RESAMPLING METHODS FOR LONGITUDINAL DATA ANALYSIS

YUE LI

NATIONAL UNIVERSITY OF SINGAPORE

2005

Resampling Methods for Longitudinal Data Analysis

YUE LI

(Bachelor of Management, University of Science and Technology of China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2005

Acknowledgements

Learning is a journey. Many people have helped me on my journey here in Singapore. Now it is a good time to express my gratitude to those people.

No line of this thesis could have been written without my dear husband Zeng Tao by my side, teaching me what is really important. I thank him for his patience, understanding and company sometimes up to 3 am in the morning. I also want to thank my family back home in China. They are always there standing by me and unconditionally supporting me.

I owe much more than I can express here to my supervisor, Associate Professor You-Gan Wang for all his patience, kind encouragement and invaluable guidance. The spark of his ideas always impresses me and inspires me to learn more. I sincerely appreciate all his effort and time for supervising me no matter how busy he was with his own work. It is really a pleasure to be his student. Thanks to many other professors in the department who have helped me greatly, namely Prof Zhidong Bai, Associate Professor Zehua Chen and Professor Bruce Brown for their helpful comments, suggestions and advice.

Thanks to National University of Singapore for providing me with the research scholarship so that I could come to this beautiful country, get to know so many

kind people and learn from them. Special thanks to our department for providing a convenient studying environment and to Mrs. Yvonne Chow for the assistance with the laboratory work.

There are many other people I would like to thank: Dr. Henry Yang from Bioinformatics Institute of Singapore, who gave me guidance during my internship; my dear friends, Ms. Wenyu Li, Ms. Rongli Zhang, Ms. Min Zhu, Ms. Yan Tang, Mr. Zhen Pang, and Mr. Yu Liang, for their help, encouragement and the enjoyable time I spent with them; those young undergraduate students I have taught in statistical tutorials, for everything I have shared with and learned from them.

Last but not least, I would like to thank Mrs. Rebecca Thai for her careful proof reading of my thesis.

Yue LI, Lily

National University of Singapore

Dec 2005

Contents

List of Tables

List of Figures

Summary

Dear friend, theory is all gray,

and the golden tree of life is green.

Goethe, from "Faust".

Longitudinal data are becoming more and more common in study designs for many research areas. One of the most widely applied statistical models for longitudinal data analysis is the Generalized Estimating Equation (GEE) model (Liang and Zeger 1986), which is the basis of this work. GEEs are appropriate for modeling the population mean of continuous and discrete responses when the correlation structure is not the focus. GEEs are preferred because they can provide consistent regression parameter estimates even when the within-subject correlation structure is misspecified. As is common to many robust methods with respect to one particular aspect of model misspecification, GEEs also pay the price for robustness. That is the underestimation of the variance for regression parameters, which is due to the approximation of response variance matrix by the moment estimate, namely the products of residuals.

The purpose of this work is to apply the resampling idea to longitudinal data

for the regression parameter estimation and provide better alternatives to the common estimation procedures in GEEs, especially in terms of variance estimation and of confidence interval construction. Two types of resampling approaches are proposed. The first approach is "smooth bootstrap," a random perturbation to the estimating algorithms, which provides a simple way to produce bootstrapped copies of parameter estimates. Two versions of smooth bootstrap methods are investigated analytically and via Monte Carlo simulation. One version retains the robustness to the misspecification of the within-subject correlation structure. The other version is model-based and hence more efficient when the covariance model is correctly specified. When compared to the commonly used sandwich estimators and some classical resampling methods applied to the longitudinal data, the smooth bootstrap methods yield more accurate variance estimates and confidence intervals for different types of data and sample sizes.

The second resampling approach proposed in this thesis is based on the estimating function rather than the parameter estimates. Several simple perturbation methods based on two versions of studentized estimating function statistics are suggested for parameter and the variance estimation. The bootstrap distribution of two versions of studentized estimating functions can be obtained. The endpoints of confidence interval for parameter estimates could be solved from only two equations defined at the quantiles of the bootstrap distribution of the studentized EF statistic. The resulting confidence intervals turn out to perform quite well for different types of data and different sample sizes. Particularly, one of the two versions of the studentized EF statistics is a first-term corrected studentized estimating function statistic obtained from the Edgeworth expansion. Bootstrapping this first-term corrected statistic will give even higher order distribution approx-

imation and lead to improved confidence intervals for the regression parameter estimates.

Chapter 1

Introduction

1.1 Longitudinal studies

1.1.1 Background

Analysis of repeated measurements has achieved great popularity in recent years because repeated measurements are commonly found in study designs of more and more research areas, such as clinical trial studies, toxicological studies and even advanced systems biological studies. In terms of "repeated measurements," we refer broadly to the data with multiple observations (such as different occasions or conditions) for each experimental unit or subject. If multiple observations are collected over a period of time, the data are known as "longitudinal data" (also called "panel data"). Repeated measurements, including longitudinal data, have many advantages for scientific studies. First, this design of data structure is the only design that is able to obtain the information that concerns individual pattern of change. Second, each unit or subject can serve as its own control,

because the measurements can be taken under both control and experimental conditions. This reduces the number of subjects, removes the variation across subjects and increases the power of analysis compared to a cross-sectional design with the same number of subjects. However, repeated measurements also raise a number of challenges for statistical analysis. The key characteristic of repeated measurement data is the possible dependence within the observations for each experimental unit or subject, which introduces correlation into the consideration, violates the common assumption of classical statistical methods and thus makes the analysis much difficult. Ignoring the correlation when it does exist could cause inefficient estimates of regression parameters and result in lower power to detect differences of interest because a number of degrees of freedom must be used to estimate the association parameters. Furthermore, in practice, when the number of the observations for each subject is not common across all the subjects or the observations are not regularly time-spaced, the data is unbalanced. For example, in the study of litter effects, the sizes of litters usually differ, or the patients under study may go back for checkups at different times. Besides imbalance, repeated measurements can also be incomplete due to some factor relevant or irrelevant to the studies. For example, in clinical trials, some patients may fail to be followed up within a certain period of time which will result in missing data. Both the unbalanced and incomplete structures of repeated measurements make the analysis of such data even harder. Therefore, appropriate statistical models and corresponding analysis methodology are in great demand to deal with such kind of data.

Table 1.1 shows the general structure of the longitudinal data which will be used throughout this dissertation (Strategies dealing with missing data are beyond

the scope of this thesis. Hence no missing data will be assumed throughout the thesis.)

Table 1.1: General structure of longitudinal data

Subject	Time	Response	Covariates		
1	1	y_{11}	x_{111}	$\cdots$	x_{11p}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	y_{1j}	x_{1j1}	$\cdots$	x_{1jp}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$
	n_1	y_{1n_1}	x_{1n_11}	$\cdots$	x_{1n_1p}
.....	.....	.....	.....	.....	.....
i	1	y_{i1}	x_{i11}	$\cdots$	x_{i1p}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	y_{ij}	x_{ij1}	$\cdots$	x_{ijp}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$
	n_i	y_{in_i}	x_{in_i1}	$\cdots$	x_{in_ip}
.....	.....	.....	.....	.....	.....
K	1	y_{K1}	x_{K11}	$\cdots$	x_{K1p}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	j	y_{Kj}	x_{Kj1}	$\cdots$	x_{Kjp}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$
	n_K	y_{Kn_K}	x_{Kn_K1}	$\cdots$	x_{Kn_Kp}

Let K be the number of subjects, n_i be the number of observations for subject i, and y_{ij} be the observations for subject i at time j, where $i = 1, \ldots, K$ and $j = 1, \ldots, n_i$. Let $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ be the vector of covariates for response y_{ij} and p be the number of covariates; hence the dimension of regression parameter of interest. The covariates could be random, random but time-independent, or completely non-stochastic. In matrix form, $Y_i = (y_{i1}, \ldots, y_{in_i})^T$ is the $n_i \times 1$ response vector for subject i and the corresponding covariates X_i being $n_i \times p$ matrix for $i = 1, \ldots, K$.

Actually, almost all types of correlated data could be expressed in such a layout. Different scales of K and n_i will result in different levels of statistical

inference. Below are some special cases of correlated data which are commonly seen, arranged in an increasing order of difficulty in statistical analysis:

- Time Series data: $K = 1$, $n_i = n$ is large.
- Multivariate data:

  $K > 1$, n_i small or moderate; independence among subjects.

  Such as longitudinal/panel data or cluster data.
- Multiple Time Series: $K > 1$, n_i is large; subjects are dependent.
- Spatial data: both K and n_i are hopefully large; rows are dependent.

The focus of this thesis will be longitudinal data that are frequently observed in biomedical and biological research areas. In longitudinal data analysis, rapid development of statistical research have been seen in recent years. Good references for overview of research relevant to longitudinal data are Diggle *et al.* (2002), Davis (2001), and Fitzmaurice *et al.*. (2004). In the following sections, some important achievements in the development of statistical analysis for longitudinal data will be reviewed.

1.1.2 Statistical models for longitudinal data

Since the second half of the 20th century, a variety of statistical approaches for longitudinal data have been studied, such as normal-theory method assuming the normality of the responses' distributions (see for example Timm 1980; Ware 1985) and weighted least squares method for categorical responses (see for example Grizzle *et al.* 1969; Koch *et al.* 1977). However, those early methods are

mainly specific to some types of variables or require strong assumption of the full distribution of the responses. Therefore their applications are very limited. To provide a unifying framework, the idea of "generalized linear models" (GLM) was introduced, which enables the analysis for both continuous and categorical data and can analyze data from different distributions: normal, gamma, binomial, Poisson, etc. (Nelder and Wedderburn 1972; McCullagh and Nelder 1989; Dobson 1990). This is one of the most important breakthroughs in statistical modeling. But it still requires the assumption for the distribution of the responses. The applicability of GLM is further extended by quasi-likelihood approach (Wedderburn 1974). The quasi-likelihood approach only needs to assume the first two moments of the distribution for the responses. Those important developments enable many related types of extensions of GLM and quasilikelihood methods towards the analysis of correlated data including longitudinal data. Such extensions are marginal models, transition models and random-effects models (Zeger and Liang 1992). The differences among those three models are the interpretations of the regression coefficients, especially for categorical outcome variables. Transition models are appropriate when it is reasonable to assume that responses follow a stochastic process depending on the subject only through the values of the measured covariates. Random effects models can model the heterogeneity among subjects; hence, the regression coefficients explain the effect of covariates, not only on the average of the population, but also on one subject's response. Marginal models only focus on the population average, which is the most common scientific objective. Furthermore, in marginal models, the association among observations from each subject is modeled separately from the regression model, while random effects and transition models model the covariate effects and within-subject correlation

through one single equation.

Marginal model is the focus of this thesis. Here we give justifications of preference for marginal model. First, statistical analysis using the marginal model is conceptually and computationally simple. For example, the marginal model can handle model structure that is the same for all the observations regardless of the number of observations for each subject. Therefore, the marginal model is simple, or "reproducible" following the terminology of Liang *et al.* (1992). Second, for the same reason, missing observations (for example, missing completely at random) are easily accommodated when using the marginal models by simply omitting those non-informative missing observations in the analysis. However, missing observations severely complicate analysis of fully or partly conditional models such as transitional model. As these observations are an explicit part of the conditional regression model for each measurement within the subject, they cannot simply be ignored in models like transitional model. Third, the marginal model can not only be applied to longitudinal data, but also to a large group of repeated measurements data, such as clustered data. When applying marginal models, cautions must be taken. For example, it is found that marginal models will tend to give biased estimates with time-dependent covariates unless working independence is assumed or a key assumption is verified (see Pepe and Anderson 1994, Emond *et al.* 1997, Pan *et al.* 2000, Schildscrout and Heagerty 2005 and references therein).

As an extension of the quasi-likelihood method to multivariate correlated responses, a marginal model approach, generalized estimating equation (GEE) method, was developed in Liang and Zeger (1986)'s seminal paper. This method

gives consistent and asymptotically normal estimates even when the correlation structure is misspecified (a comprehensive discussion about GEE can be found in Hardin and Hilbe 2002). The corresponding variance-covariance estimator, named "sandwich estimator" or "robust estimator", is robust and widely used. GEE provides us with a flexible model that can accommodate different structures of longitudinal data. In other words, the number of observations for each subject do not have to be constant, and the measuring times need not be the same across subjects. Missing data can also be accommodated under the restriction that the missing data must be MCAR (missing completely at random). All these nice properties make GEE method widely applied in correlated data analysis. The proposed new methods in this dissertation will be applied in GEE procedures and we think the application of the new methods can be easily extended to other estimating procedures with estimating equations. The details and extensive discussions about GEE procedures will be given in the next chapter.

1.2 Resampling methods

1.2.1 Introduction

To begin with, resampling methods are not panacea!

The incredibly fast development in computer power and the emergence of large numbers of friendly statistical packages have boosted the computer-intensive statistical techniques – "resampling" approaches. Resampling methods provide us with an alternative to standard statistical approaches (such as maximum likelihood estimation, etc.) for the analysis related to the sampling distribution of

estimated parameters, such as standard deviation, MSE and confidence intervals. The persistent efforts of numerous researchers have made the technique of resampling more applicable and efficient particularly in the analysis for independent data. The commonly used resampling schemes could be classified into four main categories: the permutation test developed by Fisher (1935), cross-validation proposed by Kurtz (1948), the jackknife by Quenouille (1949) and Tukey (1958), and the bootstrap proposed by Efron (1979). The bootstrap resampling scheme tends to be more versatile than the others in its wide range of applications. It is debatable whether the jackknife or bootstrap schemes are superior for their efficiency and robustness (Liu and Singh 1992, Wu 1986, Hu 2001). For a good introduction to resampling methods or bootstrap, please refer to Efron (1982), Efron and Tibshirani (1993) and Good (2001). Since resampling methods can give answers to a large class of statistical problems without strict structural assumptions on the underlying distribution of data, the applications of resampling methods have been realized in more complicated data structures, such as correlated data of structure described in Table 1.1.

1.2.2 Resampling methods for correlated data

There have been many attempts to extend the resampling methods to the correlated data in various forms and different inference problems. Lahiri (2003) provides an elaborate reference of bootstrap theory and methods for the analysis of times series and spatial data structures. Several block bootstrap methods are discussed and compared with great details in applications and theories.

Under the settings of GEE for longitudinal data, the application of various re-

sampling methods are also of great interest. Most of the existing relevant work is mainly for variance estimation, confidence interval construction or hypothesis testing. Paik (1988), Lipsitz *et al.* (1990b), Ziegler *et al.* (2000), Yan and Fine (2004) have investigated the jackknife variance estimators to the estimated parameters. Moulton and Zeger (1989) and Shermann and le Cessie (1997) proposed bootstrap methods in the variance estimation of GEE regression estimators. Those resampling variance estimators have comparable performance to the Liang and Zeger (1986)'s sandwich estimator. We refer to those methods as "classical jackknife" or "classical bootstrap" methods because they directly applied the classical idea of jackknife or bootstrap by resampling the data itself. One inherent problem for those direct jackknife or bootstrap methods is that the consistency will be affected when the sample size is not large enough. Therefore the resampling procedures may generate one or more "wild" estimators or even have singularity or convergence problem. Although "one-step" iteration is suggested in case of small number of subjects (Paik 1988, Moulton and Zeger 1989, Lipsitz *et al.* 1990b), it will still affect the accuracy of estimates and the final inference based on all the estimates. Unavoidably, due to the nature of those direct jackknife or bootstrap methods, they require great computational time, because for example, in most of the methods mentioned above, the same GEE procedure will repeat each time when deleting a subject or obtaining a bootstrap resample from all the blocks.

Rather than applying classical jackknife and bootstrap to the data itself, many researchers have investigated the problem from a totally different view: resampling based on the estimating functions. Considering that most of the estimating functions can be expressed as the sum of finite independent items, Lele (1991a) applied the jackknife idea to the items of the estimating functions; Hu and Zidek (1995)

proposed the bootstrap version in linear regression problems; Jin *et al.* (2001) bootstrapped the objective functions with U-statistic structure. Lele (1991b) further discussed the application of jackknifing or bootstrapping estimating functions for a sequence of non-independent, non-identically distributed random variables, e.g. time series data. A common characteristic of these methods is that, the finite terms of estimating functions are recombined in one way or another to obtain an estimate of the resampling distribution of parameter estimates. We refer to those methods as "EF-based resampling" methods. If there is known feature about the distribution of the estimating functions, for example, the estimating function is pivotal, a more accurate approach was proposed by Parzen, Wei and Ying (1994). They set the estimating functions equal to random values from the known pivotal distribution and obtained parameter estimates by repeating this procedure in a resampling manner. This method takes advantage of the pivotal property of estimating functions and is expected to gain more efficiency when compared to the others. But, the vital assumption of pivots here may not always be valid in practice. If one can mimic the way the estimating functions vary in their own distributions, even when the distributions are unknown, the idea of Parzen *et al.* (1994) can be extended to more general cases. The "Estimating Function Bootstrap" (EFB) proposed by Hu and Kalbfleisch (2000) is another type of "EF-based resampling" method. The EFB resorts to the bootstrapping distribution of estimating functions (EF) instead of that of estimates and inverts the quantiles of bootstrapping distribution of EF to the quantiles of parameter estimates. Their idea was applied to estimating functions with U-statistic structure by Jiang and Kalbfleisch (2004). Details of EFB will be discussed further in Chapter 5.

1.3 Aim and structure of the dissertation

The aim of this thesis is to investigate the more general application of resampling methods to the analysis of longitudinal data such as variance estimation and confidence interval construction. The basic tools used are Monte Carlo simulations and Edgeworth expansion. The proposed methods are focused on the estimating equation or estimating function with possibly unknown limiting distributions. A practical guideline for the application of resampling methods in the longitudinal data analysis is suggested.

The thesis is organized as follows. In Chapter 2, GEE procedures are studied in details to serve as the summary of common methods and different interpretations of improved approaches are provided. In Chapter 3 two versions of smooth bootstrap methods are introduced and extensive simulation studies for investigation of those methods are discussed in Chapter 4. In Chapter 5, a proposal of the other type of resampling scheme is presented, based upon the studentized EF statistics, including a first-term corrected studentized EF statistic obtained from Edgeworth expansion. In Chapter 6, concluding remarks, discussions and some topics for future research are given.

Chapter 2

GEE procedure

2.1 GEE procedure

Popularity of GEE approaches can be seen from the enormous literature on the improvement and extensions of GEE method since first proposed in Liang and Zeger (1986). There are mainly two categories: GEE1 and GEE2. The term "GEE1" refers to the methodology in which two individual estimating equations for regression and association parameters respectively are used in an iterative manner (moment estimators for the association parameters in Liang and Zeger 1986; an ad-hoc estimating equation in Prentice (1988) and Prentice and Zhao 1991). GEE1 approach requires only first and second moment assumptions and provides consistent regression parameter estimators even when the covariance model is misspecified. However GEE1 approach has the problems of inconsistent estimation of the correlation parameters and the unstable estimation for the variance of regression parameter estimates. However, choices of association estimators would also affect the asymptotic efficiency of regression estimators. Research on the estima-

tion of association parameters can be found in Lipsitz *et al.* (1990a), Park *et al.* (1998), Wang and Carey (2003, 2004), among others. "GEE2" approach allows simultaneous estimation of regression and association parameters. It requires even the third and fourth moments of the responses and produces consistent estimates of regression and association parameters only when both the mean and variance function are correctly specified (Zhao and Prentice 1990; Prentice and Zhao 1991; Liang *et al.* 1992). Extended quasi-likelihood approaches that have a close connection with GEE1 and GEE2 have been investigated by Hall and Severini (1998) and Hall (2001). In practice, since the correct model for the correlation structure is usually unknown, GEE1 may be more appropriate in terms of robustness. The discussions in this dissertation will be based on models like GEE1. The proposed resampling strategies could be extended to GEE2 framework in a straight forward manner.

One of the most important aspects in GEE is the variance estimation for the regression parameter estimates. In statistical inference it is important to obtain not only parameter estimates but also their asymptotic covariances. A very popular variance estimator, robust sandwich variance estimator, was proposed in Liang and Zeger (1986). As long as the mean and variance functions in the marginal model are correctly specified, the sandwich estimator can give consistent estimation even if the working correlation is misspecified. However, this sandwich estimator generally underestimates the true variance of the parameters and even is inconsistent in some cases. Much literature has discussed the performance of this sandwich estimator in different scenarios (for example, Paik 1988, Shermann and le Cessie 1997), and improved versions have been investigated (Mancl and DeRouen 2001, Kauermann and Carroll 2001, Pan 2001).

Next, we will introduce the GEE, the construction of the sandwich estimator and its properties. Referring to the data structure in Table 1.1, y_{ij} and x_{ij} are the response and the covariate vector for subject i at time j respectively, for $1 \leq i \leq K$ and $1 \leq j \leq n_i$. The response y_{ij} has marginal mean $\boldsymbol{\mu}_{ij}$ and marginal covariance $\boldsymbol{\phi}\,\boldsymbol{\sigma}^2_{ij}$, where $\boldsymbol{\mu}_{ij}$ and $\boldsymbol{\sigma}^2_{ij}$ are known mean and variance function linked to the covariates: $\boldsymbol{\mu}_{ij} = h\,(x^T_{ij}\,\boldsymbol{\beta})$ and $\boldsymbol{\sigma}^2_{ij} = v\,(\boldsymbol{\mu}_{ij})$. The working covariance of Y_i, V_i is then in the form of $\boldsymbol{\phi}\,A_i^{1/2} R_i\,A_i^{1/2}$, where A_i is the diagonal matrix of variances, diag $(\boldsymbol{\sigma}^2_{ij})$; $R_i(\boldsymbol{\alpha})$ is the "working" correlation matrix; $\boldsymbol{\alpha}$ is the correlation coefficient and $\boldsymbol{\phi}$ is the scale parameter usually used to explain overdispersion or underdispersion. The unknown parameters are $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\phi})^T$, and the true values are denoted as $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{\phi}_0)^T$. The GEE procedure for estimating $\boldsymbol{\beta}$ for given $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ proposed by Liang and Zeger (1986) is (GEE1 procedure):

$$\mathbf{U}(\boldsymbol{\beta}) := \sum_{i=1}^{K} D_i^T(\boldsymbol{\beta})\, V_i^{-1}(\boldsymbol{\theta})\, \boldsymbol{\epsilon}_i(\boldsymbol{\beta}) = 0_{p\times 1}, \tag{2.1.1}$$

where $D_i(\boldsymbol{\beta}) = \partial\mu_i(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ and $\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) = Y_i - \mu_i(\boldsymbol{\beta})$.

Iterative algorithm is used to obtain the solution to (2.1.1) given $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\phi}}$:

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} + \left(\sum_{i=1}^{K} \widehat{D}_i^T\, \widehat{V}_i^{-1} \widehat{D}_i\right)^{-1} \left(\sum_{i=1}^{K} \widehat{D}_i^T\, \widehat{V}_i^{-1} \widehat{\boldsymbol{\epsilon}}_i\right), \tag{2.1.2}$$

where $\widehat{D}_i$ and $\widehat{\boldsymbol{\epsilon}}_i$ are evaluated at $\widehat{\boldsymbol{\beta}}^{(m)}$ and $\widehat{V}_i$ is evaluated at $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^{(m)}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}})^T$. The consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}$ is proved in Liang and Zeger (1986). The estimate of scale parameter $\boldsymbol{\phi}$ is conventionally obtained by the estimated original or Pearson residuals via method of moments. Supplementary estimating equations for $\boldsymbol{\alpha}$ in V_i in (2.1.1) are needed so that iterative process can be used to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Care must be taken in choosing an appropriate estimator or estimating functions for $\boldsymbol{\alpha}$ to avoid the problems identified by Crowder

(1995). The choice of the $\boldsymbol{\alpha}$-estimator will affect the asymptotic efficiency of the $\hat{\boldsymbol{\beta}}$ obtained from (2.1.1), unless V_i is correctly specified.

Suppose the parameter estimators are $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}})^T$. Delta method provides the covariance of $\sqrt{K}\,\widehat{\boldsymbol{\beta}}$ as:

$$\left(\frac{1}{K}\sum_{i=1}^{K} M_i\right)^{-1} \left(\frac{1}{K}\sum_{i=1}^{K} D_i^{\mathrm{T}}\, V_i^{-1}\, \bar{V}_i\, V_i^{-1}\, D_i\right) \left(\frac{1}{K}\sum_{i=1}^{K} M_i\right)^{-1}. \tag{2.1.3}$$

Under mild regular conditions (Liang and Zeger 1986), (2.1.3) will converge to

$$V_R = \mathcal{M}^{-1}\ \mathcal{V}_U\ \mathcal{M}^{-1}, \tag{2.1.4}$$

as $K \to \infty$, where $M_i(\boldsymbol{\theta}) = D_i^T(\boldsymbol{\beta})\, V_i^{-1}(\boldsymbol{\theta})\, D_i(\boldsymbol{\beta})$ is a symmetric matrix, $\bar{V}_i$ is the true covariance matrix of Y_i, and the positive definite matrices $\mathcal{M}$ and $\mathcal{V}_U$ are the limit for $\sum_{i=1}^{K} M_i / K$ and $\mathrm{Var}\,(U(\boldsymbol{\beta}_0))/\,K$ respectively. If $\mathbf{U}$ is a score function or quasi-score function, it is particularly true that $M_i = -E(\partial \mathbf{U}_i/\partial\boldsymbol{\beta})$. The "sandwich estimator" is then obtained by using the product of residuals $\widehat{\boldsymbol{\epsilon}}_i\,\widehat{\boldsymbol{\epsilon}}_i{}^T$ to estimate the true covariance of responses $\bar{V}_i$ in formula (2.1.3) (Liang and Zeger 1986), denoted as V_{LZ} hereafter. If in the equation (2.1.1), the working covariance matrix V_i is modeled correctly, the corresponding V_R reduces to the model-based version (the naive estimator), $V_M = \mathcal{M}^{-1}$. It is well known that for a general choice V_i, $V_R - V_M$ is a nonnegative definite matrix indicating that the true covariance is the optimal choice for V_i. Hence, when the covariance is correctly specified, $\widehat{V}_M$ is a more reliable estimator for $\mathrm{cov}(\widehat{\boldsymbol{\beta}})$ since the variance of $\widehat{V}_R$ will be larger. In the case of misspecified V_i, $\widehat{V}_M$ is no longer valid as it underestimates $\mathrm{cov}(\widehat{\boldsymbol{\beta}})$.

