a linear function, and the random effect follows a normal distribution:

$$\begin{aligned} y_{ij} = \quad & \alpha + \beta_1 \mathrm{Age}_i + \beta_2 \mathrm{Time}_{ij} + \beta_3 \mathrm{BMI}_{ij} + \beta_4 \mathrm{Cogscale}_{ij} + u_i + \varepsilon_{ij}, \\ & i = 1, \ldots, n, \quad j = 1, \ldots, m_i, \end{aligned} \tag{6.2}$$

where $u_i \sim N(0, \sigma_u^2)$, a random individual effect, and $\varepsilon_{ij} \sim^{iid} N(0, R_i)$, a pure error term. The response variable y_{ij} is ADL long scale, observed at time t_{ij}. The age variable is defined as $Age = (ab1 - aa3)/365.25$, where $ab1$ is a date of entry and $aa3$ is a birthdate. The time variable stands for the resident's admission time that is given by Time $= R2b_{\mathrm{start}} - R2b$, where $R2b$ is a date RN assessment coordinator signed as complete, and cogscale denotes a cognitive performance scale measurement. The unit of the time variable is a day, and the unit of age is a year. The BMI formula is

$$\text{BMI} = \frac{\text{Weight [in pounds]} \times 704.5}{\big(\text{Height [in inches]}\big)^2}. \tag{6.3}$$

In attempting to choose the best covariance structure, the likelihood test can be used when two models with the same fixed-effects parameters are fit to the data using ML estimation and one model is a constrained version of the other. The likelihood ration test can be computed by taking the difference between the -2 Res Log Likelihood values of the full and reduced models. From the expression of -2 Res Log Likelihood, it is clear that,

$$\begin{aligned} -2\log L &= -2\left[\log \max_{H_0} g(\Sigma|\text{data})\text{- }\log \max_{\textbf{unrestricted}} g(\Sigma|\text{data})\right] \\ &= \left[-2\log g(\widehat{\Sigma}_{H_0}|\text{data})\right] - \left[-2\log g(\widehat{\Sigma}_{\textbf{unrestricted}}|\text{data})\right], \end{aligned} \tag{6.4}$$

where $\widehat{\Sigma}_{H_0}$ and $\widehat{\Sigma}_{\textbf{unrestricted}}$ are the maximum likelihood estimators of Σ under H_0 and without any restriction on Σ respectively.

REML is often preferred to maximum likelihood estimation as a method of estimating covariance parameters in linear models because it takes account of the loss of degree of freedom in estimating the mean and produces unbiased estimating equations for the variance parameters. The statistical analysis system (SAS) options in the PROC MIXED procedure are included in the below parenthesis.

Here are a few selected covariance structures.

1. $\Sigma = \sigma^2 I$ (VC)
2. $\Sigma = \sigma_1^2 J + \sigma_2^2 I$ (CS)
3. Σ unstructured (UN)
4. $\Sigma = diag(\sigma_1^2, \ldots, \sigma_m^2)$: Banded main diagonal (UN(1))

5. $\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho^{m-1} \\ \rho & 1 & \cdots & \rho^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \cdots & 1 \end{bmatrix}$: Autoregressive of order 1 (AR(1))

6. $\Sigma = \sigma^2 \begin{bmatrix} \sigma_0 & \sigma_1 & \cdots & \sigma_{m-1} \\ \sigma_1 & \sigma_0 & \cdots & \sigma_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m-1} & \sigma_{m-2} & \cdots & \sigma_0 \end{bmatrix}$: Toeplitz (TOEP)

7. $\Sigma = \sigma^2 \begin{bmatrix} \sigma_0 & \sigma_1 & 0 & \cdots & 0 \\ \sigma_1 & \sigma_0 & \sigma_1 & \cdots & 0 \\ 0 & \sigma_1 & \sigma_0 & \sigma_1 & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma_1 & \sigma_0 \end{bmatrix}$: Two Bands Toeplitz (TOEP(2))

8. $\Sigma = \sigma^2(\rho_{ij}^{d_{ij}}),\ \rho_{ii} = 1$: Spatial Power or Marcov (SP(POW)(c))

9. $\Sigma = (\sigma_{ij}),\ \sigma_{ij} = \frac{\sigma_{ii}+\sigma_{ij}}{2} - \lambda$, if $i \neq j$: Huynh-Feldt (HF).