Despite the overwhelming popularity of GEE procedures, Pepe and Anderson (1994) pointed out there is an important underlying assumption for GEE method. Unless this assumption is satisfied, only working independence structure can give

consistent estimates. The assumption is non-trivial when the covariates vary over time, i.e., the estimating equation (2.1.1) is unbiased for $\boldsymbol{\beta}$ either when working independence covariance should be used or the following conditional expectation must be valid:

$$E\left(Y_{it} \mid X_{it}\right) = E\left(Y_{it} \mid X_{i1}, X_{i2}, \ldots, X_{i\,n_i}\right). \tag{2.1.5}$$

Emond *et al.* (1997) and Pan *et al.* (2000) provided analytic calculations for bias with continuous response data. Pan *et al.* (2000) proved that the bias of GLS estimator is proportional to $\boldsymbol{\beta}$.

2.2 A closer look at sandwich estimator

In GEE models, the more interesting research topic is the variance estimator, namely the "sandwich estimator." The reason why the sandwich estimator V_{LZ} is preferred and widely applied is that it is consistent in terms of misspecification of the working correlation structure and asymptotically normal. These properties are believed to be enough for the inference about variance in most situations. However, it may not be consistent in some cases and generally underestimates the true variance of parameter estimates (Efron 1992 raised the question that delta method tends to underestimate standard errors). Even if it is consistent, the price paid for the consistency is increased variation, that is, the variance of the sandwich estimator is generally larger than the model-based estimate.

2.2.1 The bias of the sandwich estimator

The bias of V_{LZ} is found to be introduced by the approximation of $\bar{V}_i$ by $\widehat{\boldsymbol{\epsilon}}_i \widehat{\boldsymbol{\epsilon}}_i{}^T$, and the bias could be substantial when the sample size is small especially for binary responses (Paik 1988; Sherman and le Cessie 1997; Mancl and DeRouen 2001). Many bias-corrected version of sandwich estimators have been suggested (Mancl and DeRouen 2001, Kauermann and Carroll 2001). Pan (2001) gave a pooled covariance estimator based on the residuals of all the subjects under the assumption of correctly specified variance function and common correlation structures throughout subjects.

Kauermann and Carroll (2001) suggested to substitute the estimated residual $\widehat{\boldsymbol{\epsilon}}_i$ by the leverage adjusted residual $\widetilde{\boldsymbol{\epsilon}}_i = (I - H_{ii})^{-1/2}\widehat{\boldsymbol{\epsilon}}_i$, where I is an identity matrix and H_{ii} is a hat matrix defined later in the next section. Mancl and DeRouen (2001) gave an approximate bias-correction based on the following argument: they considered the first-order expansion of the residual $\widehat{\boldsymbol{\epsilon}}_i$ for $1 \leq i \leq K$,

$$\widehat{\boldsymbol{\epsilon}}_i \approx \boldsymbol{\epsilon}_i + \frac{\partial \boldsymbol{\epsilon}_i}{\partial \boldsymbol{\beta}^T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \boldsymbol{\epsilon}_i - D_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Take square of both sides of the equation above and then take the expectation. Substitute $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ by $\left(\sum_{i=1}^{K} M_i\right)^{-1} \sum_{i=1}^{K} D_i^T V_i^{-1} \boldsymbol{\epsilon}_i$, and omit the higher order of $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ then one can have:

$$\mathbf{E}\left[\widehat{\boldsymbol{\epsilon}}_i \widehat{\boldsymbol{\epsilon}}_i{}^T\right] = (I_i - H_{ii})\ \mathrm{cov}(Y_i)\ (I_i - H_{ii})^T + \sum_{j \neq i} H_{ij}\ \mathrm{cov}(Y_j)\ H_{ij}^T, \qquad (2.2.1)$$

where $H_{ij} = D_i \left(\sum_{l=1}^{K} M_l\right)^{-1} D_j^T V_j^{-1}$ is a matrix of dimension $n_i \times n_j$, I_i is an identity matrix of the same dimension as H_{ii} [note there is a typographical error in expression of (4) in Mancl and DeRouen 2001: the $\mathrm{cov}(y_i)$ in the last term should

be cov(y_j)]. H_{ii} may not be a symmetric matrix unless the working correlation matrix is independent. Mancl and DeRouen (2001) assume that the contribution of the last term to the bias of the sum in expression (2.2.1) is negligible.

We have carried out some simulations to compare the different versions of sandwich estimators: V_{NV}, the naive estimator; V_{LZ}, the first sandwich estimator; V_P, Pan (2001); V_{MD}, Mancl and DeRouen (2001); V_{KC}, Kauermann and Carroll (2001). The performances are compared in terms of relative efficiency, i.e.

$$\mathbf{RE} = \left(\frac{V_X}{V_{true}} - 1\right) \times 100\%,$$

where V_X is any one of the five variance estimates, and V_{true} is obtained from the $\widehat{\boldsymbol{\beta}}$'s estimated from 1000 simulations. Normal and Poisson responses are used (for details of generating such multivariate correlated data please refer to Chapter 4) with different sample sizes K, 40 and 20. The true correlation structure is AR1, and different working correlation structures are used, AR1, EXC (exchangeable) and IND (independent), to see the effect of misspecification of the correlation structure. Furthermore, different values of correlation coefficient are used: 0.1, 0.5 and 0.9 (0.8 for Poisson), to see the effect of the strength of correlation.

Table 2.1 and 2.2 show the results of the five sandwich-formed variance estimators for the slope estimate for normal and Poisson responses respectively (in both types of simulation one intercept and one slope were fitted in the linear predictor). The majority of the values are negative, meaning that most of those variance estimators tend to underestimate the true variance of $\widehat{\boldsymbol{\beta}}$. It is easy to see that, no matter in case of misspecification of correlation structure, or in case of strong correlation, or in case of small sample size, Mancl and DeRouen (2001)'s bias-corrected sandwich estimates outperform all the others, followed by Kauer-

Table 2.1 Relative Efficiency of five sandwich estimators (standard errors), for *normal* responses (in %)

	K=20									K=40								
	$\rho = .1$			$\rho = .5$			$\rho = .9$			$\rho = .1$			$\rho = .5$			$\rho = .9$		
	AR1	EXC	IND	AR1	EXC	IND	AR1	EXC	IND	AR1	EXC	IND	AR1	EXC	IND	AR1	EXC	IND
V_{NV}	-2.8	-5.5	-3.5	-3.9	-19.7	-2.5	-7.5	-26.6	63.9	-1.6	-5.1	-2.6	-2.6	-20.0	-2.5	-2.0	-26.9	63.4
	(.17)	(.18)	(.16)	(.17)	(.16)	(.19)	(.15)	(.09)	(.28)	(.09)	(.09)	(.08)	(.08)	(.08)	(.10)	(.07)	(.04)	(.13)
V_{LZ}	-5.0	-4.6	-4.6	-6.4	-6.0	-6.1	-5.9	-6.3	-6.1	-2.6	-2.3	-2.1	-4.6	-4.5	-4.5	-1.9	-3.1	-3.7
	(.32)	(.32)	(.32)	(.31)	(.32)	(.32)	(.17)	(.17)	(.19)	(.16)	(.16)	(.16)	(.16)	(.16)	(.16)	(.08)	(.09)	(.10)
V_P	-4.8	-4.5	-4.4	-5.8	-5.6	-5.6	-5.1	-6.1	-5.2	-2.5	-2.1	-1.9	-4.0	-4.1	-4.1	-1.0	-2.8	-3.1
	(.29)	(.30)	(.30)	(.29)	(.30)	(.30)	(.14)	(.17)	(.16)	(.15)	(.15)	(.15)	(.14)	(.15)	(.14)	(.07)	(.09)	(.08)
V_{MD}	**0.4**	**0.8**	**0.8**	**-1.1**	**-0.7**	**-0.7**	**-0.7**	**-1.1**	**-0.2**	**0.1**	**0.4**	**0.6**	**-2.0**	**-1.9**	**-1.8**	**0.8**	**-0.5**	**-0.8**
	(.34)	(.34)	(.34)	(.33)	(.34)	(.34)	(.18)	(.18)	(.21)	(.17)	(.17)	(.17)	(.16)	(.17)	(.17)	(.09)	(.09)	(.10)
V_{KC}	-2.4	-2.0	-1.9	-3.8	-3.4	-3.5	–	–	–	-1.2	-0.9	-0.7	-3.3	-3.2	-3.2	–	–	–
	(.33)	(.33)	(.33)	(.32)	(.33)	(.33)	–	–	–	(.16)	(.16)	(.16)	(.16)	(.16)	(.16)	–	–	–

– : the method by Kauermann and Carroll (2001) tends to produce complex values in variance estimates probably due to the small values of $(I - H_{ii})^{-1/2}$ when the correlation is strong.

Table 2.2 Relative Efficiency of five sandwich estimators (standard errors), for *Poisson* responses (in %)

	K=20									K=40								
	$\rho = .1$			$\rho = .5$			$\rho = .8$			$\rho = .1$			$\rho = .5$			$\rho = .8$		
	AR1	EXC	IND	AR1	EXC	IND	AR1	EXC	IND	AR1	EXC	IND	AR1	EXC	IND	AR1	EXC	IND
V_{NV}	-2.0	-5.2	-3.2	-3.5	-17.4	-3.6	-4.5	-20.7	17.8	-3.9	-7.2	-5.2	-1.8	-17.7	-3.8	1.1	-17.3	22.0
	(1.88)	(1.86)	(1.78)	(2.25)	(2.16)	(2.28)	(3.82)	(3.19)	(4.06)	(.87)	(.85)	(.81)	(1.08)	(1.04)	(1.14)	(1.78)	(1.53)	(2.09)
V_{LZ}	-4.9	-5.1	-4.8	-6.2	-5.3	-5.4	-8.3	-8.0	-7.2	-5.4	-5.2	-5.0	-3.2	-3.2	-3.8	-2.1	-1.0	-1.8
	(3.04)	(3.08)	(3.07)	(3.76)	(3.89)	(3.90)	(4.96)	(5.13)	(5.14)	(1.46)	(1.47)	(1.47)	(1.84)	(1.88)	(1.90)	(2.46)	(2.54)	(2.56)
V_P	-4.4	-4.6	-4.4	-5.6	-5.1	-5.0	-7.2	-7.6	-5.5	-5.0	-4.9	-4.6	-2.8	-3.0	-3.5	-1.2	-0.7	-1.1
	(2.92)	(2.96)	(2.95)	(3.51)	(3.76)	(3.70)	(4.63)	(5.02)	(4.82)	(1.35)	(1.37)	(1.36)	(1.69)	(1.81)	(1.78)	(2.33)	(2.48)	(2.41)
V_{MD}	**0.4**	**0.2**	**0.5**	**-0.9**	**-0.1**	**0.0**	**-3.2**	**-3.0**	**-1.8**	**-2.8**	**-2.6**	**-2.4**	**-0.6**	**-0.5**	**-1.1**	**0.5**	**0.6**	**1.0**
	(3.22)	(3.26)	(3.26)	(3.99)	(4.12)	(4.13)	(5.26)	(5.43)	(5.46)	(1.51)	(1.51)	(1.51)	(1.90)	(1.93)	(1.95)	(2.53)	(2.61)	(2.64)
V_{KC}	-2.3	-2.5	-2.2	-3.6	-2.7	-2.8	–	–	–	-4.1	-3.9	-3.7	-1.9	-1.9	-2.5	–	–	–
	(3.13)	(3.17)	(3.16)	(3.87)	(4.00)	(4.01)	–	–	–	(1.48)	(1.49)	(1.49)	(1.87)	(1.91)	(1.92)	–	–	–

– : the method by Kauermann and Carroll (2001) tends to produce complex values in variance estimates probably due to the small values of $(I - H_{ii})^{-1/2}$ when the correlation is strong.

mann and Carroll (2001)'s bias-corrected version. The latter fails to give real valued estimates when the correlation is strong because there are possibly complex values in the calculation of $(I - H_{ii})^{-1/2}$. Pan (2001)'s pooled estimator is slightly better than V_{LZ} in most of the cases but not as good as V_{MD} and V_{KC}. This might be due to the inherent bias of the residual products. All the variance estimators except the naive one are robust to the misspecification of the correlation structure. The naive estimator is efficient when the correlation structure is correctly specified as AR1. But even when the working correlation is the same as the true one, V_{NV} is not as good as V_{MD} and only comparable to V_{KC} in some cases. As for the stability of those sandwich estimators, it is clear to see that

$$\mathrm{Var}(V_{NV}) < \mathrm{Var}(V_P) \leq \mathrm{Var}(V_{LZ}) \leq \mathrm{Var}(V_{KC}) \leq \mathrm{Var}(V_{MD}). \tag{2.2.2}$$

This is consistent with the following theorem (refer to the proof in the appendix) for large K.

Theorem 2.1 *Under mild regularity conditions,* $cov\{vec(M_P)\} - cov\{vec(M_{LZ})\}$, $cov\{vec(M_{LZ})\} - cov\{vec(M_{KC})\}$ *and* $cov\{vec(M_{KC})\} - cov\{vec(M_{MD})\}$ *are non-negative definite with probability tending to 1 as* $K \to \infty$.

Under this theorem, asymptotically we have the relationship (2.2.2). The proof is included in Appendix I for reference. Furthermore, the simulations also show this result seems to hold also for small K.

2.2.2 Another justification for V_{MD}

Below we give another interpretation for the bias-correction by Mancl and DeRouen (2001). This interpretation relies on the commonly accepted results in or-

dinary linear regression and serves as a better understanding and support for V_{MD}. Here we introduce the concept of "generalized hat matrix." Denote the total number of observations $N = \sum_{i=1}^{K} n_i$. Let $\mathcal{Y}$ be the $N \times 1$ response vector, $\widehat{\mathcal{U}}$ be the $N \times 1$ estimated mean vector, $\mathcal{X} = (X_1^T, \ldots, X_K^T)^T$ be the $N \times p$ design matrix, $\mathcal{D} = (D_1^T, \ldots, D_K^T)^T$ be a $N \times p$ matrix, and $\mathcal{V}$ be a $N \times N$ block diagonal matrix whose ith block being V_i, corresponding to the ith subject. Then $\sum_{l=1}^{K} M_l = \mathcal{D}^T \mathcal{V}^{-1} \mathcal{D}$. The iterative algorithm (2.1.2) for obtaining $\widehat{\boldsymbol{\beta}}$ can be rephrased as:

$$\widehat{\boldsymbol{\beta}}_{new} = \widehat{\boldsymbol{\beta}}_{old} + \left(\mathcal{D}^T \mathcal{V}^{-1} \mathcal{D}\right)^{-1} \mathcal{D}^T \mathcal{V}^{-1} (\mathcal{Y} - \widehat{\mathcal{U}}_{old}), \tag{2.2.3}$$

and it can be further expressed as:

$$D\widehat{\boldsymbol{\beta}}_{new} = \mathcal{H}\,[\, D\,\widehat{\boldsymbol{\beta}}_{old} + (\mathcal{Y} - \widehat{\mathcal{U}}_{old})\,], \tag{2.2.4}$$

where $\mathcal{H} = \mathcal{D}\left(\mathcal{D}^T \mathcal{V}^{-1} \mathcal{D}\right)^{-1} \mathcal{D}^T \mathcal{V}^{-1}$, an $N \times N$ asymmetric and idempotent projection matrix. This matrix $\mathcal{H}$ is the analogue of the "hat" matrix in ordinary least squares (OLS). It maps the current value of $\mathcal{Z} = \mathcal{D}\widehat{\boldsymbol{\beta}}_{old} + (\mathcal{Y} - \widehat{\mathcal{U}}_{old})$ into the updated values of $\mathcal{D}\widehat{\boldsymbol{\beta}}_{new}$ (which is actually some transformation of the linear predictor $\mathcal{X}\widehat{\boldsymbol{\beta}}$), i.e.

$$\mathcal{D}\,\widehat{\boldsymbol{\beta}} \approx \mathcal{H}\,\mathcal{Z}. \tag{2.2.5}$$

Most of the properties of hat matrix in OLS are also valid in this scenario: $\mathcal{H}^2 = \mathcal{H}$; $\mathrm{tr}(\mathcal{H}) = p$; the diagonal elements h_{kl} , for $1 \le k, l \le N$, could be interpreted as how much influence or "leverage" exerted on the fitted values by the original response; the average of h_{kl} is p/N. However, different from the hat matrix in OLS, $\mathcal{H}$ here may not be symmetric unless the working correlation matrix is independent in which case the whole procedure reduces to OLS. We call this $\mathcal{H}$ a *generalized hat matrix*, and the hat matrix in OLS is a special case of $\mathcal{H}$ when

the link function is linear link and the correlation structure is independence. The generalized hat matrix $\mathcal{H}$ can be divided into blocks corresponding to different subjects. The leverage of the ith subject is contained in the ith diagonal block, $H_{ii} = D_i \left(\mathcal{D}^T \mathcal{V}^{-1} \mathcal{D}\right)^{-1} D_i^T V_i^{-1}$ of dimension $n_i \times n_i$, and the off-diagonal blocks are $H_{ij} = D_i \left(\mathcal{D}^T \mathcal{V}^{-1} \mathcal{D}\right)^{-1} D_j^T V_j^{-1}$ of dimension $n_i \times n_j$, for $1 \leq i, j \leq K$.

Recall that in OLS, the hat matrix provides nice interpretation of the variance of residuals. In our case, there is such an analogue as well. From equation (2.2.5), it can be obtained that $\widehat{\mathcal{E}} = (I - \mathcal{H})\mathcal{Z}$, where $\widehat{\mathcal{E}} = \mathcal{Y} - \widehat{\mathcal{U}}$ is $N \times 1$ residual vector. Therefore,

$$\mathrm{Var}(\widehat{\mathcal{E}} | \, \widehat{\boldsymbol{\beta}}_{old}) = (I - \mathcal{H}) \, \mathrm{Var}(\mathcal{Y}) \, (I - \mathcal{H})^T, \tag{2.2.6}$$

where I is an $N \times N$ identity matrix and $\mathrm{Var}(\mathcal{Y})$ is the covariance matrix for $N \times 1$ response vector $\mathcal{Y}$ with the variance-covariance matrices of individual subjects being the diagonal blocks V_i and the covariance matrices among subjects being the off-diagonal blocks V_{ij}, for $1 \leq i, j \leq K$ that are assumed to be diagonal matrices because of the independence among subjects under GEE settings. Look into the ith diagonal block of $\mathrm{Var}(\mathcal{Y})$ and $\mathrm{Var}(\widehat{\mathcal{E}})$, one can obtain

$$\mathrm{Var}(\widehat{\boldsymbol{\epsilon}}_i | \, \widehat{\boldsymbol{\beta}}_{old}) = (I - H_{ii}) \mathrm{Var}(Y_i) (I - H_{ii})^T. \tag{2.2.7}$$

From here it is easy to conclude that using the product of residuals to approximately estimate the true variance of responses is biased. The bias could be corrected by substituting $\bar{V}_i$ in formula (2.1.3) by $(I - H_{ii})^{-1} \widehat{\boldsymbol{\epsilon}}_i \, \widehat{\boldsymbol{\epsilon}}_i{}^T (I - H_{ii}^T)^{-1}$, which resulting the bias-corrected version of V_{MD}.

2.2.3 Why resampling ?

In the two versions of bias-corrected sandwich estimators, V_{MD} and V_{KC}, the bias from substituting the residual products for the variance of responses is reduced by using better estimation for the true covariance of the responses. The bias is corrected partially and V_{MD} seems to work quite well in different situations. However, in both bias-corrected sandwich estimators, new problems may be introduced because of the near singular matrix, $(I - H_{ii})$, especially when the sample size is small and the correlation is strong.

An alternative to the biased-corrected variance estimators could be those based on resampling methods. Although asymptotically they do not provide a different parameter estimate from the usual methods, resampling methods such as bootstrap and jackknife have been observed to correct bias in one way or another. More importantly, the resampling methods could generally provide with an estimate for the distribution of the estimated parameters which can serve as a good source for bias correction, variance estimation, confidence interval construction, hypothesis testing and even higher order inferences. Furthermore, resampling methods can have more flexible applicability in the analysis of repeated measurements. Hence, the investigations of resampling methods for longitudinal data analysis are of great interest. There have been a large amount of research done in this area and some of the relevant work have been briefly reviewed in the second part of Chapter 1. In the next few chapters, we explore some new proposals for the application of resampling idea to the GEE procedure in longitudinal data analysis.

Chapter 3

Smooth Bootstrap

A large amount of literature has discussed the application of the classical jackknife and bootstrap methods to longitudinal data. Their performance is limited due to the conventional resampling schemes, and hence the estimators there are strongly model dependent. "EF-based resampling" methods seem to have gained more attention in recent years. The notable characteristics of the method of estimating functions are as follows: first, it only depends on a few features (for example, mean and variance) of the underlying probability model; and second, it is easy to handle nuisance parameters. The less dependence on the model yields standard errors and confidence regions that are less dependent on the model. Hence, resampling methods such as bootstrap and jackknife become the natural candidates for obtaining such standard errors and confidence intervals. Since most of the estimating functions can be expressed as sum of finite terms, those terms naturally become the subject to be resampled.

In GEE procedure for longitudinal data analysis, the estimating function is also in form of sum of finite independent items, therefore we keep our interest

on "EF-based resampling" approaches which are expected to gain more efficiency compared with the classical resampling methods and bootstrap. We mainly deal with the methods for regression parameters and those methods can be extended to deal with nuisance parameter problems where the inference is based on the generalized score function and its bootstrap analog.

To have a better understanding of resampling methods to the correlated longitudinal data, we begin with the idea of bootstrap and the application of EF-based bootstrap to independent data.

3.1 Analytical discussion in independent cases

3.1.1 The idea of Bootstrap

The basic principle underlying bootstrap methods in various settings and in different forms is quite straight forward: it attempts to mimic the relationship between the "population" and the "sample" by the relationship between the "sample" and the appropriately generated "resample". It enables statisticians to avoid dealing with the unknown "population" directly; instead, it utilizes the "resample" and the original "sample" that are known or have known distributions to do statistical inferences.

This method is most applicable in case of independent and identically distributed data. Assume $X = \{X_1, X_2, \ldots, X_n\}$ is an i.i.d. random sample from an unknown probability distribution F and the parameter of interest $\boldsymbol{\theta} = \boldsymbol{\theta}(F) \in \boldsymbol{\Theta}$. The corresponding estimator based on the sample is $\widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}(F_n)$, where F_n is

the sample distribution. The inferences about some characteristic of the sampling distribution of $\widehat{\boldsymbol{\theta}}_n$, such as variance estimation and confidence intervals, depend on the population distribution F that is generally unknown. The idea of bootstrap goes as follows: first from the sample construct an estimator of F, say $\widehat{F}_n$, which can serve as a good representative of F (the most commonly used $\widehat{F}_n$ is the empirical distribution function of the sample, $\widehat{F}_n(u) = n^{-1}\sum_{i=1}^{n} I(X_i \leq u)$, where $I(\cdot)$ is an indicator function); second, draw i.i.d. resamples $X^* = (X_1^*, \ldots, X_n^*)$ from $\widetilde{F}_n$; and third, obtain "bootstrap version" of the estimator $\widehat{\boldsymbol{\theta}}_n$, $\boldsymbol{\theta}_n^*$ by replacing X with X^*, and then the "bootstrap version" of $\boldsymbol{\theta}(F)$ is given by $\boldsymbol{\theta}(\widetilde{F}_n)$. For a reasonable choice of $\widetilde{F}_n$, the bootstrap version accurately mimics those corresponding characteristics of the population and the sample which determine the sampling distribution of parameters of interest. This mimicking nature of the bootstrap method is one of its most salient features.

There are basically two ways to calculate the bootstrap distribution. One is analytical attempts to express explicitly the form of $\boldsymbol{\theta}(\widetilde{F}_n)$. This is limited to a few cases. The other approach is Monte Carlo approximations and is theoretically available for all cases, but requires a fair amount of computing time. Note that the parameter estimated from GEE procedure has no closed form, but it is usually obtained in an iterative manner and so that perturbing the estimation algorithm may greatly simplify the calculations needed in resampling (more in later sections).

3.1.2 Smooth bootstrap for independent data

Most of the estimating functions in statistical models can be expressed as the sum of finite independent or nearly independent items. One can consider these items as

the realizations of some random variable from some common distribution known or unknown so that one can apply the idea of bootstrap to those items, treating them as the original sample from some population. Therefore the inference about the sampling distribution of parameter estimators would become easily based on the sampling distribution of the estimating functions. We start with a simple model for data structure in Table 1.1, assuming independence both within and between subjects, $Y_i = X_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$, for $1 \leq i \leq K$, where $\boldsymbol{\beta}$ is the $p \times 1$ parameter vector of interest, Y_i and $\boldsymbol{\epsilon}_i$ are $n_i \times 1$ vectors, X_i is $n_i \times p$ covariate matrix, or the model can be expressed as $\mathcal{Y} = \mathcal{X}\boldsymbol{\beta} + \mathcal{E}$, the form in Section 2.2.2. Besides the assumption of independence, we further assume that the common variance within subjects, i.e. $\text{cov}(\boldsymbol{\epsilon}_i) = \Sigma_i = \text{diag}\{\sigma_i^2, \ldots, \sigma_i^2\}_{(n_i \times n_i)}$; and we allow the heterogeneity across subjects, i.e. Σ_i might not be equal to Σ_j for $i \neq j$. Then $\text{cov}(\mathcal{E}) = \text{diag}\{\Sigma_1, \ldots, \Sigma_K\}$. The estimation procedure reduces to the ordinary least squares problem and the usual estimator $\widehat{\boldsymbol{\beta}} = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\mathcal{Y}$. The normal estimating equation is:

$$\sum_{i=1}^{K} X_i^T \boldsymbol{\epsilon}_i = 0, \tag{3.1.1}$$

where $\boldsymbol{\epsilon}_i = Y_i - X_i\boldsymbol{\beta}$. Various classical resampling schemes can be applied here, for example, delete-one jackknife, residual bootstrap or paired-bootstrap. Wu (1986) discussed in full details about their performances under independence assumption. One possible EF-based bootstrap is the direct application of Hu and Zidek (1995) to this multivariate data, where $\widehat{\boldsymbol{\beta}}^*$ is obtained from resampling $\{X_i^T\widehat{\boldsymbol{\epsilon}}_i\}_{1 \leq i \leq K}$. Their case can be viewed as the special case with $n_i = n = 1$. Let $z_i = X_i^T\widehat{\boldsymbol{\epsilon}}_i$, then the bootstrap estimates for $\boldsymbol{\beta}$ can be obtained from :

$$\widehat{\boldsymbol{\beta}}^*_{(b)} = \widehat{\boldsymbol{\beta}} + \left(\sum_{i=1}^{K} X_i^T X_i\right)^{-1} \sum_{i=1}^{K} z^*_{i\,(b)}, \tag{3.1.2}$$

for $b = 1, \ldots, B$, where $\{z_i^*\}_{1 \le i \le K}$ being bootstrap resample from $\{z_i\}_{1 \le i \le K}$. More generally, consider the weighted version of the estimating equations:

$$\sum_{i=1}^{K} w_i \, X_i^T \boldsymbol{\epsilon}_i = 0, \tag{3.1.3}$$

where $(w_i)_{i=1}^K$ are weights which are independent of the data. Usually the weights sum up to 1, but since here the right side of the estimating equation is zero, it will not matter whether we have such a restriction or not. Practically it is natural and convenient to choose weights with mean 1 and variance 1 because the first two moments of the weighted estimating functions will remain the same as the original unweighted estimating functions. The performance of such weights are quite satisfactory as seen in later Monte Carlo simulations. Note that the weights $(w_i)_{i=1}^K$ suggested here are generated *independently* of the data. In contrast in some other research work the weights are chosen as data dependent. The main reason for the preference for the random weights is that in widely-used GEE approach for non-normal responses, the estimating functions are usually nonlinear and the inclusion of random weights will not make the statistical inference more difficult (since higher order derivatives may be involved otherwise).

From (3.1.3), the bootstrap-mannered estimating procedure can be carried out:

$$\widehat{\boldsymbol{\beta}}_{(b)}^* = \widehat{\boldsymbol{\beta}} + \left(\sum_{i=1}^{K} w_{i(b)} \, X_i^T X_i \right)^{-1} \sum_{i=1}^{K} w_{i(b)} \, z_i, \tag{3.1.4}$$

therefore, the procedure (3.1.2) is a special case of (3.1.4) when the weights are obtained from a multinomial distribution. We refer to this resampling scheme as "smooth bootstrap," because, in general, the form of (3.1.3) can usually be regarded as a "smoothed" version of bootstrapping when w_i is a continuous random variable such as $\mathbf{N}(1, 1)$ or Exp(1), exponential with unit parameter. The

rationale behind the smooth bootstrap scheme is as follows. The first term on the right-hand side of (3.1.4) reflects the value of the true parameter in the bootstrap space, which is $\widehat{\boldsymbol{\beta}}$, and the second term represents the random fluctuation of the bootstrap replicate around this value, which follows from the perturbation to the estimating functions. A possible extension to the second and third order efficiency approximations can be done in a way discussed in Claeskens et al (2003), in which the authors suggested a second order Taylor approximation of the estimating equation and eventually gave a quadratic form of the bootstrap.

Denote the expectation with respect to the distribution induced by resampling as E_*. The common conditions needed are:

Assumption 3.1 *For the design matrix* $\mathcal{X}$*,*

$$\max_{1\le i\le K, 1\le j\le n_i} x_{ij}^T (\sum_{i=1}^{K} X_i^T X_i)^{-1} x_{ij} \le c/n,$$

for some scalar $c > 0$ *independent of* n*.*

Assumption 3.2 *For the error variances,* $\max_{1\le i\le K} \sigma_i^2 < \infty$.

The consistency of the covariance matrix of $\widehat{\boldsymbol{\beta}}^*$ for this independent multivariate data can be established easily from (3.1.4):

Theorem 3.1

$$E(v_*) = cov(\widehat{\boldsymbol{\beta}})\{1 + O(n^{-1})\},$$

where $v_* = E_*(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})^T$.

This is because

$$(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})^T = \left(\sum_{i=1}^{K} w_{i(b)}\, X_i^T X_i\right)^{-1} \left(\sum_{i=1}^{K} w_{i(b)}^2 X_i^T \widehat{\boldsymbol{\epsilon}}_i \widehat{\boldsymbol{\epsilon}}_i^T X_i\right) \left(\sum_{i=1}^{K} w_{i(b)}\, X_i^T X_i\right)^{-1},$$

and $E(\widehat{\boldsymbol{\epsilon}}_i \widehat{\boldsymbol{\epsilon}}_i^T) = \mathrm{diag}(\boldsymbol{\sigma}_i^2) + O(n^{-1})$.

For the asymptotic normality of $\widehat{\boldsymbol{\beta}}^*$, more assumptions are needed to establish the next theorem.

Assumption 3.3 *The residuals $\boldsymbol{\epsilon}_{ij}$ are independent with distributions F_i having mean 0 and variance $\boldsymbol{\sigma}_i^2$, both F_i and $\boldsymbol{\sigma}_i^2$ being unknown, for $1 \le i \le K$.*

Assumption 3.4 *There exists a positive definite matrix V such that*

$$V_K = K^{-1} \sum_{i=1}^{K} X_i^T X_i \to V \text{ as } K \to \infty.$$

Assumption 3.5 *There exists a positive definite matrix W such that*

$$W_K = K^{-1} \sum_{i=1}^{K} X_i^T \, diag(\boldsymbol{\sigma}_i^2)\, X_i \to W.$$

Assumption 3.6 *x_{ij} and $E(\boldsymbol{\epsilon}_{ij}^4)$ are uniformly bounded for $1 \le i \le K$.*

Theorem 3.2 *For the regression model (3.1.1), suppose assumptions (3.3) - (3.6) hold, then given data $(\mathcal{Y}, \mathcal{X})$,*

$$\sqrt{K}\,(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \sim N_p\,(0, V^{-1} W V^{-1}).$$

Proof. From (3.1.4), we have:

$$\sqrt{K}\,(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) = (\sqrt{K}\, V_{K,w})^{-1} \sum_{i=1}^{K} w_{i(b)}\, z_i,$$

where $V_{K,w} = K^{-1} \sum_{i=1}^{K} w_{i(b)} X_i^T X_i \to V$ as $K \to \infty$ as well.

Let $\boldsymbol{\xi}_i = l^T w_i z_i, 1 \le i \le K$, for any fixed p dimensional vector l with $||l|| = 1$. Conditional on the original sample, $(l^T z_i)_{i=1}^K$ are independent and identically distributed random variables with zero means and common standard deviation $\boldsymbol{\sigma}_K = (l^T W_{K,w} l)^{1/2} = (K^{-1} \sum_{i=1}^{K} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T)^{1/2}$, where $W_{K,w}$ is the weighted version of W_K hence converge to W as well. The Berry-Essen theorem implies that

$$\sup_x \left| P_K \left((\sqrt{K}\, W_{K,w})^{-1} \sum_{i=1}^{K} l^T w_i z_i \le x \right) - \boldsymbol{\Phi}(x) \right| \le A \boldsymbol{\gamma}_K \sqrt{K}^{-1}, \qquad (3.1.5)$$

where $\boldsymbol{\gamma}_K = K^{-1} \sum_{i=1}^{K} |\boldsymbol{\xi}_i|^3 / (K^{-1} \sum_{i=1}^{K} |\boldsymbol{\xi}_i|^2)^{3/2}$. From Assumption (3.5), $K^{-1} \sum_{i=1}^{K} |\boldsymbol{\xi}_i|^2 \to l^T W l > 0$ almost everywhere. Moreover, $n^{-3/2} \sum_{i=1}^{K} |\boldsymbol{\xi}_i|^3 \to 0$ as $n \to \infty$ almost everywhere, due to the assumption (3.6) and strong law of large numbers. Finally, since the inequality (3.1.5) is valid for all l, this completes the proof. □

The results stated above are similar to Hu and Zidek (1995). We do not distinguish random or fixed covariates because these two different cases will lead to essentially equivalent results. Regardless of random or fixed covariates, the results stated in the previous theorems are valid.

In summary, the smooth bootstrapping scheme in (3.1.4) has the following properties:

- the sample covariance of $\widehat{\boldsymbol{\beta}}^*$ is a consistent estimator for $\mathrm{cov}(\widehat{\boldsymbol{\beta}})$; and
- the asymptotic distribution of the bootstrap estimator $\widehat{\boldsymbol{\beta}}^*$ is normal whether the covariates are fixed or random.

3.2 Smooth bootstrap for longitudinal data

Following the argument in the previous section, the smooth bootstrap of estimating function procedure can be applied to more general longitudinal settings allowing correlation within subjects. Note that the weights are added to the subject level, and in longitudinal data subjects are independent of each other. Therefore, the inclusion of correlation structure in each summand will not affect the consistency of such a procedure. In this section, two smooth bootstrap approaches for correlated longitudinal data are suggested (Wang and Li 2005). One approach is "robust" to the misspecification of correlation structure; the other approach is "model-based" counterpart, which is supposed to be more efficient than the robust version when the working correlation is correctly specified. The smooth bootstrap estimators are shown to be more efficient in terms of variance estimation and confidence interval coverage in the later simulation studies.

3.2.1 Robust version of Smooth bootstrap

Recall the GEE procedure for the longitudinal data introduced in Chapter 2,

$$\mathbf{U}(\boldsymbol{\beta}) := \sum_{i=1}^{K} D_i^T(\boldsymbol{\beta})\, V_i^{-1}(\boldsymbol{\theta})\, \boldsymbol{\epsilon}_i(\boldsymbol{\beta}) = 0_{p\times 1} \tag{3.2.1}$$

and the iterative algorithm is given by

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} + \left(\sum_{i=1}^{K} \widehat{D}_i^T\, \widehat{V}_i^{-1} \widehat{D}_i\right)^{-1} \left(\sum_{i=1}^{K} \widehat{D}_i^T\, \widehat{V}_i^{-1} \widehat{\boldsymbol{\epsilon}}_i\right). \tag{3.2.2}$$

We carry out the same smooth bootstrap procedure as in (3.1.3), that is,

$$\sum_{i=1}^{K} w_i\, D_i^T(\boldsymbol{\beta})\, V_i^{-1}(\boldsymbol{\theta})\, \boldsymbol{\epsilon}_i(\boldsymbol{\beta}) = 0_{p\times 1} \tag{3.2.3}$$

and similarly we have the following algorithm

$$\widehat{\boldsymbol{\beta}}^{*(m+1)} = \widehat{\boldsymbol{\beta}}^{*(m)} + \left(\sum_{i=1}^{K} w_i \, \widehat{D}_i^{*T} \, \widehat{V}_i^{*-1} \widehat{D}_i^{*}\right)^{-1} \left(\sum_{i=1}^{K} w_i \, \widehat{D}_i^{*T} \, \widehat{V}_i^{*-1} \widehat{\boldsymbol{\epsilon}}_i^{*}\right), \qquad (3.2.4)$$

where $\widehat{D}_i^*$ and $\widehat{\boldsymbol{\epsilon}}_i^*$ are evaluated at $\widehat{\boldsymbol{\beta}}^{*(m)}$; $\widehat{V}_i^*$ is evaluated at $\widehat{\boldsymbol{\theta}}^* = (\widehat{\boldsymbol{\beta}}^{*(m)}, \widehat{\boldsymbol{\alpha}}^*, \widehat{\boldsymbol{\phi}}^*)^T$ and $\widehat{\boldsymbol{\alpha}}^*$ and $\widehat{\boldsymbol{\phi}}^*$ are respectively the intermediate estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ using $\widehat{\boldsymbol{\beta}}^{*(m)}$, the * in $\widehat{\boldsymbol{\alpha}}^*$ and $\widehat{\boldsymbol{\phi}}^*$ indicating the resampling manner for $\boldsymbol{\beta}$. The weights w_i for $1 \leq i \leq K$ are random values from distribution with mean one and unit variance. The consistency and asymptotic normality of the resulting estimators can be established in a similar way as in the previous section. Considering the cumbersome form of (3.2.4) might make this resampling scheme rather complicated and too computer-intensive, here we present a proposition to realize an easy implementation of such smooth bootstrap idea.

Let $C_i(\boldsymbol{\theta})_{p\times n_i} = D_i^T(\boldsymbol{\beta}) V_i^{-1}(\boldsymbol{\theta})$ in the estimating equation (3.2.1). The following proposition shows that if at each iteration step in (3.2.2) or (3.2.4) the parameter $\boldsymbol{\theta}$ in C_i are fixed at values of $\widehat{\boldsymbol{\theta}} + o_p(1)$, such as $\widehat{\boldsymbol{\theta}}$ or the true values $\boldsymbol{\theta}_0$, the consistency and asymptotic covariance of the resulting $\boldsymbol{\beta}$-estimator will not be affected. This implies that in the estimating equation (3.2.1), if we replace $C(\boldsymbol{\theta})$ with $C(\widehat{\boldsymbol{\theta}})$ or $C(\boldsymbol{\theta}_0)$, the resulting $\boldsymbol{\beta}$-estimators will have the same asymptotic variance. Jiang and Zhang (2001) also claimed a similar result. The proof of the following proposition is included in Appendix I.

Proposition 3.1 *Suppose $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ are respectively the solutions to the unbiased estimating equations $g_1 : \sum_{i=1}^{K} C_i(\boldsymbol{\beta})\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) = 0$ and $g_2 : \sum_{i=1}^{K} C_i(\bar{\boldsymbol{\beta}})\boldsymbol{\epsilon}_i(\boldsymbol{\beta}) = 0$, where $\bar{\boldsymbol{\beta}}$ is also a consistent estimate for $\boldsymbol{\beta}$ in the sense of $\bar{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + o_p(1)$. Therefore, we have $K \cdot cov(\widetilde{\boldsymbol{\beta}}) = K \cdot cov(\widehat{\boldsymbol{\beta}}) + o(1) = V_R + o(1)$.*

Given the proposition, we can rephrase our iterative steps in (3.2.4) as follows:

$$\widetilde{\boldsymbol{\beta}}_R^{(m+1)} = \widetilde{\boldsymbol{\beta}}_R^{(m)} + \left(\sum_{i=1}^{K} w_i \widehat{M}_i\right)^{-1} \left(\sum_{i=1}^{K} w_i D_i^T(\widetilde{\boldsymbol{\beta}}_R^{(m)}) V_i^{-1}(\widetilde{\boldsymbol{\beta}}_R^{(m)}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}}) \boldsymbol{\epsilon}_i(\widetilde{\boldsymbol{\beta}}_R^{(m)})\right) \tag{3.2.5}$$

where we fix the parameters in M_i. If we begin the iteration with a consistent parameter estimate, for example, the GEE estimator $\widehat{\boldsymbol{\beta}}$, assuming the resulting $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\phi}}$ are as well $\sqrt{K}$ consistent, we can fix these nuisance parameter as well. Hence we will focus on making replicates of $\widehat{\boldsymbol{\beta}}$ by giving random weights in (3.2.5).

The motivation of this smooth bootstrap procedure can be presented in another way. Suppose that $F(u)$ is the distribution of $U(\boldsymbol{\beta}_0)$. Parzen, Wei and Ying (1994) considered the case when $F(u)$ is pivotal and suggested to obtain "bootstrapped" copies of $\widehat{\boldsymbol{\beta}}$, $\boldsymbol{\beta}^*$, by solving $U(Y;\boldsymbol{\beta}) = R$, where R is generated from the distribution $F(u)$. Parzen et al (1994) showed that the conditional distribution of $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$ is asymptotically identical to the distribution of $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$. This resampling approach relies on the key fact that independent realizations of R are available. Therefore, the pivotal assumption here is vital.

In many cases, the distribution of $U(\boldsymbol{\beta}_0)$ may not be known but may well depend on both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. To this end, we suggest the following stochastic realizations for $U(\boldsymbol{\beta})$, $\sum_{i=1}^{K} w_i \widehat{C}_i \widehat{\boldsymbol{\epsilon}}_i$, in which $w_1, w_2, ..., w_K$ are independent realizations from a distribution with arbitrary mean but unit variance, and the hat symbol $\widehat{\ }$ indicates evaluation at $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}})$.

This may suggest us to use $\sum_{i=1}^{K} w_i \widehat{C}_i \widehat{\boldsymbol{\epsilon}}_i$ as the random realizations of $F(u)$:

$$U(\boldsymbol{\beta}) = 0 \leftarrow \sum_{i=1}^{K} w_i \widehat{C}_i \widehat{\boldsymbol{\epsilon}}_i \tag{3.2.6}$$

Further, we are motivated to solve the following perturbed GEE version of $U(\boldsymbol{\beta})$ conditional on $\widehat{\boldsymbol{\beta}}$ to produce smooth bootstrap estimates of $\widehat{\boldsymbol{\beta}}$ which can be produced in an iterative way,

$$\widetilde{U}(\boldsymbol{\beta}) := \sum_{i=1}^{K} w_i \, \widehat{C}_i \, \boldsymbol{\epsilon}_i(\boldsymbol{\beta}) = 0 \tag{3.2.7}$$

Note that $\widehat{C}_i$ is evaluated at the known estimates $\widehat{\boldsymbol{\theta}}$. By the proposition introduced above, the asymptotic covariance of the resulting $\boldsymbol{\beta}$-estimator is not affected if $\widehat{C}_i$ is replaced by C_i as in (3.2.7). For example, the modified version of $\widetilde{U}(\boldsymbol{\beta})$,

$$\sum_{i=1}^{K} w_i \, D_i^T(\boldsymbol{\beta}) \, V_i^{-1}(\boldsymbol{\beta}; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}}) \, \boldsymbol{\epsilon}_i(\boldsymbol{\beta}),$$

is asymptotically equivalent to $\widetilde{U}(\boldsymbol{\beta})$. This can lead to the iterative steps as in (3.2.5).

More specifically, the iterative steps go as follows:

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}_R}^{(0)} &= \widehat{\boldsymbol{\beta}} \\
\widetilde{\boldsymbol{\beta}_R}^{(1)} &= \widetilde{\boldsymbol{\beta}_R}^{(0)} + \left(\sum_{i=1}^{K} w_i \, \widehat{M}_i \right)^{-1} \left(\sum_{i=1}^{K} w_i \, D_i^T(\widetilde{\boldsymbol{\beta}_R}^{(0)}) \, V_i^{-1}(\widetilde{\boldsymbol{\beta}_R}^{(0)}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}}) \, \boldsymbol{\epsilon}_i(\widetilde{\boldsymbol{\beta}_R}^{(0)}) \right) \\
\widetilde{\boldsymbol{\beta}_R}^{(2)} &= \widetilde{\boldsymbol{\beta}_R}^{(1)} + \left(\sum_{i=1}^{K} w_i \, \widehat{M}_i \right)^{-1} \left(\sum_{i=1}^{K} w_i \, D_i^T(\widetilde{\boldsymbol{\beta}_R}^{(1)}) \, V_i^{-1}(\widetilde{\boldsymbol{\beta}_R}^{(1)}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}}) \, \boldsymbol{\epsilon}_i(\widetilde{\boldsymbol{\beta}_R}^{(1)}) \right) \\
&\quad ...
\end{aligned}$$

Note that $\widetilde{\boldsymbol{\beta}_R}^{(l)}$, for $l = 1, 2, ...$, are all asymptotically equivalent. It is easy to see that $\widetilde{\boldsymbol{\beta}_R}^{(l)} \to \widehat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$ given the initial value $\widehat{\boldsymbol{\beta}}$ being the consistent GEE parameter estimate and $\sqrt{K}$ consistency of $\widehat{\alpha}$ and $\widehat{\boldsymbol{\phi}}$. For any l, the covariance of the estimates from the perturbing method conditional on the observed data is given by

$$\widehat{\text{var}}(\widetilde{\boldsymbol{\beta}_R}^{(l)}) = \left(\sum_i \widehat{M}_i \right)^{-1} \left(\sum_i (\widehat{C}_i \, \widehat{\boldsymbol{\epsilon}}_i \, \widehat{\boldsymbol{\epsilon}}_i^T \, \widehat{C}_i^T) \right) \left(\sum_i \widehat{M}_i \right)^{-1}, \tag{3.2.8}$$

which is essentially the estimated var($\widehat{\boldsymbol{\beta}}$) by plugging in $\widehat{\boldsymbol{\beta}}$. As we can see, this smooth bootstrap estimator can be easily obtained without solving any extra estimating equations compared with that in Parzen et al (1994), each time generating a random value from the pivotal distribution one new estimating equation has to be solved.

Theorem 3.3 *Under mild regularity conditions, conditional on the data and given a consistent GEE estimate* $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\alpha}}$ *and* $\widehat{\boldsymbol{\phi}}$, *the smooth bootstrap estimates from the resampling procedure (3.2.5) satisfy*

$$\sqrt{K}\,(\widetilde{\boldsymbol{\beta}_R} - \widehat{\boldsymbol{\beta}}) \sim N_p\,(0, \mathcal{M}^{-1}\,\mathcal{V}_U\,\mathcal{M}^{-1}),$$

where $\mathcal{M}$ and $\mathcal{V}_U$ are as defined in Chapter 2. The proof of this theorem closely resembles Theorem 3.2 with the exception that there is correlation matrix involved in V.

Below are remarks about this smooth bootstrap method.

Remark 1. It might be expected that the estimator can be improved even further by iterating the above procedure. However, our experience based on simulation studies indicates that further iterations do not result in significant difference. Therefore, we recommend using $\widetilde{\boldsymbol{\beta}_R}^{(1)}$ or $\widetilde{\boldsymbol{\beta}_R}^{(2)}$ as the smooth bootstrap estimator of $\widehat{\boldsymbol{\beta}}$.

Remark 2. During the iterative steps we have kept the information matrix fixed at the initial consistent parameter estimates. The information matrix can be updated accordingly at each step. The reason for not updating the information matrix at each step is that the computation time will be greatly reduced without

affecting the consistency and asymptotic variance. Furthermore, when the sample size is small, perturbing the information matrix may result in unstable variance estimates because the perturbed information matrix may become close to singular. Hence, in such cases, caution needs to be taken when applying the smooth bootstrap methods.

Remark 3. The proposed method is robust to the misspecification of the correlation structure given that mean function is correctly specified and the variance function is correct or at least can be efficiently estimated. The "robust" is not in the sense of outlier-resistance.

Remark 4. This perturbation method has a close link to the classical bootstrap method. If we let $(w_1, w_2, ..., w_K)$ follow a multinomial distribution with the population parameter K and probability vector $(1/K, 1/K, ..., 1/K)$, the above resampling method will coincide with the traditional bootstrap method. In this case w_i has mean 1 and variance $1 - 1/K$, close to 1 when K is large. Moulton and Zeger (1989) applied this idea for repeated measures extended from the GLM setting.

Remark 5. One might want to use covariate-dependent weights or even parameter-dependent weights. However, the asymptotic covariance matrix of the estimator will be the same as the one we suggest here. Thus, the complication may not result in a real gain in efficiency from this point of view.