Potential problems of using the LR test to compare covariance models include parameters that may be on the boundary of the parameter space and that the models being compared may not be nested since the comparison may be inconsistent, depending upon which model was taken as the full model.

To address these problems in model selection, two other model selection criteria have been used. Two information criteria frequently used in repeated measures analysis are Akaike's (1973) Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC) (1978). Both the AIC and BIC penalize the log-likelihood for the number of parameters and number of observations. The model with the smallest AIC (BIC) is best. Table 11 presents model fitting estimates of covariance parameters for the first-order autoregressive covariance structure fitted to the MDS with MS patient data. As Table 12 shows, the results are following that as the age of the subjects increases, their ADL long scale decreases meaning that they stayed less in the nursing home. The age of residents are negatively correlated with the ADL long scale. The time and cogscale effects are positively correlated with the ADL long scale, namely, when the subjects have a large time effect and the high cognitive scale effect, they have a high ADL scale. The BMI index has a negative coefficient. This indicates that a higher BMI index person has a tendency to a lower ADL long scale. As Table 13 shows, we see that if the AIC and the BIC are used to select the covariance structure from the TOEP(2), AR(1), CS candidate models, then the AR(1) structure would be selected as it has the smallest values. The unstructured (UN) correlation presents the number of parameters $190 = \frac{(19)}{2}$ that need to be estimated, the Hessian matrix is not positive definite, and the convergence is not met.

Table 11: Model fitting estimates of covariance parameters for the first-order autoregressive covariance structure fitted to the MDS with MS patient data.

Effects	Estimates	Standard Error	z-value	p-value
Intercept	35.088	0.503	69.81	$< .0001$
AR(1)	0.282	0.008	33.68	$< .0001$
Residual	10.627	0.120	88.90	$< .0001$

Table 12: Model fitting estimates of fixed effects parameters for the first-order autoregressive covariance structure fitted to the MDS with MS patient data.

Effects	Estimates	Standard Error	z-value	p-value
Intercept	23.294	0.272	85.63	$< .0001$
Time	0.001	0.0001	12.54	$< .0001$
Age	-0.049	0.004	-12.47	$< .0001$
BMI	-0.100	0.005	-20.49	$< .0001$
Cogscale	1.147	0.017	67.77	$< .0001$

Table 13: Information criteria results for unstructured (UN), banded structure (TOEP(2)) first-order autoregressive (AR(1)), and compound symmetry (CS) structures fitted to the MDS with MS patient data.

Model	Covariance Parameters	-2 Log likelihood	AIC	BIC
TOEP(2)	2	293773.4	293779.4	293801.8
AR(1)	2	293529.7	293525.7	293548.1
CS	2	294814.7	294820.3	294842.7

CHAPTER VII

CONCLUSION

7.1 Summary

The main goals of this dissertation are to examine the theoretical and empirical justifications of moving block bootstrap techniques in longitudinal data that consist of a large number of replications for relatively small number of subjects, and the diagonal optimal weights for unbalanced longitudinal designs having a large number of subjects and a small number of replications. In Chapter III, we presented standard statistical models for repeated measurement data when the response variable is continuous.

In Chapter IV, moving block bootstrap methods are used for analyzing longitudinal data in which a small number of subjects have a large number of replications over time by investigating the efficacy and utility of the methodology, theoretically and empirically, through a small simulation study. Those have second order optimality in the case of dependent stationary data, under regular conditions.

In Chapter V, we presented a way to find diagonal optimal weights for unbalanced longitudinal data in terms of the asymptotic mean squared error of regression coefficient. The performance of diagonal optimal weights was investigated via a simulation study.

In Chapter VI, we provided a detailed examination of the data set concerning nursing home residents with multiple sclerosis, which was obtained from a large database termed the minimum data set. Using the AIC and BIC criterion, we selected the correlation structure for each patient and made an inference for fixed effects allowing for a random intercept factor.