3.2.2 Model-based version of smooth bootstrap

Just like the sandwich estimator, the perturbation method introduced above is appropriate even when V_i is misspecified. In the case when V_i is correctly specified, we expect to have a more efficient counterpart corresponding to the model-based estimator $\widehat{V}_M$. *This option of variance estimator is hardly seen in existing research.* We now introduce another resampling procedure to provide such option when $V_i = \bar{V}_i$.

Consider the Cholesky decomposition of the variance matrix. Let $\mathbf{\Gamma}_i(\boldsymbol{\theta}) = V_i^{-1/2}(\boldsymbol{\theta})$ (a lower triangular matrix) such that $V_i^{-1}(\boldsymbol{\theta}) = \mathbf{\Gamma}_i^T(\boldsymbol{\theta})\,\mathbf{\Gamma}_i(\boldsymbol{\theta})$, $\widehat{\mathbf{\Gamma}}_i = \mathbf{\Gamma}_i(\widehat{\boldsymbol{\theta}})$, $Z_i(\boldsymbol{\beta};\boldsymbol{\alpha}_0,\boldsymbol{\phi}_0) = \mathbf{\Gamma}_i(\boldsymbol{\beta};\boldsymbol{\alpha}_0,\boldsymbol{\phi}_0)\,\boldsymbol{\epsilon}_i(\boldsymbol{\beta})$. The idea here is to make the residuals of one subject, $\boldsymbol{\epsilon}_{ij}, 1 \le j \le n_i$, "independent" of each other so that an analogue of residual bootstrap in i.i.d. case could be done. The correct covariance model is the key to the consistency of this approach. Therefore, when V_i is correctly specified, $\mathrm{var}(Z_i) = \boldsymbol{\phi}\, I_{n_i}$, where I_{n_i} is an $n_i \times n_i$ identity matrix, we define

$$\widetilde{U}_M(\boldsymbol{\beta}) : = \sum_{i=1}^{K} \widehat{D}_i^T\, \widehat{\mathbf{\Gamma}}_i^T\, W_i\, Z_i(\boldsymbol{\beta};\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\phi}}), \tag{3.2.9}$$

where W_i is a $n_i \times n_i$ diagonal matrix with diagonal elements $\{w_{ii}\}_1^{n_i}$ that are independent realizations from a distribution with mean 1 and variance 1. Here the perturbation to the i-th subject is based on a weighted sum of n_i random variables $\{w_{ii}\}_1^{n_i}$, while in (3.2.7) the subject i is perturbed only by one random number w_i.

Again, the iterative procedure can be carried out as follows

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}_M}^{(0)} &= \widehat{\boldsymbol{\beta}} \\
\widetilde{\boldsymbol{\beta}_M}^{(1)} &= \widetilde{\boldsymbol{\beta}_M}^{(0)} + \left(\sum_{i=1}^{K} \widehat{D}_i^T \widehat{\boldsymbol{\Gamma}}_i^T W_i \widehat{\boldsymbol{\Gamma}}_i \widehat{D}_i\right)^{-1} \left(\sum_{i=1}^{K} \widehat{D}_i^T \widehat{\boldsymbol{\Gamma}}_i^T W_i Z_i(\widetilde{\boldsymbol{\beta}_M}^{(0)}; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}})\right) \\
\widetilde{\boldsymbol{\beta}_M}^{(2)} &= \widetilde{\boldsymbol{\beta}_M}^{(1)} + \left(\sum_{i=1}^{K} \widehat{D}_i^T \widehat{\boldsymbol{\Gamma}}_i^T W_i \widehat{\boldsymbol{\Gamma}}_i \widehat{D}_i\right)^{-1} \left(\sum_{i=1}^{K} \widehat{D}_i^T \widehat{\boldsymbol{\Gamma}}_i^T W_i Z_i(\widetilde{\boldsymbol{\beta}_M}^{(1)}; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}})\right) \\
&\ \ \ ...
\end{aligned}$$

To further simplify the computing, there could be only Z_i that updates at each iterative step without any effect to the consistency. Again, $\widetilde{\boldsymbol{\beta}_M}^{(l)}$, for $l = 1, 2, ...$, are all asymptotically equivalent. When the correlation is correctly specified, we have $\widetilde{\boldsymbol{\beta}_M}^{(l)} \to \widehat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$. It can be seen that, for any l, the covariance of the estimates from the perturbing method conditional on the observed data is given by

$$\widehat{\text{var}}(\widetilde{\boldsymbol{\beta}_M}^{(l)}) = \left(\sum_{i=1}^{K} \widehat{M}_i\right)^{-1} \left(\sum_{i=1}^{K} (\widehat{D}_i^T \widehat{\boldsymbol{\Gamma}}_i^T \,\text{diag}\,(\widehat{\boldsymbol{\Gamma}}_i \widehat{\boldsymbol{\epsilon}}_i \widehat{\boldsymbol{\epsilon}}_i^T \widehat{\boldsymbol{\Gamma}}_i^T)\, \widehat{\boldsymbol{\Gamma}}_i \widehat{D}_i)\right) \left(\sum_{i=1}^{K} \widehat{M}_i\right)^{-1}. \tag{3.2.10}$$

When V_i is correctly specified, $E[\widehat{\text{var}}(\widetilde{\boldsymbol{\beta}_R}^{(l)})] = E[\widehat{\text{var}}(\widetilde{\boldsymbol{\beta}_M}^{(l)})] = \mathcal{M}^{-1}$, and the variation of $\widehat{\text{var}}(\widetilde{\boldsymbol{\beta}_M}^{(l)})$ is expected to be smaller than that of $\widehat{\text{var}}(\widetilde{\boldsymbol{\beta}_R}^{(l)})$, which is verified in the simulation studies. From this point on, the procedures for $\widetilde{\boldsymbol{\beta}_R}^{(l)}$ and $\widetilde{\boldsymbol{\beta}_M}^{(l)}$ will be referred to as "Robust Resampling" (RR) and "Model-based Resampling" (MR) procedures. For MR procedures, similar remarks can be given as for RR procedure.

To summarize, the smooth bootstrap approaches capture the the first order and second order moments. Intuitively, higher order moments such as skewness can also be captured if appropriate w_i can be chosen. Hu and Zidek (1995) established the consistency of the third and fourth moments in the linear model for

independent data. In the context of time series, Lele (1991a) considered how such random sequences may be chosen. Future research along this direction is of great interest. There are more aspects which tend to affect the performance of perturbation by adding weights. One must be cautious when introducing random weights. Usually borrowing information outside the sample may risk robustness. Although the consistency of the estimator's moments can be assured, it may possibly be achieved at the expense of the estimation accuracy for overall distribution. Particularly when K is not sufficiently large, the distribution of weighted estimating function will be sensitive to the choice of the distribution for the weights. Thus, data-dependent weights in the application of GEE models will be an interesting subject for future research.

Chapter 4

Simulation studies for smooth bootstrap

4.1 Correlated data generation

The generation of correlated random variates from multivariate distributions has gained little interest. However, it is important to generate appropriate correlated data for Monte Carlo simulations to investigate statistical methods for the analysis of correlated data. In this section different approaches for generating correlated data, namely normal, overdispersed Poisson and lognormal, will be discussed briefly. More information is included in the Appendix II.

Considering the characteristics of longitudinal data, in most of the cases in our simulation studies, the covariates x_{ij} for $1 \leq i \leq K$ and $1 \leq j \leq n_i$ are generated from a uniform distribution on $(j\,, j+1)$ to make them have some time trend. K is the number of subjects, which were chosen to be 20 and 40 (the corresponding

simulation and bootstrap times in the simulation studies are 1000 for sample size 20 and 500 for sample size 40 respectively, considering smaller variation for larger sample size). n_i is the number of observations for the ith subject. Both common and different n_i's are considered as balanced and unbalanced structure of longitudinal data. For balanced data, the number of observations per subject is $n_i = n = 4$. For unbalanced data, the number of observations per subject varies as follows: the proportions for $n_i = 2$, 3, 4, 5, and 6 are 20%, 30%, 20%, 15% and 15% respectively.

4.1.1 Correlated normal data

The normal data are generated from the following model, $y_{ij} = \boldsymbol{\mu}_{ij} + \boldsymbol{\epsilon}_{ij}\,, 1 \leq i \leq K, 1 \leq j \leq n_i\,$, where identity link is used, $\boldsymbol{\mu}_{ij} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{ij}\,$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)^T = (0,1)^T$. For each subject i, $(\boldsymbol{\epsilon}_{i1}, \boldsymbol{\epsilon}_{i2}, ..., \boldsymbol{\epsilon}_{in_i})$ are correlated with a first order autoregressive structure (AR(1)), correlation coefficient being 0.5.

4.1.2 Correlated lognormal data

To generate non-normal data with long tail distributions, we use the approach provided in Wang and Carey (2004) for generating correlated lognormal data with given mean and covariance. Log link is used, $\boldsymbol{\mu}_{ij} = \exp(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{ij})$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)^T = (0,1)^T$. Gamma variance function is used for this type of data in the simulation, i.e., variance matrix being a diagonal matrix diag($\boldsymbol{\mu}_i^2$). AR(1) correlation structure is used with correlation coefficient 0.5.

4.1.3 Overdispersed Poisson data

For the overdispersed Poisson data, three methods with different variance functions are provided here. The relevant R code and justification are given in Appendix II.

Method 1): AR(1) Poisson with $\mathrm{Var}(\boldsymbol{\mu}) = \boldsymbol{\mu} + \boldsymbol{\lambda}\boldsymbol{\mu}^2$.

y_{ij} is sampled from a Poisson distribution with mean $\boldsymbol{\mu}_{ij}\,\boldsymbol{\xi}_{ij}$, where $\boldsymbol{\mu}_{ij} = \exp(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{ij})$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)^T = (0,1)^T$. Here x_{ij} is from uniform distribution on $(j\,, j+1)$ and then log-transformed, and $\boldsymbol{\xi}_{ij}$ is a gamma random variable with mean 1 and variance λ. In our case we chose $\lambda = 0.5$ so that $\mathrm{var}(y_{ij}) = \boldsymbol{\mu}_{ij} + 0.5\,\boldsymbol{\mu}_{ij}^2$. Within subject correlations are introduced through $\boldsymbol{\xi}_{ij}$ by allowing $\boldsymbol{\xi}_{ij}$ to be correlated for $1 \le j \le n_i$, correlation coefficient being 0.3. The induced covariance of Y_i is $V_i = \mathrm{diag}\,(\boldsymbol{\mu}_i) + \mathrm{diag}\,(\boldsymbol{\mu}_i)\,\mathrm{cov}\,(\boldsymbol{\xi}_i)\,\mathrm{diag}\,(\boldsymbol{\mu}_i)$, from which we can determine the required correlation matrix for $\boldsymbol{\xi}_i$, $R_\xi = (r_{kl})$. The correlated gamma variables with the desired correlation matrix R_0 can be easily generated from $\boldsymbol{\xi}_{ij} = \sum_{l=1}^{d} z_{lj}^2 \,/\, d$, where $d = 2/\boldsymbol{\lambda} = 4$ and each of the four n_i-variate $(z_{l\cdot}), 1 \le l \le d$ are independently sampled from $\mathrm{N}(\mathbf{0}, R_0)$, where $R_0 = (\bar{r}_{kl}) = (\sqrt{r_{kl}})$ (Henderson and Shimakura, 2003; Davis, Dunsmuir and Wang, 2000).

Method 2): AR(1) Poisson with $\mathrm{Var}(\boldsymbol{\mu}) = \boldsymbol{\mu}/(1-\boldsymbol{\rho}^2)$.

For subject i, generate an initial observation y_{i0} is generated from a Poisson distribution with mean $\boldsymbol{\mu}_{i1}$ and variance $\boldsymbol{\mu}_{i1}$. Then the observations $y_{ij}\,, 1 \le j \le n_i$ are generated from Poisson distribution with parameter $\boldsymbol{\mu}_{i(j-1)} + \boldsymbol{\rho}\sqrt{\dfrac{\boldsymbol{\mu}_{ij}}{\boldsymbol{\mu}_{i(j-1)}}}(y_{i(j-1)} - \boldsymbol{\mu}_{i(j-1)})$,

so that $E(y_{ij}) = \boldsymbol{\mu}_{ij}$, $\text{Var}(y_{ij}) = \dfrac{\boldsymbol{\mu}_{ij}}{1-\boldsymbol{\rho}^2}$, and $\text{Cor}(y_{ij}, y_{il}) = \boldsymbol{\rho}^{|j-l|}$. The mean vector $\boldsymbol{\mu}_i$ is generated in the same way as in the first method.

Method 3): AR(1) Poisson with $\text{Var}(\boldsymbol{\mu}) = a\boldsymbol{\mu} + b\boldsymbol{\mu}^2$.

This is a more general approach to generate AR(1) Poisson data using gamma frailty model. For subject i, the first observation y_{i1} is generated from a Poisson distribution with parameter $\boldsymbol{\mu}_{i1} * t_1$, where t_1 is a random value from a gamma(a_1, a_1) and $a_1 = \dfrac{\boldsymbol{\mu}_{i1}}{(a-1) + b\boldsymbol{\mu}_{i1}}$. The following observations y_{ij} for $2 \le j \le n_i$ are generated from Poisson distribution with parameter $\boldsymbol{\eta}_j * t_j$, where $\boldsymbol{\eta}_j = \boldsymbol{\mu}_{ij} + \boldsymbol{\rho}\sqrt{\dfrac{a\boldsymbol{\mu}_{ij} + b\boldsymbol{\mu}_{ij}^2}{a\boldsymbol{\mu}_{i(j-1)} + b\boldsymbol{\mu}_{i(j-1)}^2}}\left(y_{i(j-1)} - \boldsymbol{\mu}_{i(j-1)}\right)$, and $t_j = \dfrac{\boldsymbol{\mu}_{ij}^2 + \boldsymbol{\rho}^2(a\boldsymbol{\mu}_{ij} + b\boldsymbol{\mu}_{ij}^2)}{(a\boldsymbol{\mu}_{ij} + b\boldsymbol{\mu}_{ij}^2)(1-\boldsymbol{\rho}^2) - \boldsymbol{\mu}_{ij}}$. Therefore, we have $\mathbf{E}(y_{ij}) = \boldsymbol{\mu}_{ij}$, $\text{Var}(y_{ij}) = a\boldsymbol{\mu}_{ij} + b\boldsymbol{\mu}_{ij}^2$ and $\text{Cor}(y_{ij}, y_{il}) = \boldsymbol{\rho}^{|j-l|}$. The two constants a and b in the variance function are not arbitrary, $a > 0$ and $1 - 1/(a + b\boldsymbol{\mu}) > \rho^2$ for all $\boldsymbol{\mu}$.

4.2 Simulation Results

The objectives of our simulation studies are two-fold: first, to investigate the consistency of the variance estimators (in terms of the relative efficiency of estimated standard errors against the empirical true standard deviation of $\widehat{\boldsymbol{\beta}}$ obtained from the simulation); and second, to evaluate the accuracy of 80% and 95% confidence interval coverage probabilities.

We use random weights from different distributions to see the performances of proposed resampling methods when using different weights. The RR procedure using $Normal(1, 1)$ random values is denoted as "RR-norm" and the MR

procedure using $Normal(1,1)$ is denoted as "MR-norm." The same notation rule applies to the random values from other distributions. The following table lists the distributions from which come all the random values under investigation. The first four cumulants might be relevant to the performances and are listed. The first two cumulants are all 1.

Table 4.1: Different distributions to generate weights in simulation studies

Dist.	Mean	Var	Skewness	Kurtosis	Has negative value?
Norm(1,1)	1	1	0	0	Y
Double Exp$(1,\sqrt{2})$	1	1	0	3	Y
Unif$(1-\sqrt{3},1+\sqrt{3})$	1	1	0	-1.2	Y
Gamma(4,2)-1	1	1	1	1.5	Y
Pois(1)	1	1	1	1	N
Exp(1)	1	1	2	6	N
Chisq(2)/2*	1	1	2	6	N

Note *: Distribution chisq(2)/2 is essentially Exp(1).

The simulation studies compare the performances of RR and MR procedures, with commonly used variance estimators (V_{NV}, V_{LZ} and V_{MD}) and classical resampling variance estimators: the "approximate jackknife" estimator (denoted as AJ) improved by Yan and Fine (2004); and the "All Block Bootstrap" estimator by Sherman and le Cessie (1997) (denoted as BB).

4.2.1 Consistency of variance estimates

The results for examining the consistency of variance estimates are presented in side-by-side boxplots, in terms of the relative efficiency (ratios) of the standard errors from different methods to the "true" standard deviation of $\widehat{\boldsymbol{\beta}}$, which was obtained from the simulations. Generally, the relative performances of those competitors are similar in different simulation studies. We will discuss the results for normal responses first, followed by additional comments for the simulation

results for lognormal and Poisson responses and ending with a brief summary of the highlights.

Figure 4.1 and 4.2 show the results for normal responses with sample size $K = 40$ and $K = 20$ respectively. In each figure, the left three panels are for the balanced data, and the right three panels are for the unbalanced data (refer to the beginning of the previous section for the details about n_i). In either panel, three individual graphs represent different working correlation structures, AR(1), exchangeable and independence, where the true correlation structure of the data is AR(1) with correlation coefficient .5. Boxplots have the advantage of showing the main characteristics of a distribution, such as median which tells us the central tendency, interquartile range (IQR) which is a good measure of variation, and the whiskers and outliers which can tell us the skewness and the tail of the distribution. We also add the mean value to the boxplot with a "+" sign. Comparing Figure 4.1 and 4.2, the IQRs of those boxplots increase as the sample size decreases, which is as expected. Similarly, an increase of the variation of those standard deviation estimators can be seen comparing the results for balanced data with those for unbalanced data.

In each of the six plots in Figure 4.1, there are three groups of boxplots. The first group consists of five estimators: the naive estimator, LZ, MD and the two classical resampling estimators, AJ and BB. The second group of boxplots display the results for RR estimators suggested in Section 3.2, using different random variates to perturb the estimating algorithm, in the order of RR-norm, RR-de, RR-unif, RR-gamma, RR-pois and RR-exp. The last group consists of the corresponding MR counterparts following the same order. The horizontal lines are

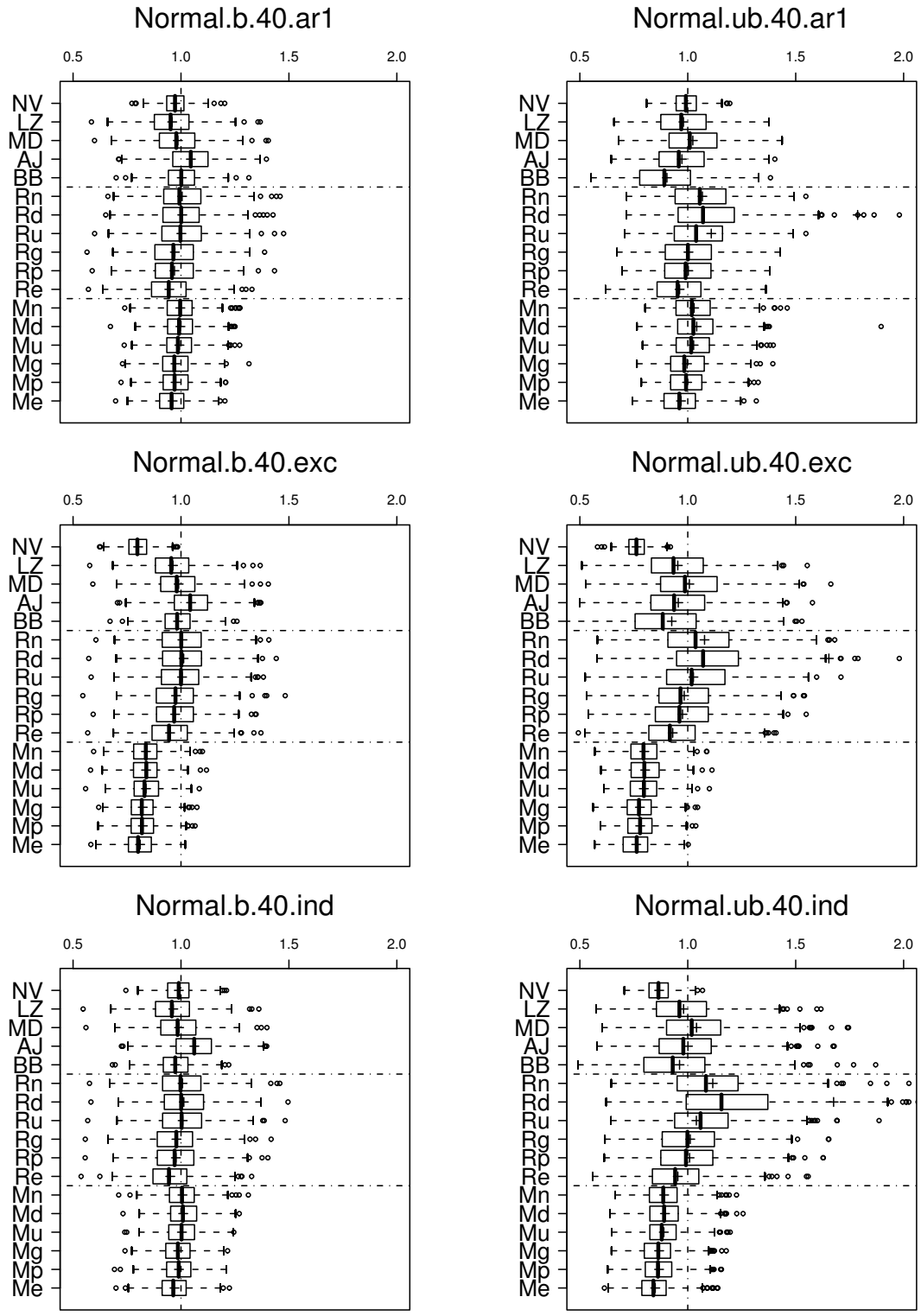

Figure 4.1: Relative efficiency of std. dev. estimates for normal data, K=40, sim=500, boot=500

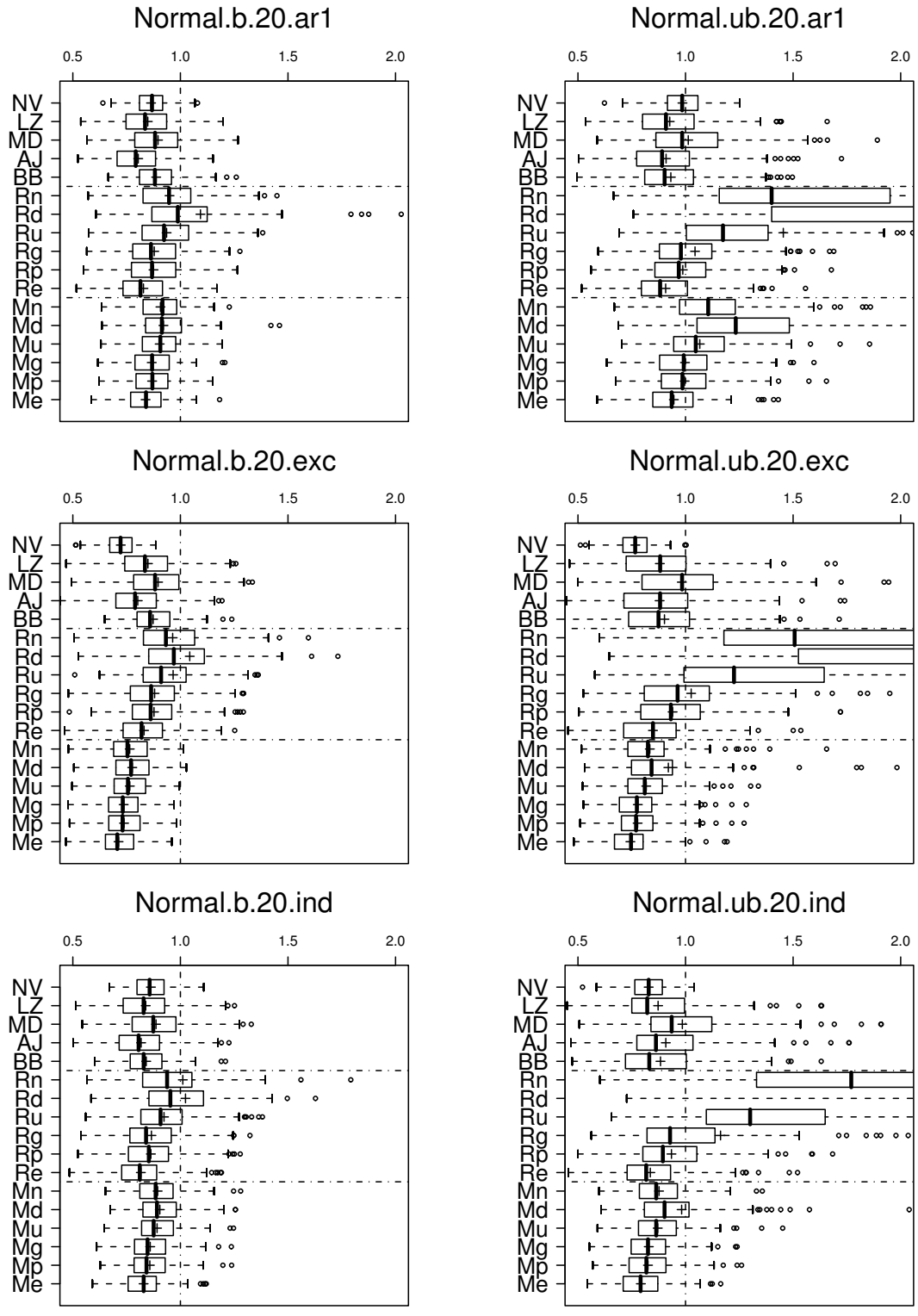

Figure 4.2: Relative efficiency of std. dev. estimates for normal data, K=20, sim=1000, boot=1000

used to separate the three groups for clearer presentation. As we can see, most of the estimators underestimate var($\hat{\boldsymbol{\beta}}$) since most of the median lines are below the vertical line "ratio=1", which represents the same value as the true standard deviation on that line. Focusing on one single plot, "Normal.b.40.ar1." It is quite obvious that the median lines of the first three RR and the first three MR estimators nearly touch the vertical line, meaning that they give very accurate estimation for the variance of parameter estimates. As for the other three types of random variates, gamma and Poisson seem to perform similarly in RR and MR procedures. RR/MR-gamma and RR/MR-pois are less accurate than the previous three, namely normal, double exponential and uniform, but more accurate than RR/MR-exp respectively. All the RR estimators have larger variations than the MR estimators which confirms our previous discussion. ABB approach gives accurate estimates in this case as well; while AJS approach tends to overestimate the standard deviation of the parameter estimates. As for the "sandwich"-type estimators, the naive estimator has the smallest variation and is consistent in this case, because the correlation model is correctly specified; the LZ sandwich estimator underestimates the standard deviation obviously; the MD estimator corrects the bias of LZ quite a lot, but is still outperformed by some of the RR and MR procedures, for example RR-norm. The variations of LZ and MD estimators are similar to the RR estimators and AJS estimator. The variation of ABB estimator is a little smaller, but still larger than that of the naive estimator. Similar results can be seen in "Normal.b.40.exc" and "Normal.b.40.ind" (the other two small plots in the top panel in Figure 4.1) with the exception of the performances of the naive estimator and the MR estimators. When the working correlation is EXC, the naive and MR estimators are no longer valid although they have smaller

variations. In the case of working independence assumption, i.e. Rw=IND, the naive and MR procedures are still producing very efficient estimates. This may due to the moderate correlation strength ($\boldsymbol{\rho} = 0.5$). One additional important result in these two plots is the robustness to the misspecification of the correlation structure for the LZ, MD, AJS, ABB and the RR estimators.

Looking to the right panel in Figure 4.1, in this case of unbalanced data, the ABB approach fails to yield efficient estimates and has similar large variation as the LZ, MD, AJS and the RR estimators. The AJS estimates underestimate the standard deviation this time. The MD bias-corrected estimator is still quite accurate, and the LZ is still not very accurate. The first three RR estimators tend to overestimate a bit for this unbalanced case. RR-gamma and RR-pois are very accurate in terms of variance estimation because their median lines and the mean values are among the closest to the vertical line. Another distinct result can be seen for this unbalanced data compared with those for balanced data. When the working correlation is independent, the naive and MR estimators are not as consistent as in the balanced case while the RR estimators are still valid.

In the case of small sample size presented in Figure 4.2, the superiority of the RR procedures become more obvious. In the left three plots for the balanced data, RR-norm, RR-de and RR-unif yield the most accurate standard deviation estimates among all. In case of correct working correlation, MR-norm, MR-de and MR-unif also outperform the "sandwich"-type estimators and the classical resampling estimators. The MD estimators are only comparable to RR-gamma and RR-pois, but less accurate than the first three RR estimators. On the right panel, the best estimators become RR-gamma, RR-pois and MD, since the RR

procedures as well as some MR procedures using the first three types of random variates tend to have much larger variations. The most likely reason for this is that for these three distributions, negative random values are likely to be drawn. When the sample size is small and the data is unbalanced (in our case the smallest number of observation per subject is only 2), those negative values would result in very unstable estimation when making the bootstrap copies of parameter estimates and leads to even inconsistent estimation for the variance. The gamma distribution we used also has negative values, but generating negative values is less likely than from the first three distributions. But it does have same tendency as shown in the plot "Normal.ub.20.ind." There are many outliers in the bootstrap distribution of parameter estimates by RR-gamma procedure which result in the less efficient mean estimates but the median is still satisfied. Such problems are not observed for RR-pois procedures. Therefore, in case of unbalanced data with small sample size, RR-pois procedure may be the preferable procedure for the robust smooth bootstrap estimation.

The simulation studies for Poisson and lognormal data yield generally similar results. Figure 4.3 and 4.4 are the simulation results for Poisson data with sample size 40 and 20 respectively. Figure 4.5 and 4.6 are for lognormal data. All the data are generated with AR(1) correlation with the correlation coefficient being .5. The layout of the plots are the same as normal data. As can be seen from Figure 4.3 and 4.4, the performances of different methods in terms of relative efficiency of standard deviation estimation for Poisson data are almost the same as in normal response case. As for lognormal data, there are some minor differences from the simulation results for normal and Poisson data. First, many outliers in the right tail of the boxplots can be seen for all the estimators. Second, the variations of the

standard deviation estimators for lognormal data are generally larger than those for normal and Poisson data. For lognormal data, the IQRs of the naive and MR estimators have similar lengths to the robust estimators. Therefore, these model-based estimators do not have an obvious advantage in variation as in the cases of normal and Poisson data. The most likely explanation is that the estimation via GEE model for data not from exponential family is not as efficient as that from the exponential family. Despite the larger variance of the standard deviation estimation, the superiority of RR procedures over other competitors is clearly illustrated on the graphs. Furthermore, in case of unbalanced data of sample size 20 (the right panel in Figure 4.6), the first three RR procedures are more stable than those for normal and Poisson data. In particular, in the case of AR(1) or EXC working correlation (the first two plots in the right panel in Figure 4.6), the median lines of RR-unif estimates fall on the vertical line.

To summarize, the proposed RR procedures and MR procedures in case of correct covariance model can generate more accurate variance estimates compared with the usual "sandwich"-type variance estimators and classical resampling estimators. But the choices of random weights depend on the sample size and the data structure. Here we list out some practical guidelines on the choices of weights for **the variance estimation of the parameter estimates**:

- For balanced data of moderate or large sample size, **RR-norm, RR-de and RR-unif** procedures can provide better estimates for the standard deviation of the regression parameter estimates;
- For unbalanced data, especially of small sample size, RR-norm, RR-de and RR-unif tend to produce unstable bootstrap copies of parameter estimates

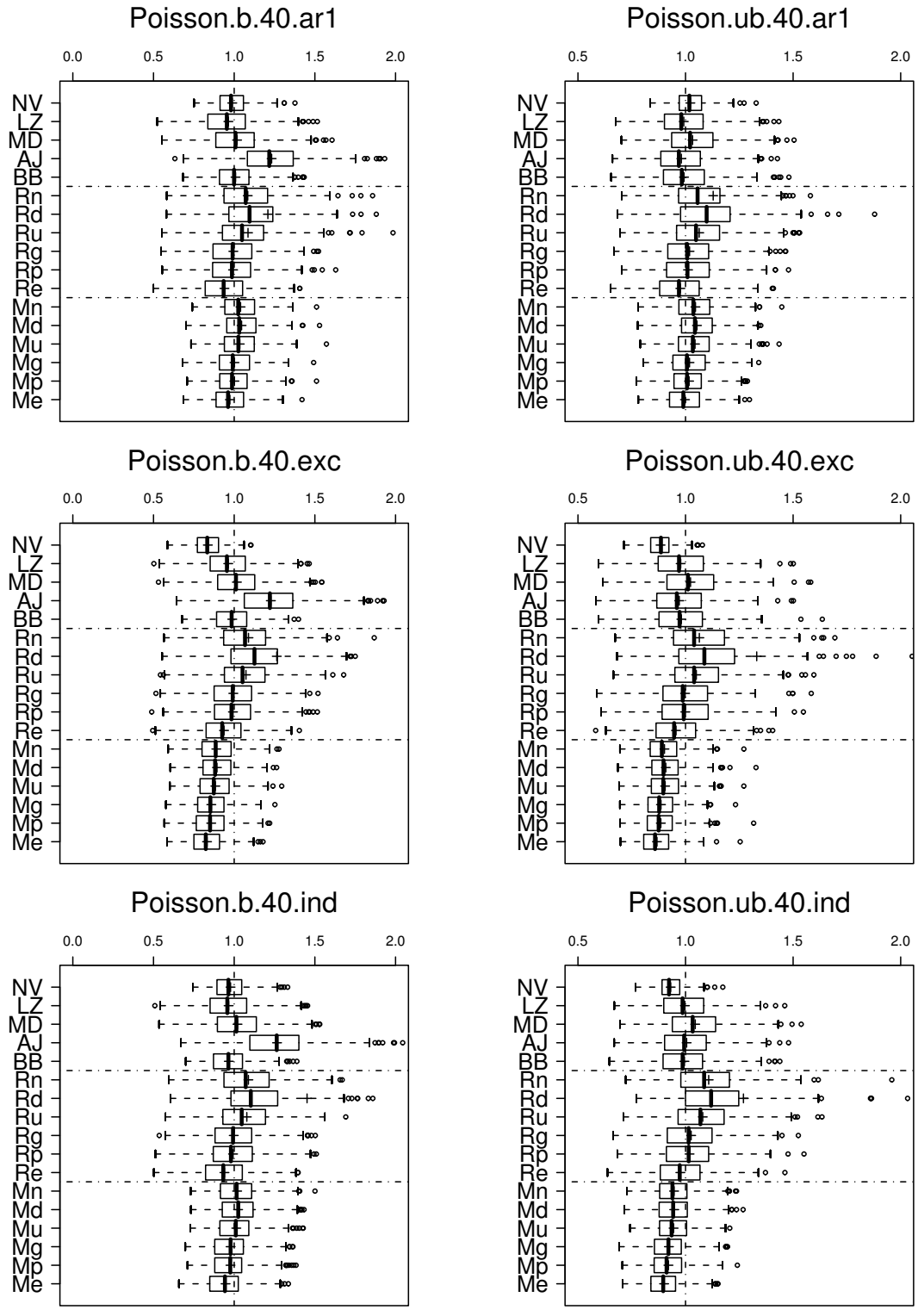

Figure 4.3: Relative efficiency of std. dev. estimates for Poisson data, K=40, sim=500, boot=500

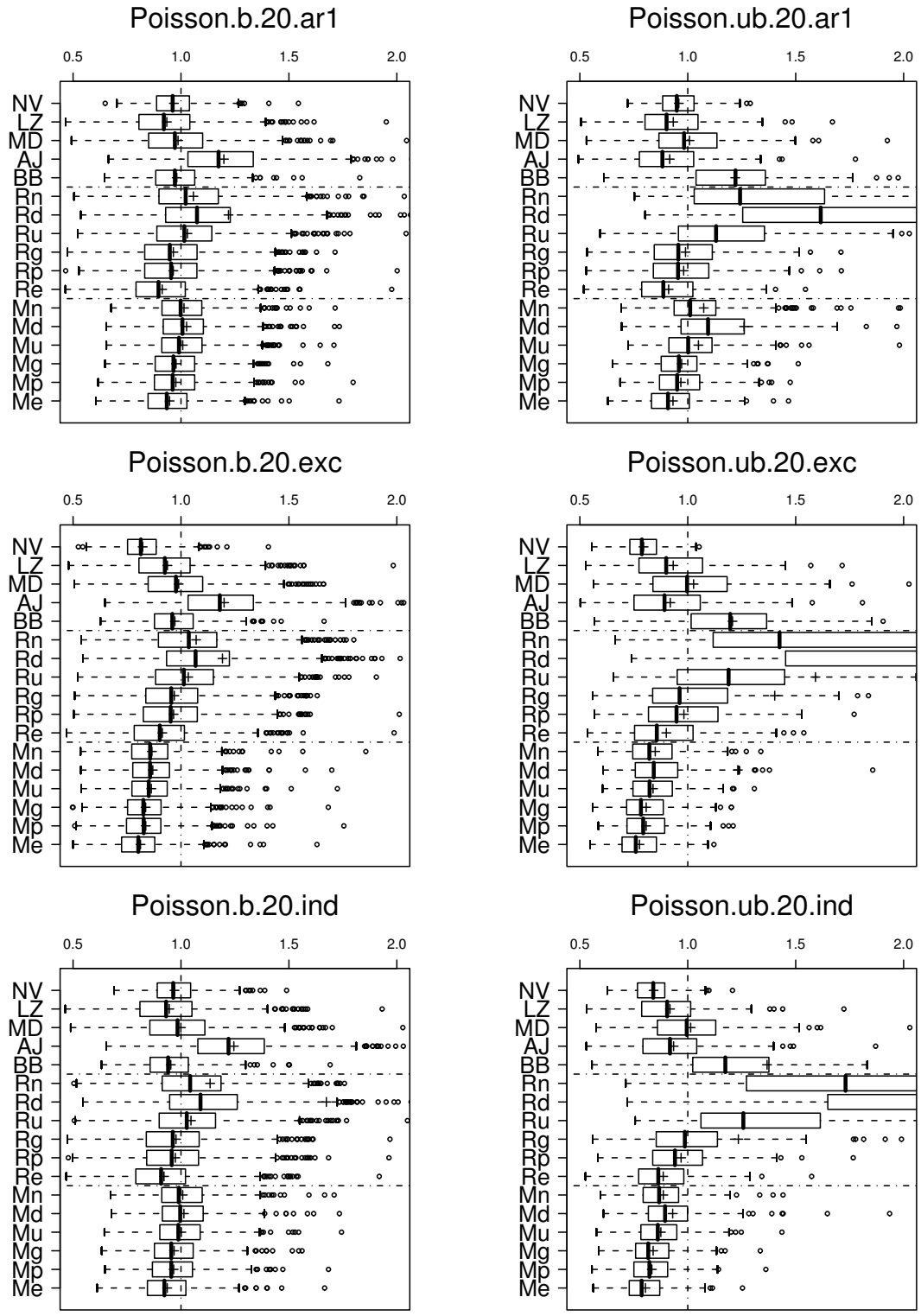

Figure 4.4: Relative efficiency of std. dev. estimates for Poisson data, K=20, sim= 1000, boot=1000

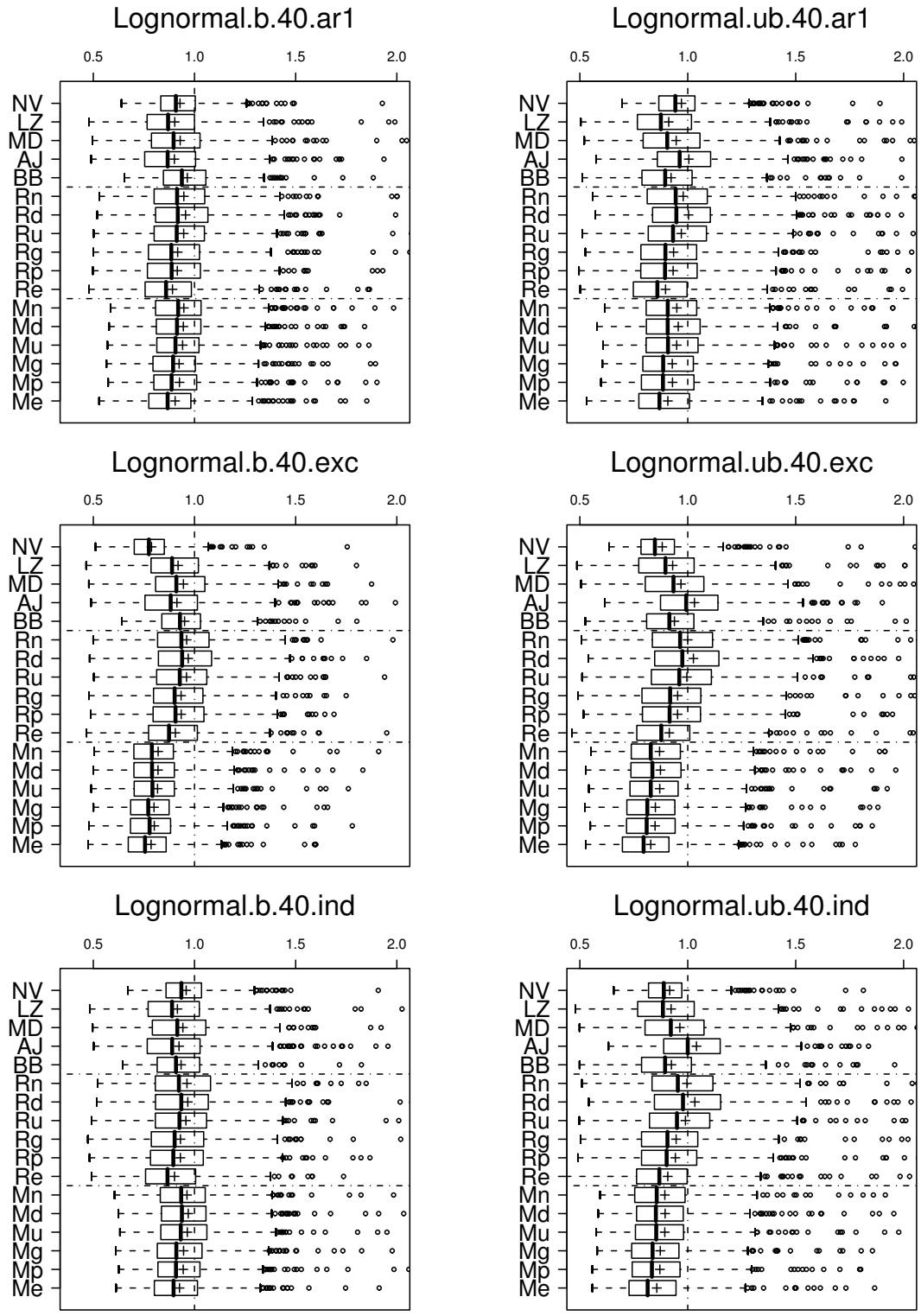

Figure 4.5: Relative efficiency of std. dev. estimates for lognormal data, K=40, sim=500, boot=500

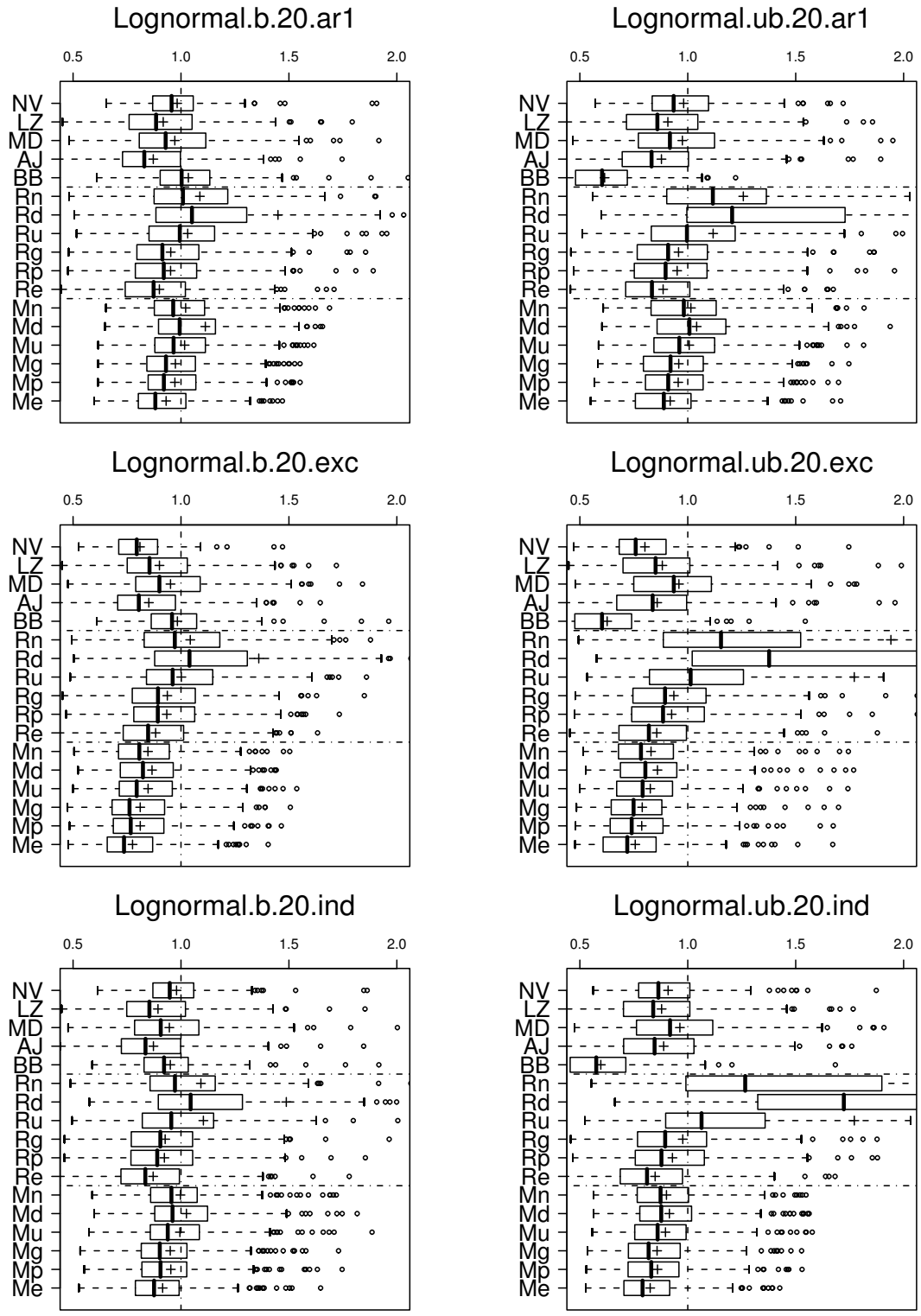

Figure 4.6: Relative efficiency of std. dev. estimates for lognormal data, K=20, sim=1000, boot=1000

and thus the resulting variance estimators may become inconsistent. **RR-gamma and RR-pois**, especially the latter, will be the efficient smooth bootstrap procedure to be applied; and

- In case of the correlation structure could be correctly specified, **MR-norm, MR-de and MR-unif** would be more efficient than the RR counterparts since they generally have smaller variance than the RR counterparts.

4.2.2 Confidence interval coverage

We have also investigated the 95% and 80% confidence interval coverage probabilities for different variance estimators. SD-type confidence intervals are constructed for the naive, LZ, MD, KC and AJS methods. That is, the two endpoints of the confidence intervals are calculated as the parameter estimates plus or minus the quantile from standard normal for specific level of significance (in our case, 0.2 or 0.05) multiplied by the standard errors. The confidence intervals in ABB are constructed using percentile methods, that is, 10% and 90% quantiles of the bootstrap distribution of the parameter estimates are used as the two endpoints of 80% confidence interval; 2.5% and 97.5% quantiles are used for 95% confidence interval. For RR and MR procedures, we construct both types of confidence intervals, SD-type and percentile.