7.2 Future Work

There are several topics to be considered beyond the works completed in this dissertation. We now discuss some possible future research work.

The circular block bootstrap (Politis and Romano, 1992), alternatives to moving block bootstrap, which have an advantage of reducing the bias of the bootstrap variance, can be extended in longitudinal data. The tapered block bootstrap method (Paparoditis and Politis, 2001; Paparoditis and Politis, 2002) in which each block end points are shrunk toward a target value before being concatenated to form a bootstrap pseudo-series, which indeed leads a more accurate variance estimator, can be used in a longitudinal setting. We need to explore the optimal block length using the other block bootstrap methods in longitudinal data. The correlation structure of different time measurements can be extended to long range dependence and nonstationary dependence.

While most of our work has focused on linear constraints in our repeated measurement model, developments with nonlinear constraints might also be possible. Other possible extensions include binary or polytomous data in repeated measurements.

The possible nonparametric or semiparametric estimation methods using block bootstrap technique in the analysis of longitudinal data pose interesting problems for future research. Furthermore, it may be desirable to develop a methodology for irregularly spaced repeated measurement data.

REFERENCES

Arnold, S. (1981). *The Theory of Linear Models and Multivriate Analysis*. New York: Wiley.

Babu, G. J. (1986). "Bootstrapping statistics with linear combinations of chi-squares as weak limit." *Sankhya Ser A.*, 56, 85–93.

Bhattacharya, R. N. and Qumsiyeh, M. (1988). "Second order and L^p comparisons between the bootstrap and empirical edgeworth expansion methodologies." *The Annals of Statistics*, 17, 160–169.

Bhattacharya, R. N. and Ranga Rao, R. (1986). *Normal Approximations and Asymptotic Expansions*. Malabar, FL: Krieger.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed. New York: Springer-Verlag.

Bühlmann, P. (1994). "Blockwise bootstrapped empirical process for stationary sequences." *The Annals of Statistics*, 22, 995–1012.

— (1997). "Sieve bootstrap for time series." *Bernoulli*, 3, 128–148.

Bühlmann, P. and Kunsch, H. R. (1995). "The blockwise bootstrap for general parameters of a stationary time series." *Scan. Journal of Statistics*, 22, 35–54.

Carlstein, E. (1986). "The use of subseries methods for estimating the variance of a general statistic from a stationary time series." *The Annals of Statistics*, 14, 1171–1179.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. New York: Springer-Verlag.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge, UK: Cambridge University Press.

Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. New York: Oxford University Press Inc.

Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*. 2nd ed. New York: Chapman & Hall/CRC.

Efron, B. (1979). "Bootstrap methods: Another look at the jackknife." *The Annals of Statistics*, 7, 1–16.

Feng, Z., McLerran, D., and Grizzle, J. (1996). "A comparison of statistical methods for clustered data analysis with Gaussian error." *Statistics in Medicine*, 15, 1793–1806.

Goncalves, S. and White, H. (2002). "The bootstrap of the mean for dependent heterogeneous arrays." *Econometric Theory*, 18, 1367–84.

Granger, C. W. and Joyuex, R. (1980). "An introduction to long-memory time series models and fractional differencing." *J. Time Series Anal.*, 1, 15–29.

Graybill, F. A. (1976). *Theory and Application of Linear Model*. MA: Duxbury Press.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, P., Horowitz, J., and Jing, B. Y. (1995). "On blocking rules for the bootstrap withe dependent data." *Biometrika*, 82, 561–574.

Hand, D. J. and Crowder, M. J. (1996). *Practical Longitudinal Data Analysis*. London, UK: Chapman & Hall.

Hartley, H. O. and Rao, J. N. K. (1967). "Maximum likelihood estimation for the mixed analysis of variance model." *Biometrika*, 54, 93–108.

Hocking, R. R. (1985). *The Analysis of Linear Models*. Pacific Grove, CA: Brooks Cole.

Hosking, J. R. M. (1981). "Fractional differencing." *Biometrika*, 68, 165–176.