- **SD-type CI for smooth bootstrap methods**

Considering the accuracy of the variance estimation observed in the first aspect of the simulation, we think and see SD-type confidence interval by the proposed RR and MR approaches (especially using the first three types of random vari-

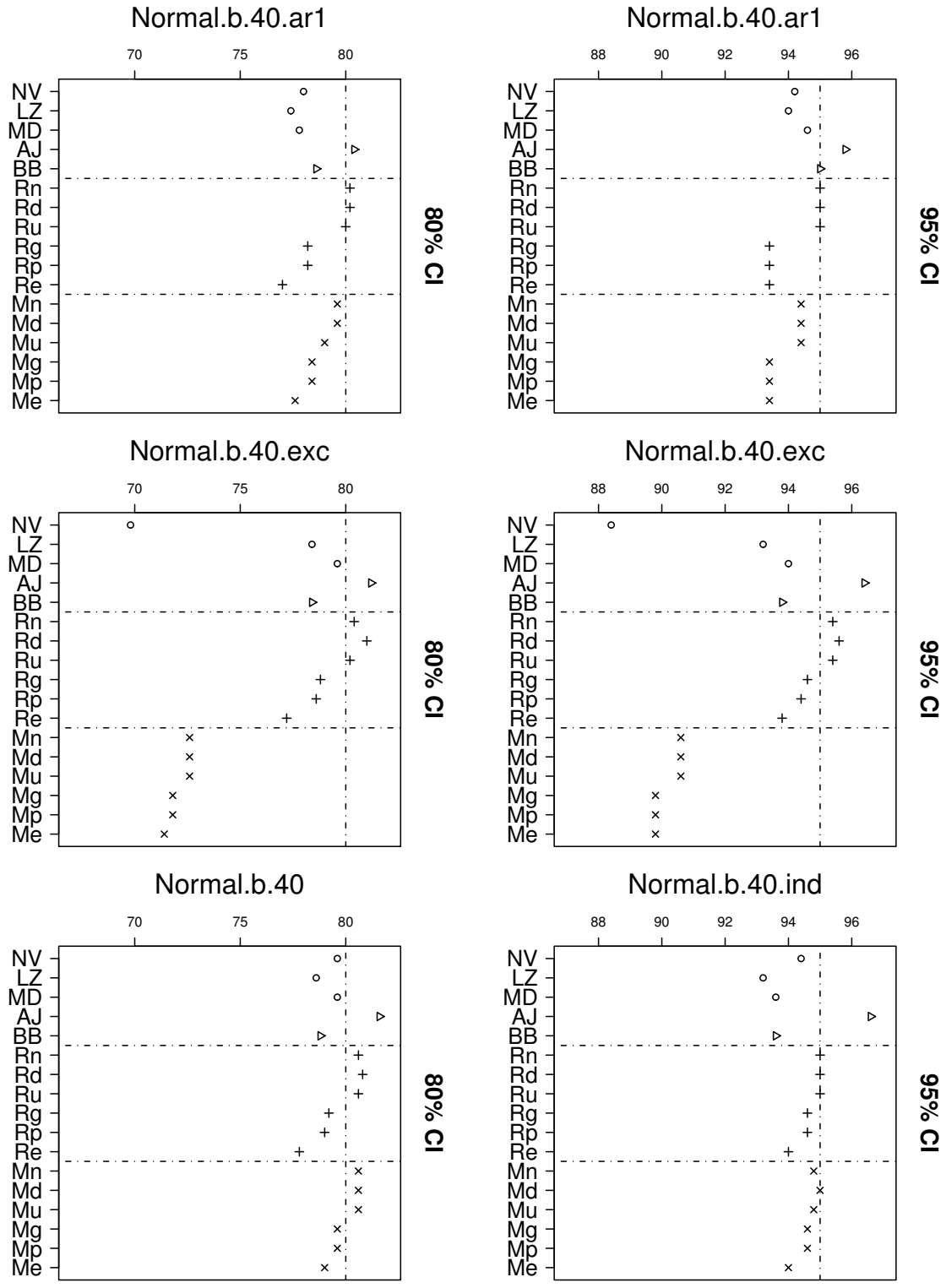

Figure 4.7: 80% and 95% CI coverage probabilities for normal balanced data, K=40, sim=500, boot=500 (SD-type CI used for smooth bootstrap methods)

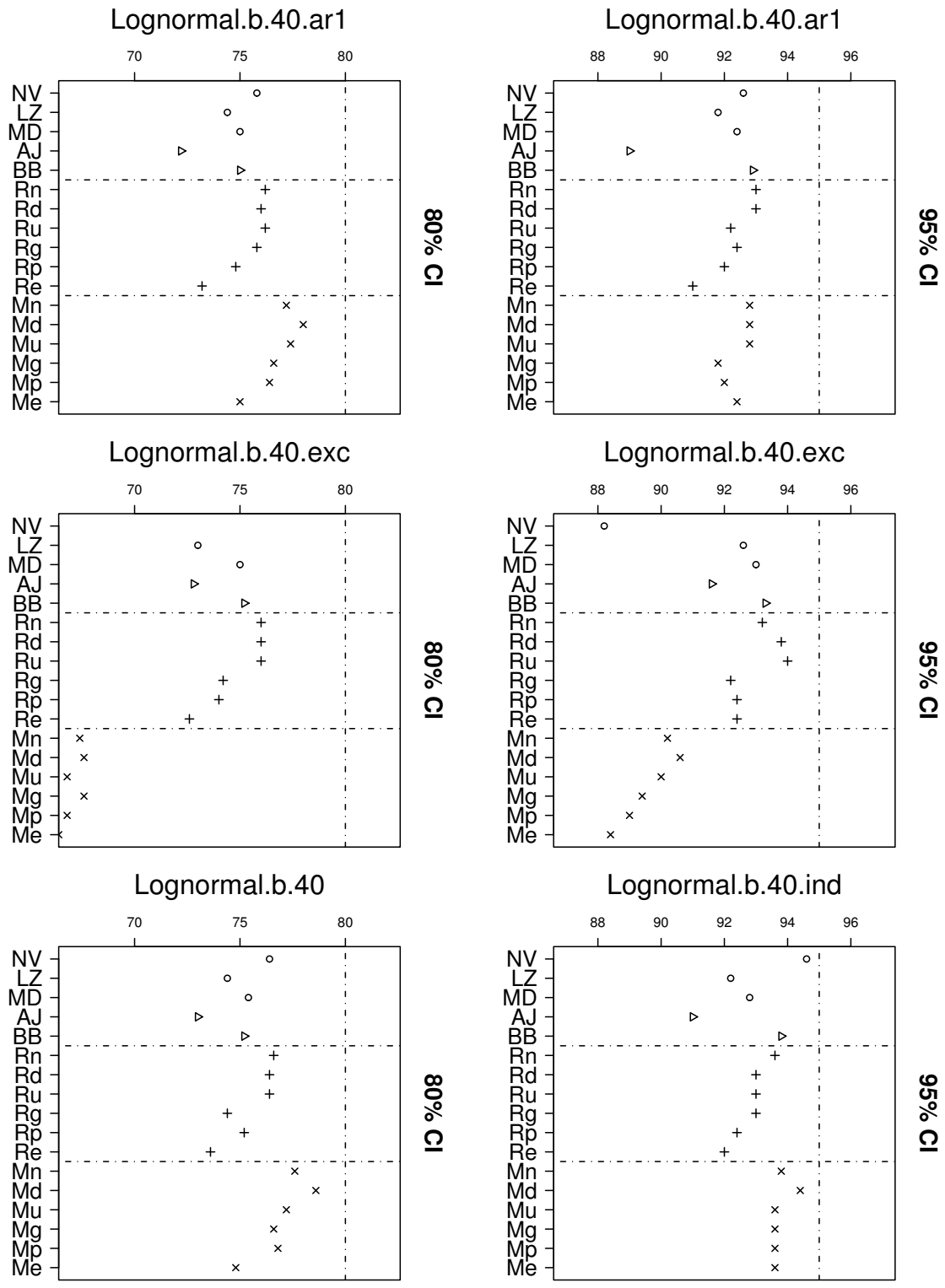

Figure 4.8: 80% and 95% CI coverage probabilities for lognormal balanced data, K=40, sim=500, boot=500 (SD-type CI used for smooth bootstrap methods)

ates) should easily beat the other candidates, because they correct the bias of the sandwich variance estimator and are robust to the misspecification of correlation structure. It is reasonable to conclude from the results that in the previous simulation studies that the RR and MR approaches outperform most other candidates in most cases. Here we only illustrate the simulation results for balanced normal data of sample size 40 in Figure 4.7 and balanced lognormal data of sample size 40 in Figure 4.8. All the other cases lead to almost the same conclusion. The left panel presents the results for 80% confidence interval while the right panel for 95% confidence interval. The vertical lines in the plots are the nominal values: 80% and 95%. In each individual plot of the six in Figure 4.7, the candidate methods are classified into the same three groups and are laid out in the same way as in the previous boxplots. For clearer presentation, besides including the two horizontal lines, different symbols for data points are used to differentiate the results from different methods: the first three "sandwich" estimators, the naive, LZ and MD, are presented using "$\circ$"; the two classical resampling methods use "$\triangleright$"; the six RR procedures use "+"; and the MR procedures use "$\times$". As seen in Figure 4.7, the 80% and 95% confidence interval coverage probabilities by RR-norm, RR-de and RR-unif are the closest to the nominal values in all cases, even identical to the nominal values in some cases. The differences in the smooth bootstrap methods when using different random variates can be clearly seen from this Figure. They can be classified into three groups: the first three perform similarly; followed by those using Gamma and Poisson random variates; and finally is the procedure using unit exponential random variates. When SD-type confidence intervals are used, the relationship of confidence interval length together with the variations of the lengths by different methods is the same as what we observe for the variance

estimation.

- **Percentile CI for the smooth bootstrap methods**

We also construct the percentile confidence intervals for the proposed smooth bootstrap methods. This is more interesting because we can check whether the bootstrap distribution by the proposed procedure well approximate the true distribution of the parameter estimates **without using normal approximation**. The answer turns out to be a promising "yes"! The approximation of the parameter estimate distribution by the bootstrap distribution obtained from the proposed procedures is quite satisfactory especially when using random variates from the specific normal, double exponential and uniform distribution in Table 4.2. A simple illustration is shown in Figure 4.9. It presents the comparison of the distribution of the parameter estimates with the bootstrap distribution obtaining from RR-norm procedure for unbalanced Poisson data of sample size 20 under the working independence assumption. As illustrated for this unbalanced small data set, the approximation is quite good and the two curves from simple kernel fitting are bell-shaped. The good approximation can also be seen from constructing percentile confidence intervals. All the results for percentile confidence intervals are presented here for reference. We mainly explain the results for normal responses. Similar results can be seen in the graphs for Poisson and lognormal data.

Figure 4.10 and 4.11 show the 80% and 95% confidence interval coverage probabilities (in %) for normal data of sample size 40, one for balanced and the other for unbalanced structure. Figure 4.12 and 4.13 are the results for normal data of sample size 20, balanced and unbalanced respectively. The same plot layout is used as in Figure 4.7 and 4.8. From these graphs, we can see that there is

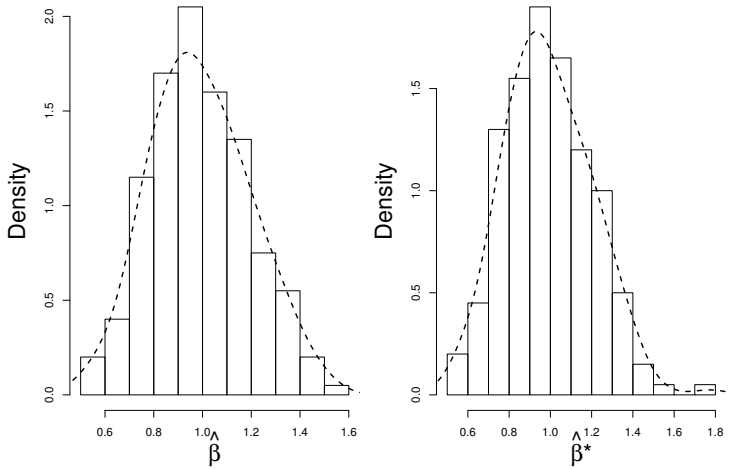

Figure 4.9: Histograms for parameter estimate and the bootstrapped estimates for unbalanced Poisson data of sample size 20

an analogy between the performance of these "sandwich" and resampling methods in terms of coverage probabilities and variance estimation. Figure 4.10 and 4.11 illustrate that for normal data of moderate sample size, the classical resampling methods generally perform only as good as LZ estimators in constructing confidence intervals, especially when the data is unbalanced. The coverage probabilities by AJS methods for balanced data are always larger than the nominal values. The MD estimators perform better than LZ because of the bias-correction. The RR-norm, RR-de and RR-unif procedures have higher or at least comparable coverage probabilities to the MD estimators, but have better performance for 95% confidence intervals than for 80% CIs. RR-gamma and RR-pois perform not as good as the first three RR procedures but are generally comparable to the LZ estimator. The RR-exp does not have very satisfactory performance. The naive estimator and MR procedures still perform the best with correctly specified correlation structure and satisfactorily with working independence for balanced data, but are less efficient in other cases. The same conclusions can be drawn from Figure 4.12 and 4.13 for normal responses of sample size 20.

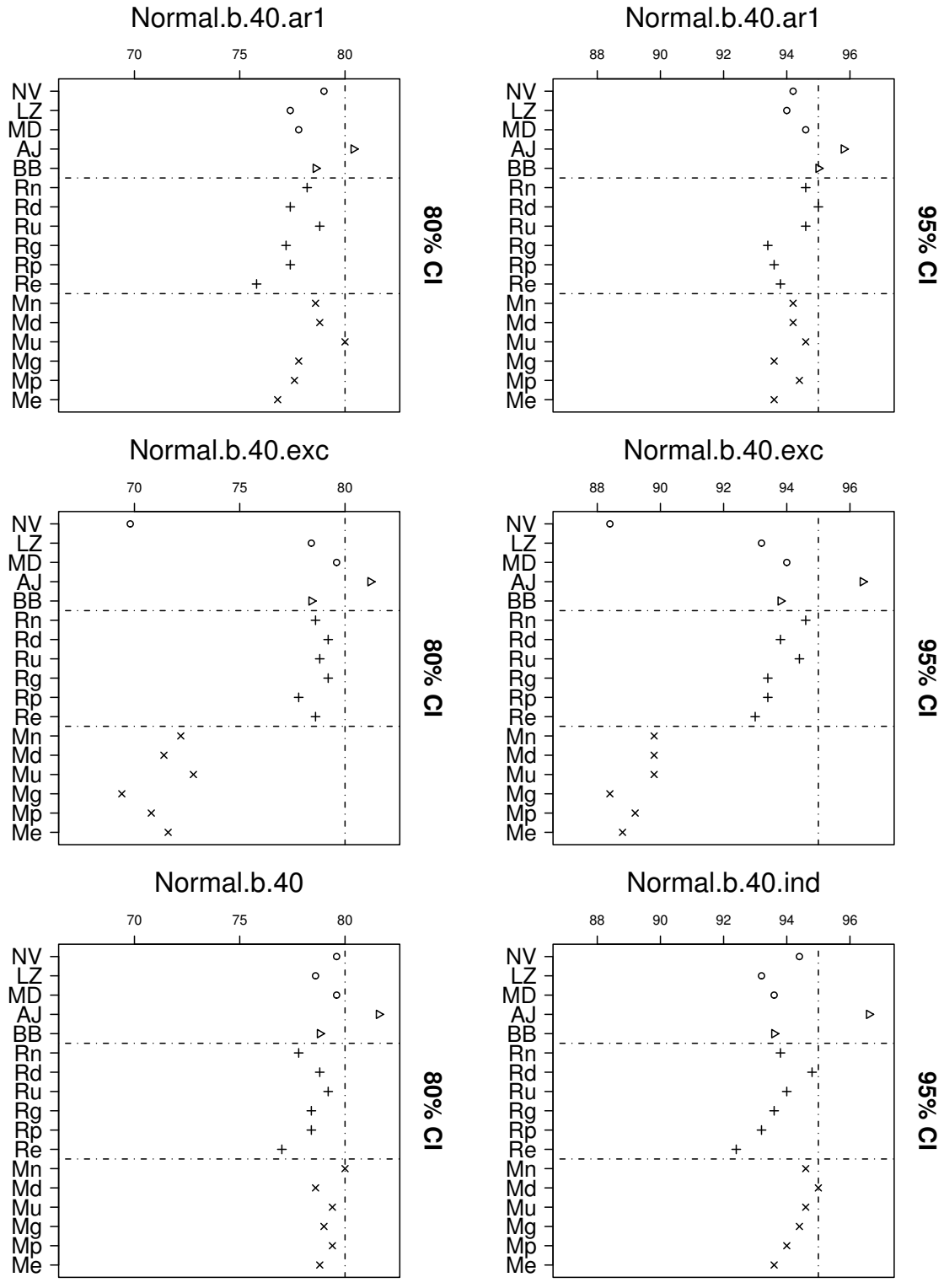

Figure 4.10: 80% and 95% CI coverage probabilities for normal balanced data, K=40, sim=500, boot=500

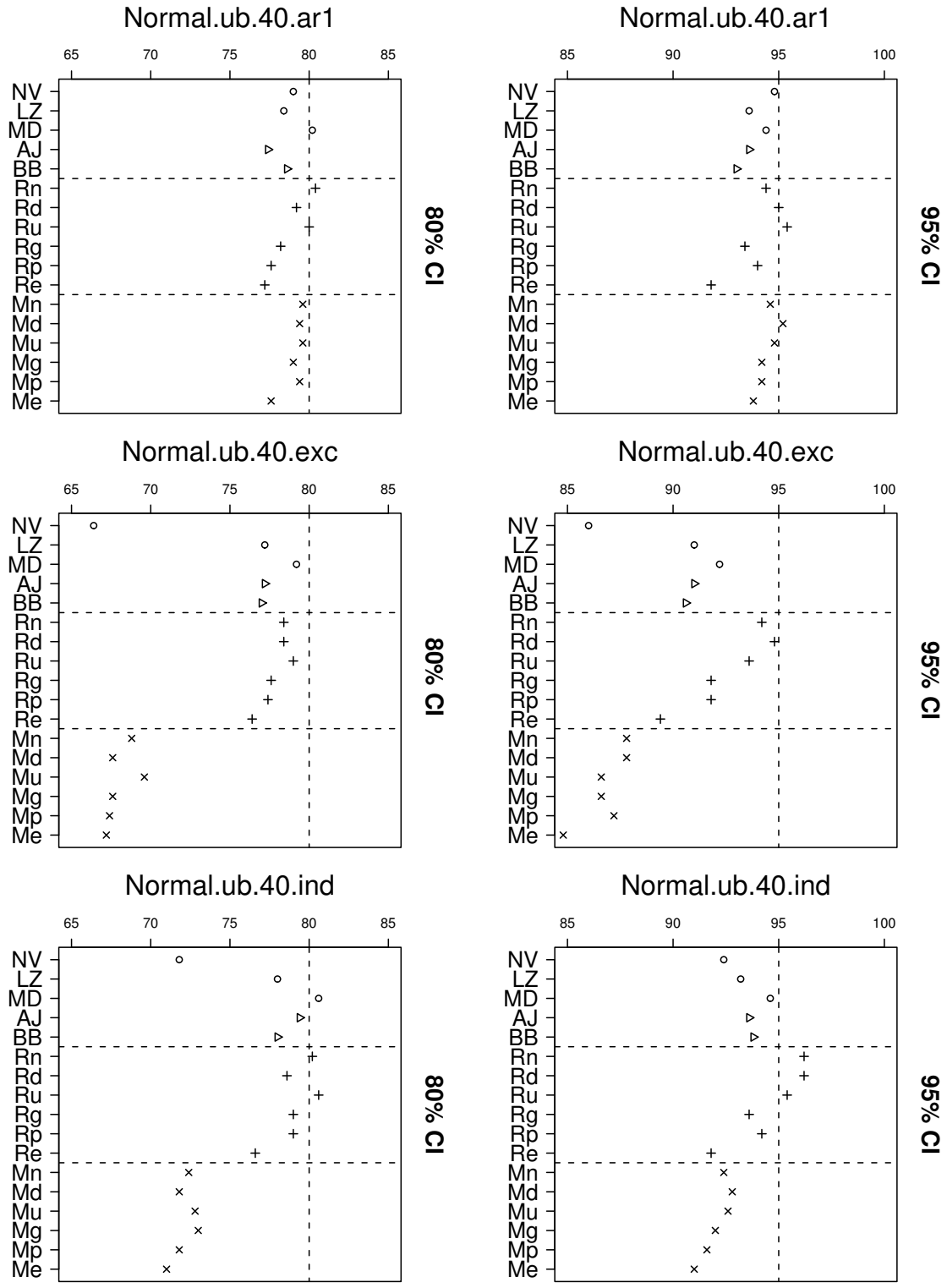

Figure 4.11: 80% and 95% CI coverage probabilities for normal unbalanced data, K=40, sim=500, boot=500

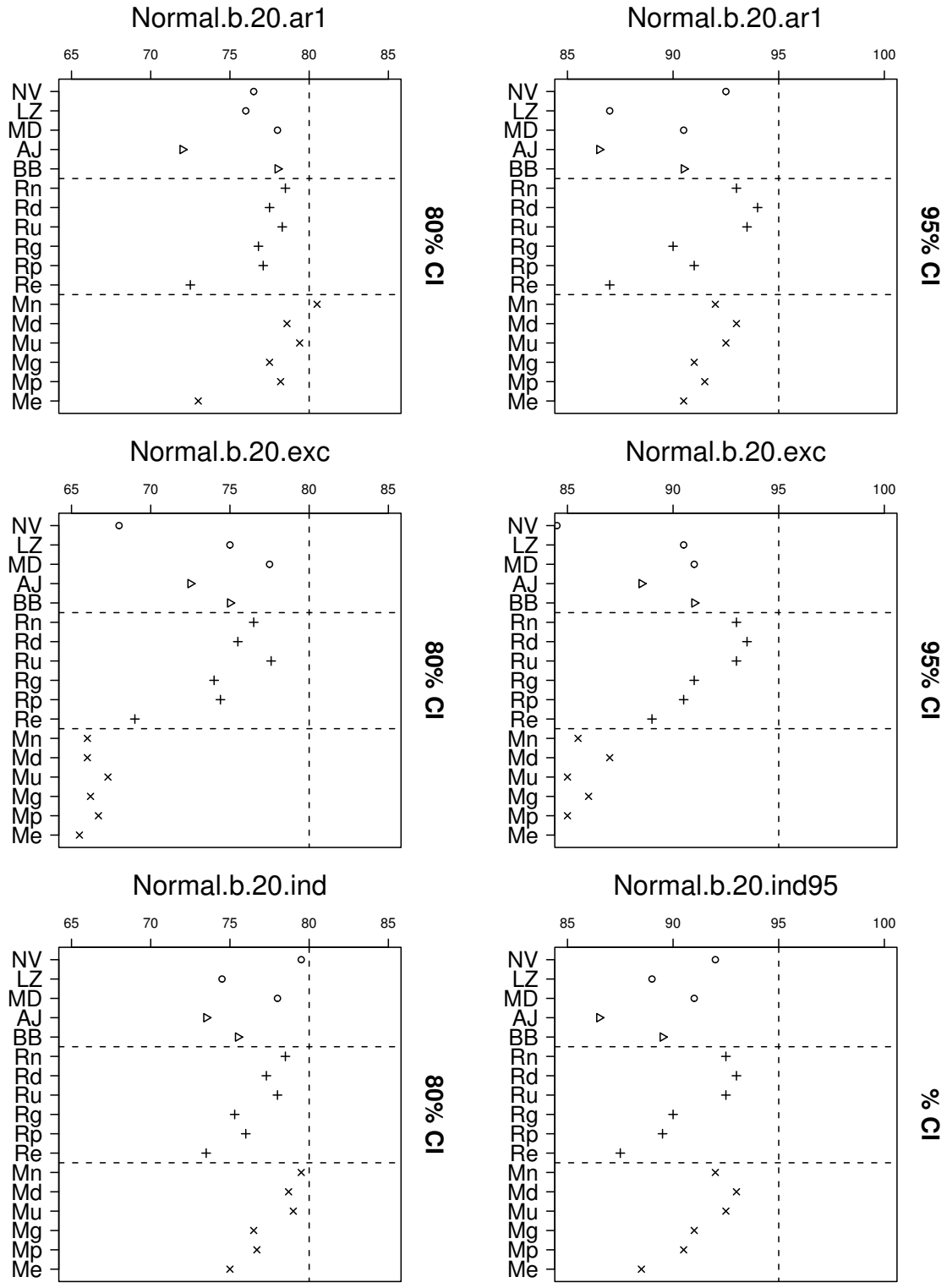

Figure 4.12: 80% and 95% CI coverage probabilities for normal balanced data, K=20, sim=1000, boot=1000

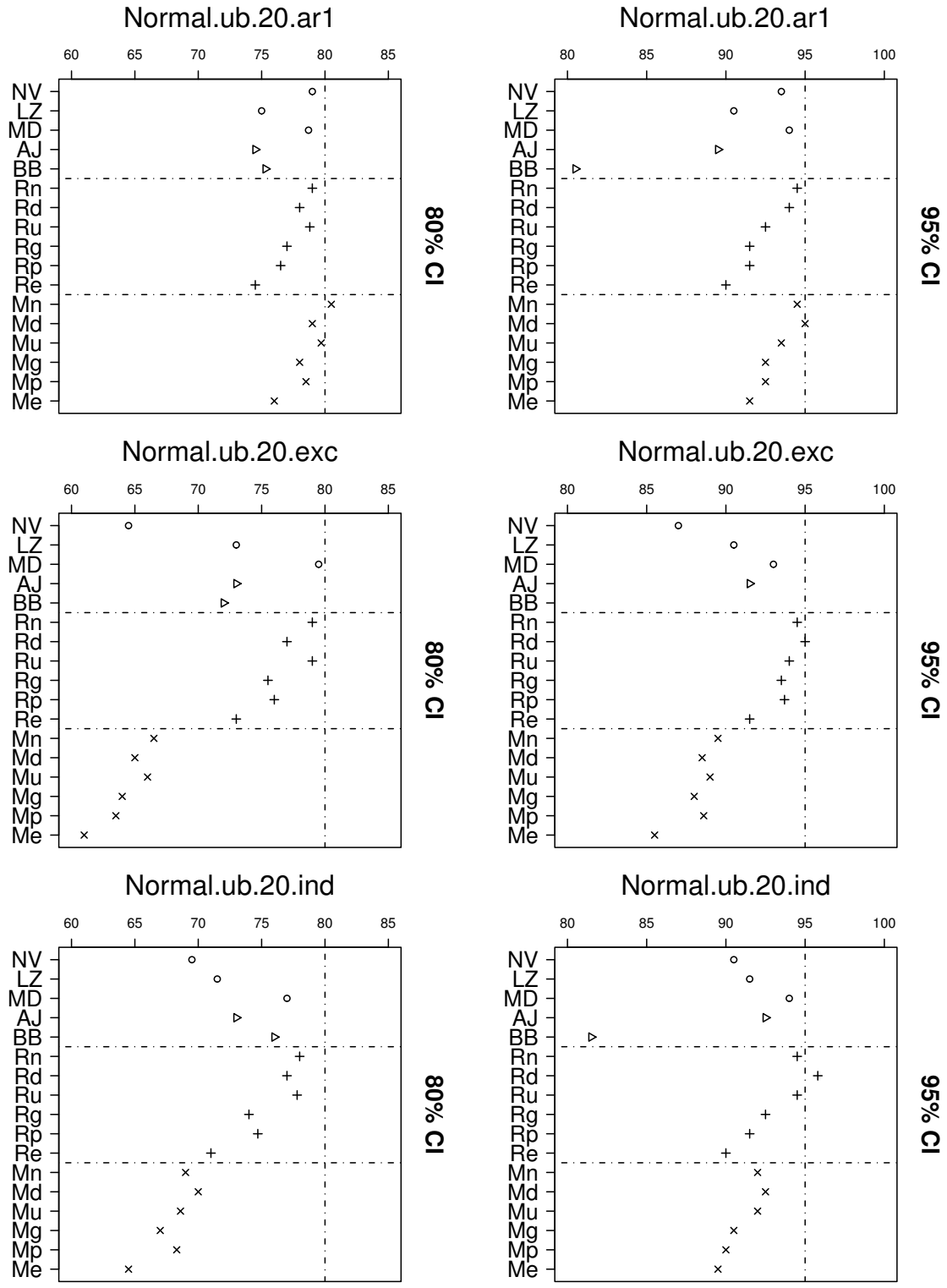

Figure 4.13: 80% and 95% CI coverage probabilities for normal unbalanced data, K=20, sim=1000, boot=1000

As for Poisson and lognormal data, the simulation results about confidence interval coverage when percentile CIs are used for smooth bootstrap procedures are presented in Figure 4.14 - 4.21. The most preferable confidence intervals in most of the cases are also MD and the first three RR procedures: RR-norm, RR-de and RR-unif. Therefore, different from the guideline for variance estimation, the first three RR procedures are recommended in constructing **percentile** confidence intervals for all cases. For SD-type confidence intervals, the guidelines for variance estimation can be applied.

4.3 Real data application

4.3.1 Leprosy study

In this section we apply the smooth bootstrap method to a real dataset from a clinical trial of antibiotics for Leprosy (page 312, Fitzmaurice et al; 2004). The data consists of counts from a randomized and placebo-controlled clinical trial of 20 patients with leprosy at the Eversley Childs Sanitorium in the Philippines. Each patient was randomly assigned to either of two antibiotics (denoted treatment drug A and B) or a placebo (denoted treatment drug C). The numbers of leprosy bacilli at six sites on the body were recorded before and after the treatment. The scientific interest of this study is whether treatment with antibiotics (drugs A and B) reduces the abundance of leprosy bacilli at the six sites on the body when compared to placebo (drug C). This dataset displays substantial overdispersion under a Poisson model. We used the following marginal model:

$$\log E(Y_{ij}) = \log(\boldsymbol{\mu}_{ij}) = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \,\mathrm{time}_{ij} + \boldsymbol{\beta}_3 \,\mathrm{time}_{ij} \times \mathrm{trt}_{1i} + \boldsymbol{\beta}_4 \,\mathrm{time}_{ij} \times \mathrm{trt}_{2i}\,, \quad (4.3.1)$$

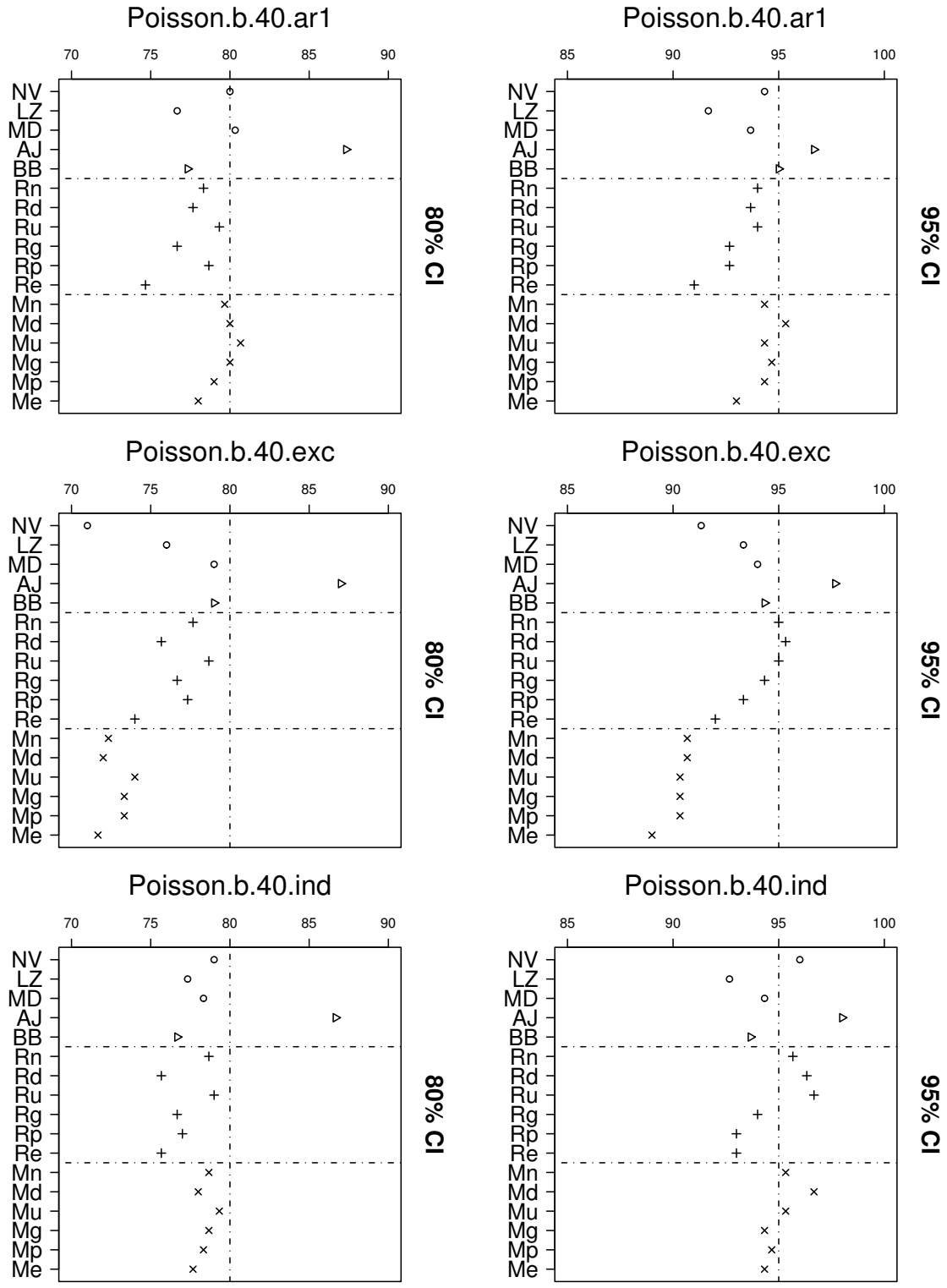

Figure 4.14: 80% and 95% CI coverage probabilities for Poisson balanced data, K=40, sim=500, boot=500

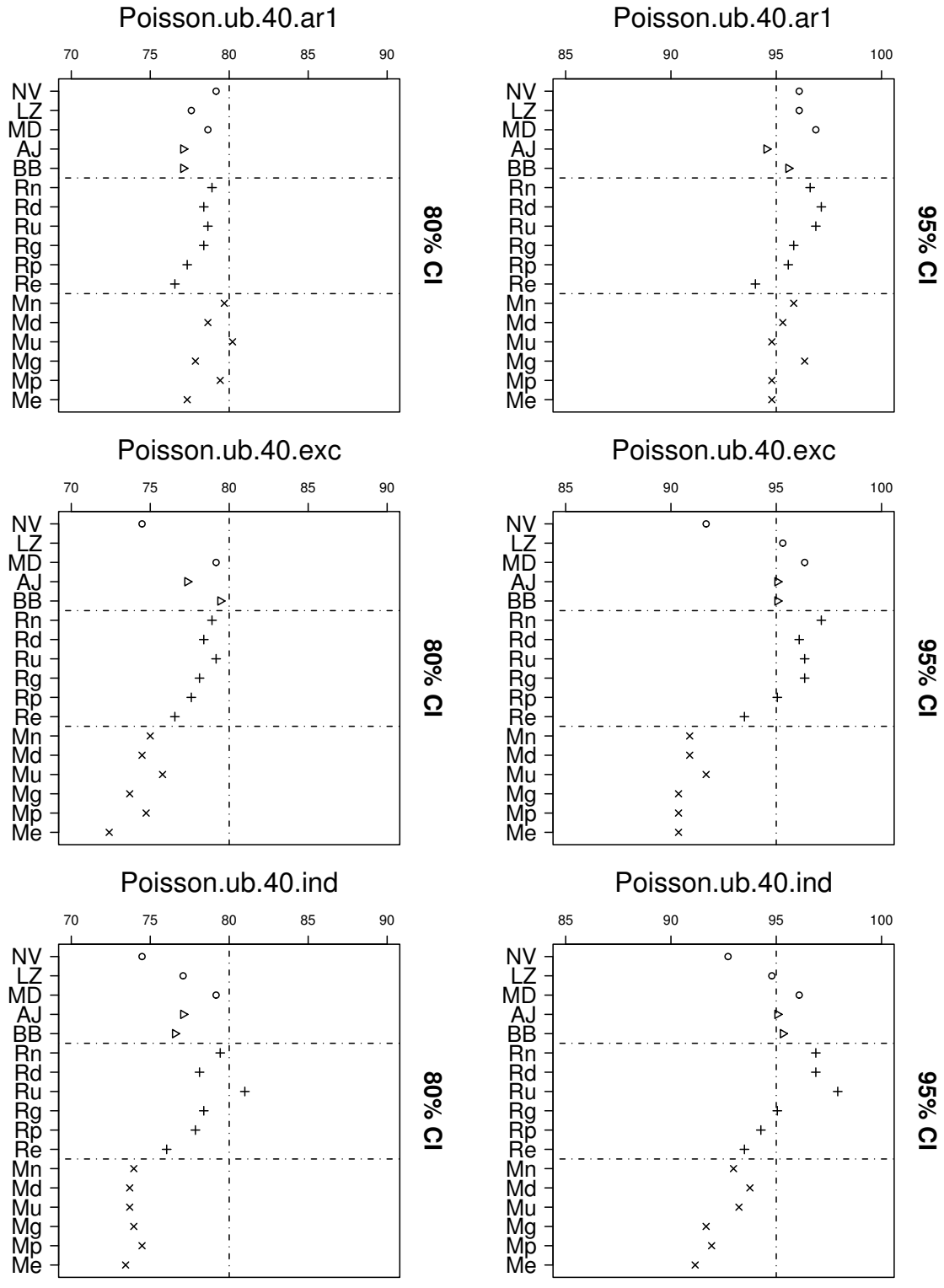

Figure 4.15: 80% and 95% CI coverage probabilities for Poisson unbalanced data, K=40, sim=500, boot=500

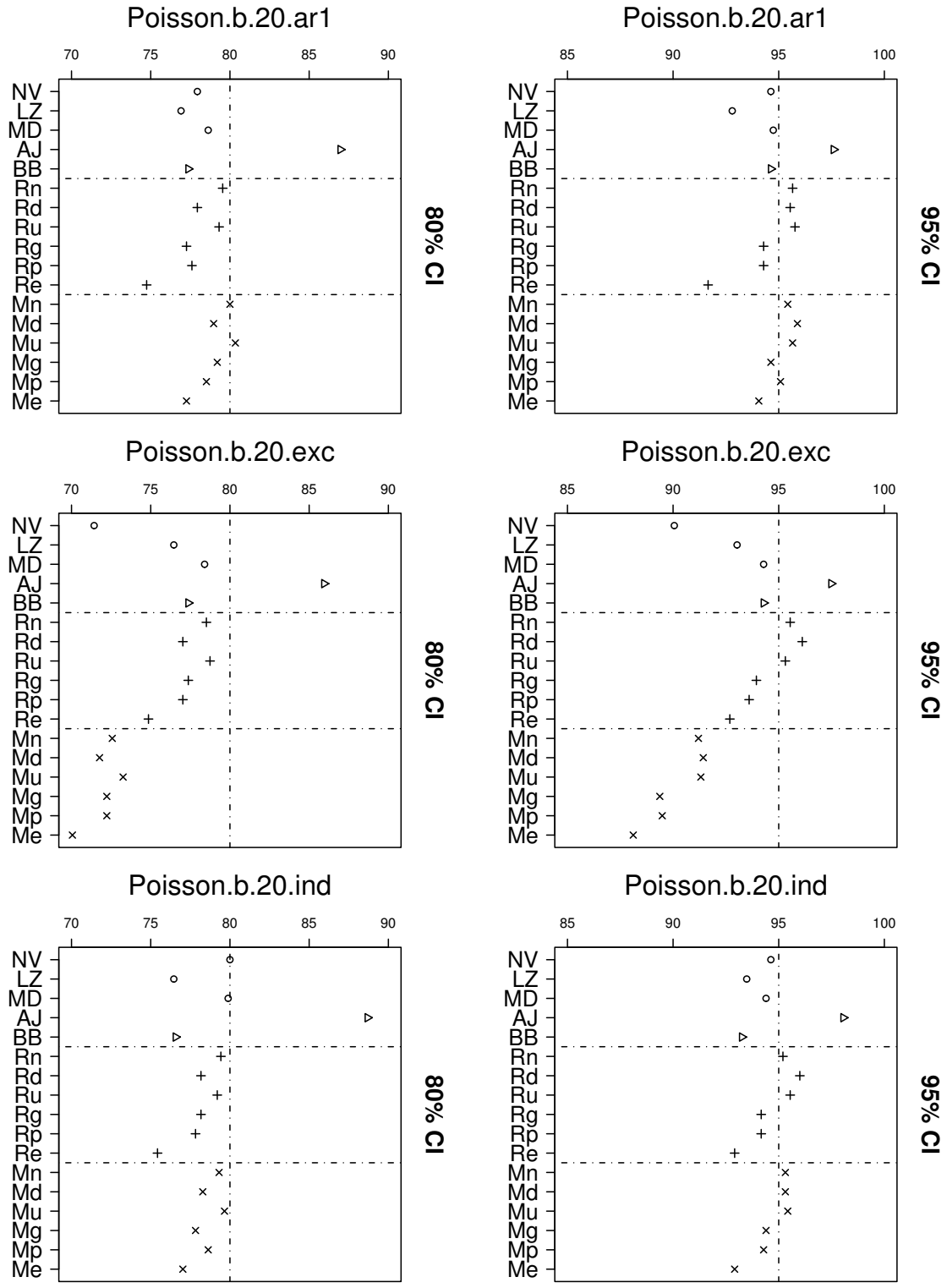

Figure 4.16: 80% and 95% CI coverage probabilities for Poisson balanced data, K=20, sim=1000, boot=1000

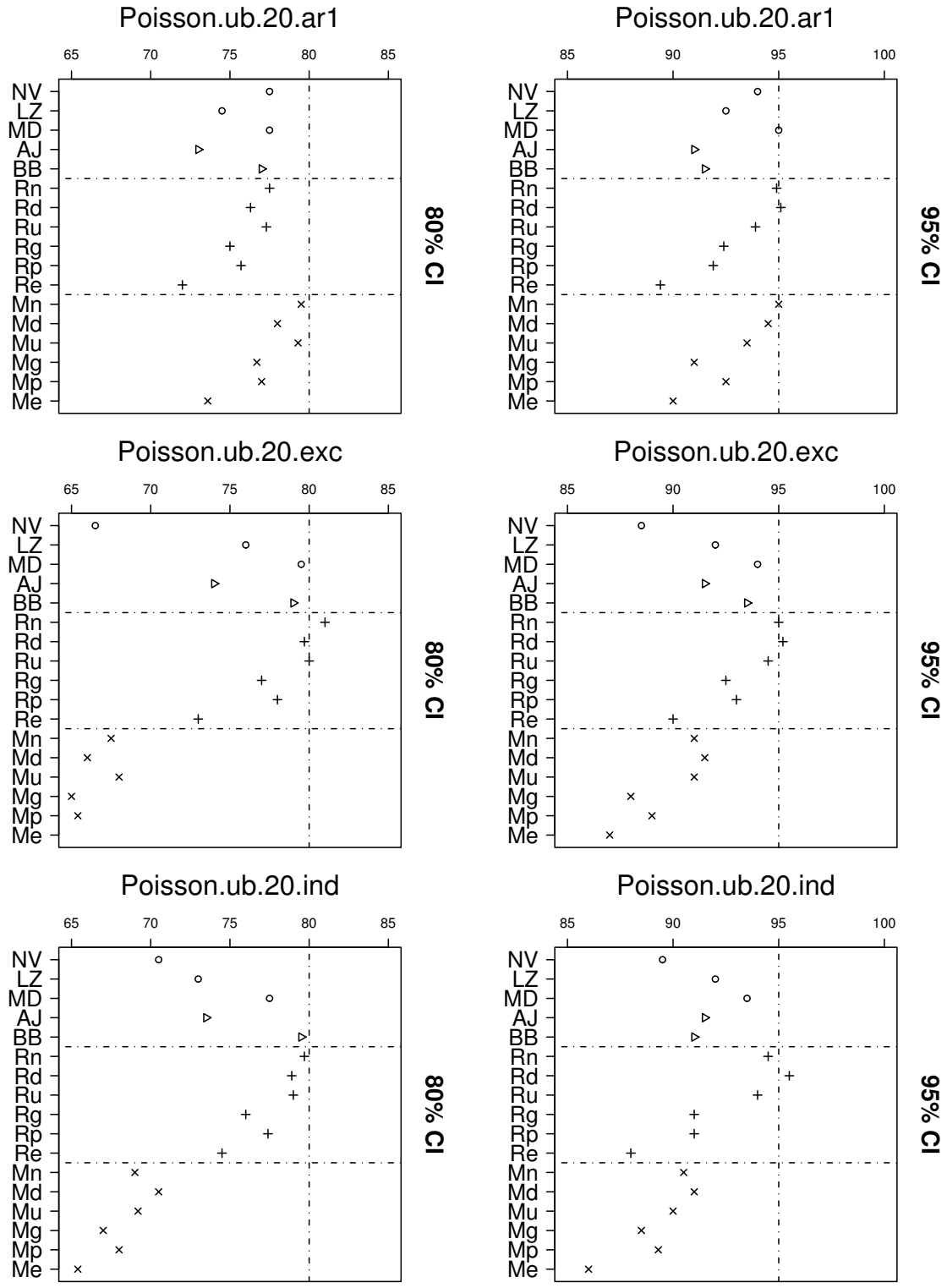

Figure 4.17: 80% and 95% CI coverage probabilities for Poisson unbalanced data, K=20, sim=1000, boot=1000

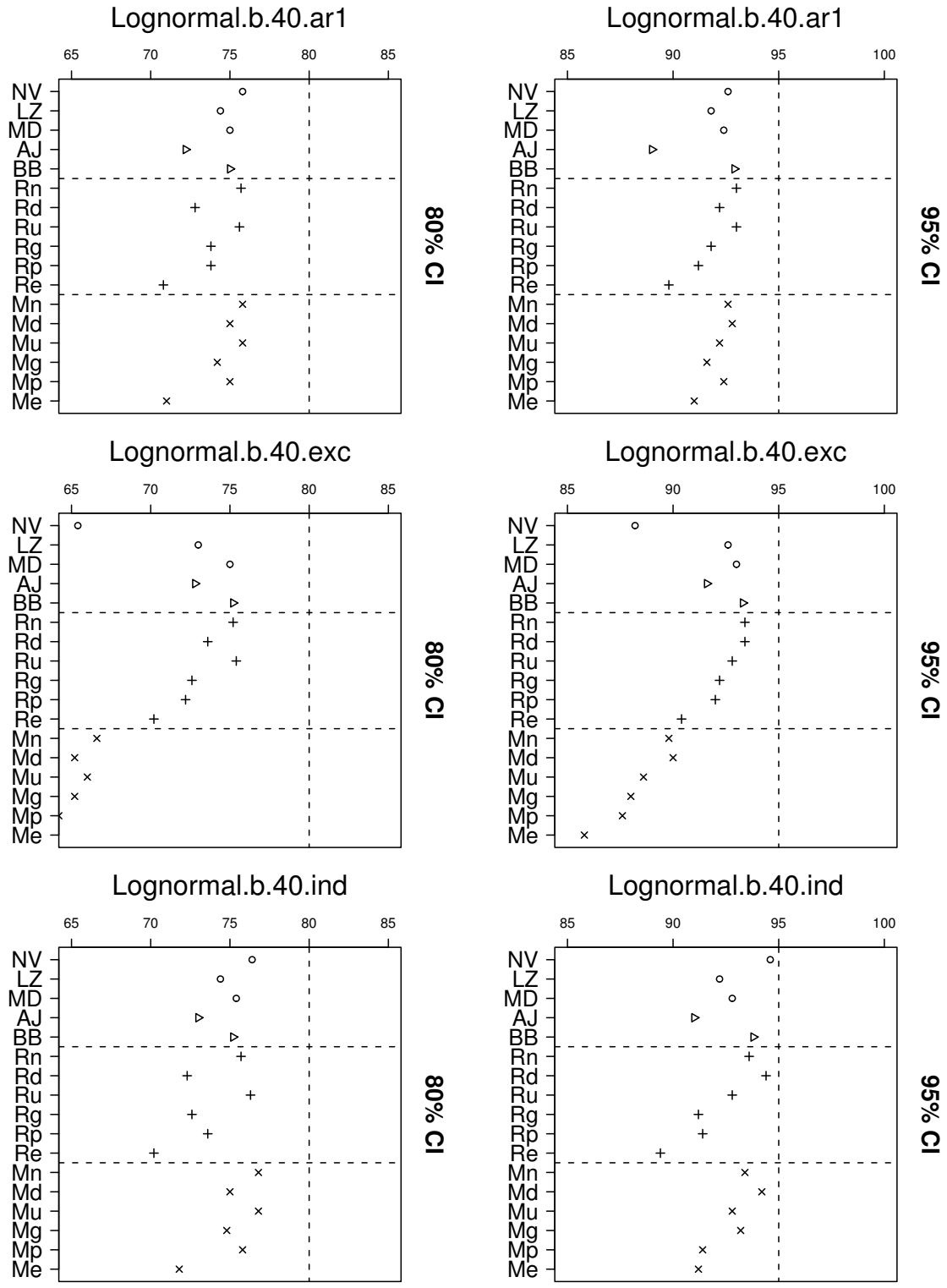

Figure 4.18: 80% and 95% CI coverage probabilities for lognormal balanced data, K=40, sim=500, boot=500

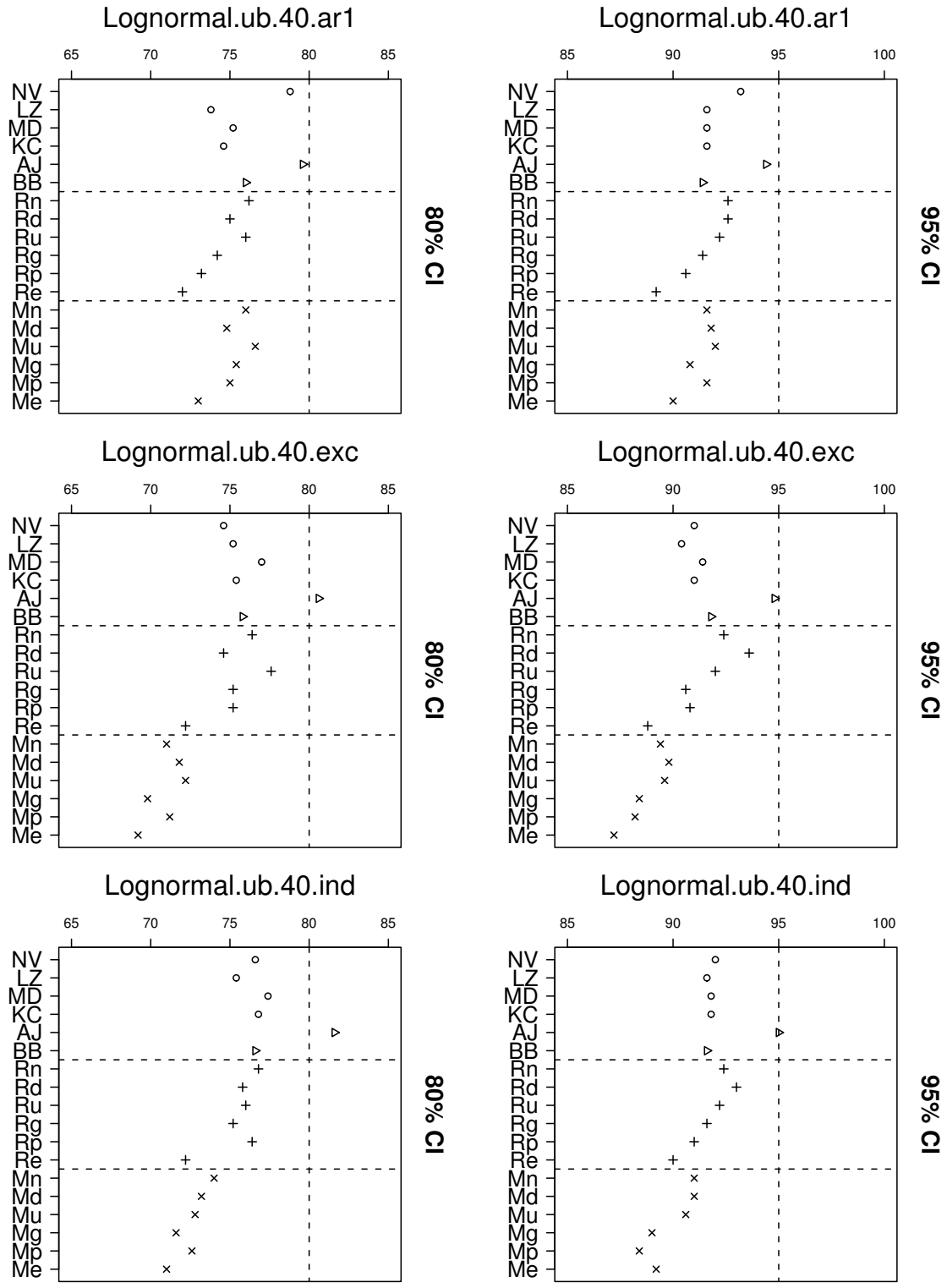

Figure 4.19: 80% and 95% CI coverage probabilities for lognormal unbalanced data, K=40, sim=500, boot=500