Huang, J. Z., Wu, W. O., and Zhou, L. (2002). "Varying-coefficient models and basis function approximations for the analysis of repeated measurements." *Biometrika*, 89, 111–128.

Keselman, H. J., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1998). "A comparision of two approaches for selecting covariance structure in the analysis of repeated measurements." *Communication and Statistics., Series. B*, 27, 591–604.

Künsch, H. R. (1989). "The jackknife and the bootstrap for general stationary observations." *The Annals of Statistics*, 17, 1217–1261.

Lahiri, S. N. (1992). "Bootstrapping m-estimators of a multiple linear regression parameter." *The Annals of Statistics*, 20, 1548–1570.

— (1993). "On the moving block bootstrap under long range dependence." *Statistics and Probability Letters*, 18, 405–413.

— (1994). "On two-term Edgeworth expansions and bootstrap approximations for studentized multivariate m-estimators." *Sankhya Ser A.*, 56, 201–226.

— (1995). "On the asymptotic behaviour of the moving block bootstrap for normalized sums of heavy-tailed random variables." *The Annals of Statistics*, 23, 1331–1349.

— (1996). "On Edgeworth and moving block bootstrap for studentized m-estimators in multiple linear regression models." *Journal of Multivariate Analysis*, 56, 42–59.

— (1999). "Theoretical comparisons of block bootstrap methods." *The Annals of Statistics*, 27, 386–404.

— (2001). "Effects of block lengths on the validity of block resampling methods." *Probability Theory and Related Fields*, 121, 73–97.

Laird, N. M. and Ware, J. H. (1982). "Random-effects models for longitudinal data." *Biometrics*, 38, 963–974.

Liang, K. Y. and Zeger, S. L. (1986). "Longitudinal data analysis using generalized linear models." *Biometrika*, 73, 13–22.

Lindsey, J. K. (1999). *Models for Repeated Measurements*. 2nd ed. New York: Oxford University Press Inc.

Liu, R. Y. and Singh, K. (1992). "Moving blocks jackknife and bootstrap capture weak dependence." In *Exploring the Limits of Bootstrap*. New York: Wiley, pp. 225-248.

Paparoditis, E. and Politis, D. N. (2001). "Tapered block bootstrap." *Biometrika*, 88, 1105–1119.

— (2002). "Tapered block bootstrap for general statistics from stationary sequences." *Econometric Journal*, 5, 131–148.

Politis, D. N. and Romano, J. P. (1992). "A circular block resampling procedure for stationary data." In *Exploring the Limits of Bootstrap*. New York: Wiley, pp. 263-270.

Porell, F., Caro, F. G., Silva, A., and Monane, M. (1988). "A longitudinal analysis of the nursing home outcomes." *Health Services Research*, 33, 835–865.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York: John Wiley & Sons, Inc.

Royall, R. M. (1986). "Model robust confidence intervals using maximum likelihood estimators." *International Statistical Review*, 54, 221–226.

Searle, S. R. (1997). *Linear Models*. Classic ed. New York: Wiley.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. New York: Wiley.

Seber, G. A. F. (1977). *Linear Regression Analysis*. New York: Wiley.

Singh, K. (1981). "On the asymptotic accuracy of the Efron's bootstrap." *The Annals of Statistics*, 9, 1187–1195.

Xie, M. and Yang, Y. (2003). "Asymptotics for generalized estimating equations with large cluster sizes." *The Annals of Statistics*, 31, 13–22.

VITA

Hyunsu Ju, son of Sung Moon Ju and Il Sun Yang, was born on October 7, 1970, in Junnam, Korea. He received a Bachelor of Science in statistics in February of 1996 from Chung-Ang University. In February 1998, he received a Master of Arts in statistics from Seoul National University. He went to America to pursue his Ph.D. in August of 1999 in the Department of Statistics at Texas A&M University. He received his Ph.D. in statistics under the direction of Dr. Suojin Wang in December 2004. His permanent address is

530-62 Sadang 1 Dong SeoChoGU

Seoul, Korea, 130-061