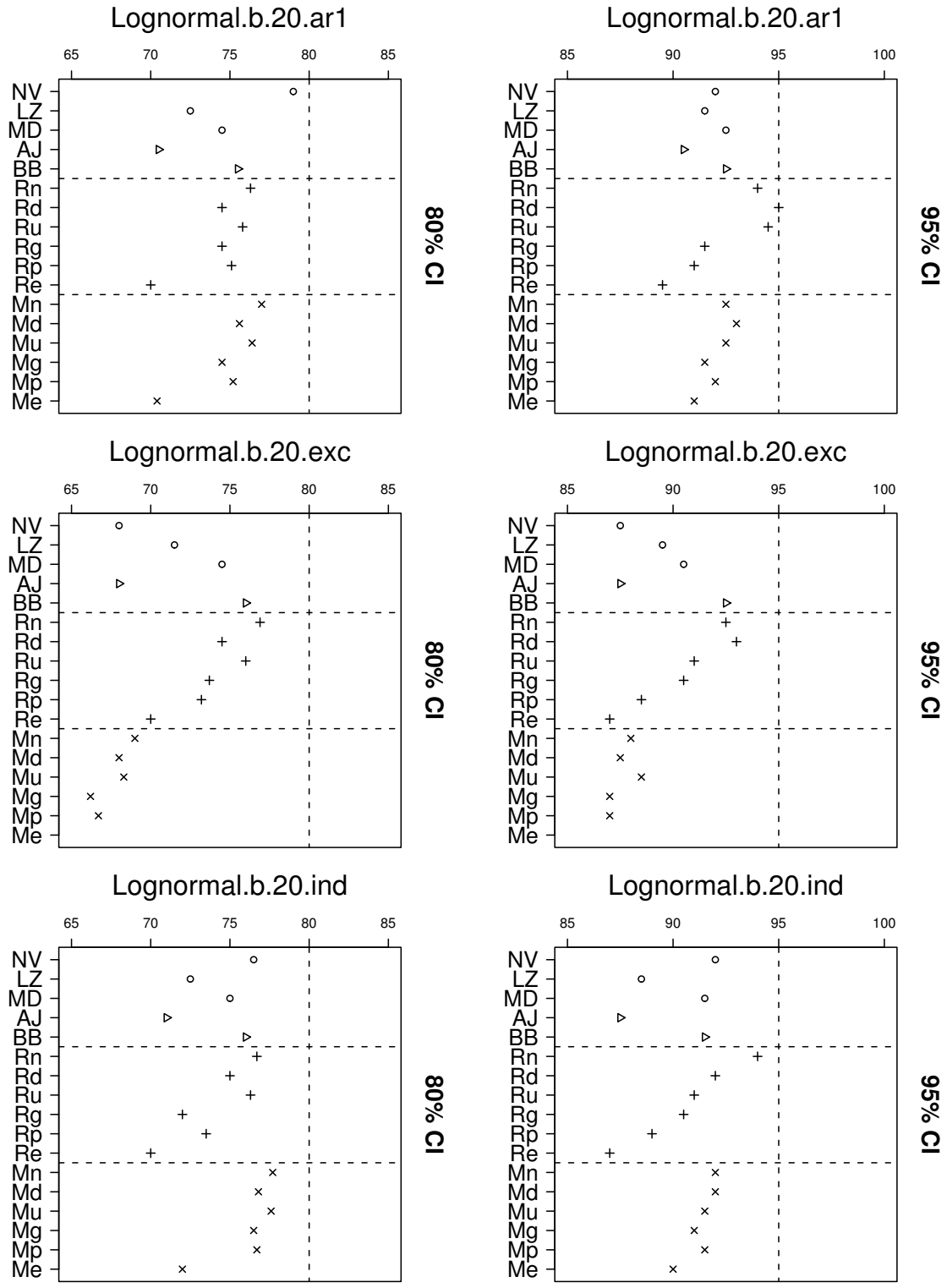

Figure 4.20: 80% and 95% CI coverage probabilities for lognormal balanced data, K=20, sim=1000, boot=1000

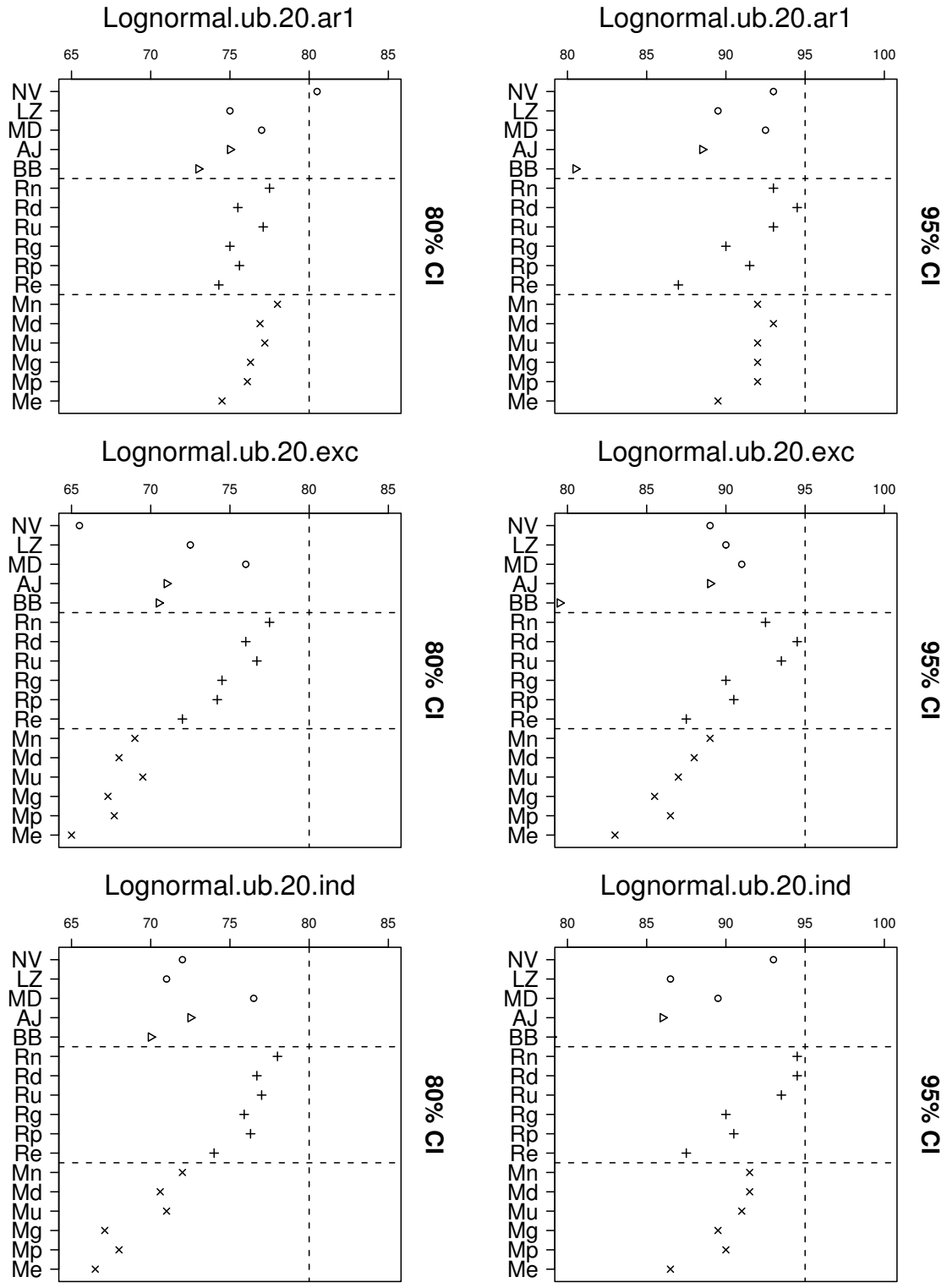

Figure 4.21: 80% and 95% CI coverage probabilities for lognormal unbalanced data, K=20, sim=1000, boot=1000

where Y_{ij} is the count of the number of leprosy bacilli for the ith patient in the jth observation($j = 1, 2$ representing before and after treatment). The variables trt_1 and trt_2 are dummy variables for drugs A and B respectively, with $\text{trt}_1 = 1$ if the patient took drug A and $\text{trt}_1 = 0$ otherwise, and $\text{trt}_2 = 1$ if the patient took drug B and $\text{trt}_2 = 0$ otherwise. The binary variable is time which denotes before- (time = 0) or after-treatment (time = 1). Note that there is no treatment effect term in the model, because the mean number of the leprosy bacilli is assumed to be the same among the three groups before treatment (Fitzmaurice et al; 2004). Hence, the main interest is the coefficients $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$, which represent the difference of responses between drug A, B and C respectively. For example, significant $\boldsymbol{\beta}_3$ implies important changes in the responses when comparing drug A and the placebo C.

Estimated standard errors from seven different methods for $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$ from the marginal log-linear regression model are listed in Table 4.2. Also listed are the 95% confidence intervals together with their lengths for $\hat{\boldsymbol{\beta}}_3$ and $\hat{\boldsymbol{\beta}}_4$. We choose $Normal(1, 1)$ and $Poisson(1)$ random variates to illustrate the applications of the smooth bootstrap methods. The working correlation used in the estimation was AR(1). Since there are only 2 observations for each of the 20 patients, only one correlation relationship needs to be identified which means any working correlation structure except independence should be very close to the truth. In this case, MR procedure is expected to perform most efficiently. This can be seen from Table 4.2. MR-norm and MR-pois procedures gave the smallest variance estimates and the narrowest confidence interval especially MR-pois procedure. The ABB method performed poorly for this small data set and failed to declare the significance of $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$. Using the MR-pois procedure, 2000 smooth bootstrap estimates for

$\hat{\boldsymbol{\beta}}_3$ and $\hat{\boldsymbol{\beta}}_4$ were produced, and the histograms with kernel fitting are presented in Figure (4.22). We can see the estimates are approximately normally distributed.

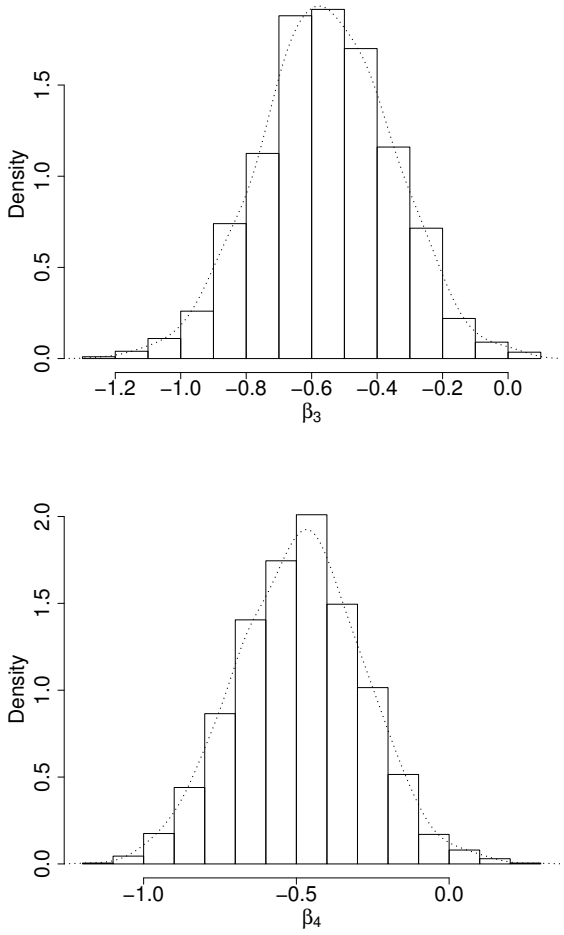

Figure 4.22: Histogram with kernel fitting for β_3 and β_4 estimated by "MR-pois" procedure. The top panel is for β_3 and the bottom is for β_4.

Table 4.2: Parameter estimates with standard errors and length of 95% confidence interval (intervals in brackets) from a Poisson model for the leprosy bacilli data. Estimates "RR-norm" and "MR-norm" represent the robust and model-based resampling estimates using Norm(1,1) for perturbation; "RR-pois" and "MR-pois" represent the two estimates using Pois(1); "ABB" represent all block boostrap estimates by Sherman and le Cessie (1997).

Variable	Methods	Estimate	SE	Z	p-value	length of 95% CI
time_{ij}	LZ	-0.0029	0.1570	-0.018	0.985	
β_2	naive		0.1530	-0.019	0.985	
	ABB		0.1890	-0.015	0.988	
	RR-norm		0.1876	-0.015	0.988	
	RR-pois		0.1693	-0.017	0.986	
	MR-norm		0.1440	-0.020	0.984	
	MR-pois		0.1331	-0.022	0.983	
$\text{time}_{ij} \times \text{trt}_{1i}$	LZ	-0.5626	0.2220	-2.534	0.011	0.871 (-0.998, -0.127)
β_3	naive		0.2498	-2.252	0.024	0.979 (-1.052, -0.073)
	ABB		0.3660	-1.537	0.124	1.434 (-1.563, -0.129)
	RR-norm		0.2580	-2.181	0.029	1.040 (-1.086, -0.046)
	RR-pois		0.2351	-2.392	0.017	0.928 (-1.005, -0.077)
	MR-norm		0.2187	-2.573	0.010	0.877 (-0.990, -0.113)
	MR-pois		0.2010	-2.798	0.005	0.772 (-0.943, -0.171)
$\text{time}_{ij} \times \text{trt}_{2i}$	LZ	-0.4953	0.2342	-2.115	0.034	0.918 (-0.954, -0.036)
β_4	naive		0.2442	-2.028	0.043	0.957 (-0.974, -0.017)
	ABB		0.4038	-1.227	0.220	1.583 (-1.522, 0.061)
	RR-norm		0.2709	-1.828	0.068	1.105 (-1.040, 0.065)
	RR-pois		0.2496	-1.984	0.047	0.978 (-0.964, 0.014)
	MR-norm		0.2220	-2.232	0.026	0.909 (-0.947, -0.038)
	MR-pois		0.2068	-2.395	0.017	0.803 (-0.894, -0.09)

Note: Interpretations of those coefficients:
β_2: difference between before and after treatment for Drug C;
β_3: difference between Drug A and C when after treatment;
β_4: difference between Drug B and C when after treatment.

Chapter 5

Bootstrap methods based on first-term corrected studentized EF statistics

Other than resampling the data in the classical way, more general idea of resampling is to perturb the estimating process in such a way that a bootstrap distribution of the regression parameter estimates could be produced. Parzen *et al.* (1994) perturbed the estimating equation by random values from the distribution of the pivotal estimating functions. Tian *et al.* (2004) perturbed an initial parameter estimate and applied importance resampling technique to generate the final estimate replicates. The smooth bootstrap approaches introduced in Section 3.2 use random weights to different levels of data through the estimating functions: RR procedure adds weights to the subject level and MR procedure adds weights to the measurement level.

Let's consider a natural idea of perturbation, $\sum_{i=1}^{K} U_i = a$, where "a" is some random variate or variable satisfying some specific conditions and can be generated in a resampling way so that many new estimating functions can be constructed and copies of the parameter estimates could be obtained by solving $\sum_{i=1}^{K} U_i = a$. When the estimating function is pivotal or asymptotically pivotal, the method in Parzen *et al.* (1994) can apply. However, for general estimating functions with unknown distributions, the picture for such an "a" is less clear. One natural choice of "a" could be $\sum_{i=1}^{K} U_i = \sum_{i=1}^{K} \widehat{U}_i^*$, where $\widehat{U}_i^*$'s are bootstrap samples from $(\widehat{U}_i)_{i=1}^{K}$ evaluated at $\widehat{\boldsymbol{\theta}}$. Repeatedly solving such equations for different bootstrap samples can result in a bootstrap estimate for the distribution of the parameter estimates. Or, one can directly obtain the bootstrap distribution of the estimating function by repeatedly bootstrapping from $\widehat{U}_i$. Based on the bootstrap distribution of the estimating function (EF), statistical inference on parameter estimates could be done, for example inverting the quantiles of the EF bootstrap distribution could produce confidence intervals for the parameter estimates. In this way only one or two equations need to be solved for the construction of confidence intervals for parameter estimates. Hu and Kalbfleisch (2000) strongly recommended the EF bootstrap and suggested to bootstrap the studentized estimating function, $V^{-1/2}\sum_{i=1}^{K} U_i$, where V is the variance of the estimating function, $\sum_{i=1}^{K} U_i$, which can be estimated for example by the $\sum_{i=1}^{K}(U_i - \overline{U})^2$.

In this chapter we resort to the theory of Edgeworth expansion and first-term correction that render us a better pivotal statistic and lead us to a way of generating the random values for $\sum_{i=1}^{K} U_i = a$ or $V^{-1/2}\sum_{i=1}^{K} U_i = a$. First we derive the results under the assumption that U_i's are identically distributed. The

extension to non-i.i.d. cases is straight forward and is discussed at the end of this chapter.

5.1 A brief introduction to Edgeworth expansion

It's necessary to give a brief introduction to the Edgeworth expansion (more comprehensive explanation on this topic could be found in Hall 1992), because this may define the scope of our discussion. The theory of Edgeworth expansion has a very long history that can date from the first paper by Chebyshev (1890). After a time of decline, a revival of research based on Edgeworth expansion has been seen in recent years, since it is useful for exploring the properties of contemporary statistical methods, among which there is bootstrap method. The Edgeworth expansion involves a great deal of analytical computations for the distribution approximations, or more specifically, normal approximations. Briefly, if $\widehat{\boldsymbol{\theta}}$ is estimated from a sample of size n, and if $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normally distributed with zero mean and variance $\boldsymbol{\sigma}^2$, then the distribution function of $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ may be expanded as a power series in $n^{-1/2}$:

$$P\{n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/\boldsymbol{\sigma} \leq x\} = \boldsymbol{\Phi}(x) + n^{-1/2}p_1(x)\boldsymbol{\phi}(x) + \ldots + n^{-j/2}p_j(x)\boldsymbol{\phi}(x) + \ldots, \tag{5.1.1}$$

where $p_j(x)$'s are polynomials with coefficients depending on the cumulants of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, $\boldsymbol{\Phi}(x)$ and $\boldsymbol{\phi}(x)$ are the standard normal distribution function and density function respectively. This expansion (5.1.1) is the so-called Edgeworth expansion. Conventionally, we are more interested in the terms after $\boldsymbol{\Phi}(x)$, hence

$n^{-1/2}p_1(x)\boldsymbol{\phi}(x)$ is considered as the first term rather than the second term. The Edgeworth series can be best explained via the characteristic functions. Let $S_n = n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/\boldsymbol{\sigma}$. Since S_n is asymptotically normally distributed, its characteristic function χ_n converges to that of the standard normal, $e^{-t^2/2}$. The Edgeworth series can be obtained from the expansion of $\chi_n(t) = \{\chi(t/n^{1/2})\}^n$, where χ is the characteristic function of $Y = (X - \boldsymbol{\mu})/\boldsymbol{\sigma}$.

Most of the results on Edgeworth expansion are established for sample mean problem. Let's consider the sum of independent random variables. Suppose X, X_1, $X_2, \ldots$ are i.i.d. random variables with mean $\boldsymbol{\theta}_0 = \boldsymbol{\mu}$ and finite variance $\boldsymbol{\sigma}^2$. An estimate for $\boldsymbol{\theta}_0$ is the sample average, i.e. $\widehat{\boldsymbol{\theta}} = n^{-1}\sum_{i=1}^{n} X_i$ with variance $n^{-1}\boldsymbol{\sigma}^2$. By the central limit theorem, $S_n = n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/\boldsymbol{\sigma}$ is asymptotically normally distributed with zero mean and unit variance. Many inference about the estimate are based on this normal approximation. For example, confidence intervals for the estimate $\widehat{\boldsymbol{\theta}}$ are constructed based on this normal approximation, $(\widehat{\boldsymbol{\theta}} - n^{-1/2}\boldsymbol{\sigma}z_\alpha, \widehat{\boldsymbol{\theta}} + n^{-1/2}\boldsymbol{\sigma}z_\alpha)$, where z_α is the quantile for significance level α from the standard normal distribution. The quality of this normal approximation is therefore of great importance. It can be investigated via the Edgeworth expansion for S_n, in the form of (5.1.1). However, the expansion for its distribution may not be available for infinite terms. One has to stop the expansion after a given number of terms and the remainder is of smaller order than the previous term,

$$P\{S_n \le x\} = \boldsymbol{\Phi}(x) + n^{-1/2}p_1(x)\boldsymbol{\phi}(x) + \ldots + n^{-j/2}p_j(x)\boldsymbol{\phi}(x) + o(n^{-j/2}). \quad (5.1.2)$$

This expansion is valid for fixed j, as $n \to \infty$. Regularity conditions need to be satisfied for the "asymptotic" expansion (5.1.2), with the remainder of the stated

order uniformly in all x:

$$E(|X|^{j+2}|) < \infty, \qquad \text{and} \qquad \limsup_{|t|\to\infty} |\chi(t)| < 1.$$

The second condition is the well-known *Cramér's condition*. It holds when the distribution of X is nonsingular, or equivalently if that distribution has nondegenerate absolutely continuous components, more specifically, if X has a proper density function. Therefore, it would be convenient to restrict our discussions for the *continuous* estimating functions. For more robust estimating function, such as quantile regression estimating functions or rank estimating functions, Edgeworth expansion may not exist. For those cases, Bahadur representation may be another useful tool and that direction of research might be worth doing.

5.2 First-term corrected EF statistics in i.i.d. cases

The estimating functions in GEE's case are also in forms of the sum of finite independent terms and the correlation which usually makes estimation complicated is within each term itself. Consider the case of a scalar parameter, i.e. $p = 1$. We have the one-dimensional estimating equation:

$$\sum_{i=1}^{K} U_i := \sum_{i=1}^{K} D_i^T V_i^{-1} \boldsymbol{\epsilon}_i = 0.$$

Denote $U_i = D_i^T V_i^{-1} \boldsymbol{\epsilon}_i \,, 1 \le i \le K$ and $\overline{U} = K^{-1} \sum_{i=1}^{K} U_i$. Assuming the independence among all the K subjects, U_i's are independent variables with common mean 0. Further assume they have common variance first, i.e. $\text{var}(U_1) = \cdots =$

$\text{var}(U_K) = \boldsymbol{\sigma}^2$. Hence, $(U_i)_{i=1}^K$ are i.i.d. random variables with mean 0 and finite variance $\boldsymbol{\sigma}^2$. We then have $E(\sum_{i=1}^{K} U_i) = 0$ and $\text{Var}\,(\sum_{i=1}^{K} U_i) = K\boldsymbol{\sigma}^2 \stackrel{def}{=} V^2$.

Consider the pivotal statistic based on the estimating function:

$$S_K = \frac{\sum_{i=1}^{K} U_i}{\sqrt{K}\boldsymbol{\sigma}}, \tag{5.2.1}$$

which is actually the standardized estimating function. Since usually $\boldsymbol{\sigma}$ is unknown, we use $\widehat{V}^2 = \dfrac{K}{K-1}\sum_{i=1}^{K}(U_i - \overline{U})^2$ (or asymptotically equivalent $\sum_{i=1}^{K} U_i^2$) to estimate the variance of $\sum_{i=1}^{K} U_i$ and consider the studentized pivotal statistic instead:

$$T_K = \frac{\sum_{i=1}^{K} U_i}{\widehat{V}}. \tag{5.2.2}$$

Hu and Kalbfleisch (2000) proposed EFB for similar studentized estimating function by bootstrapping its empirical distribution and then inverted the confidence interval to that of the parameter estimates. On the other hand, based on the central limit theorem, the asymptotic normality of T_K can be established. Some simple idea of perturbation to the estimating function arises, such as setting $T_K = a$, where "a" is a random value from the asymptotic normal distribution of T_K. Such procedures could be expected to produce consistent estimates for the parameters. Therefore, we use the Edgeworth expansion to explore the property of this normal approximation for the distribution of T_K. We have found that there are higher-order approximations for this distribution so that the corresponding perturbation procedures can lead to more efficient results. It turns out that the perturbation done to the new statistic whose distribution has higher-order approximation to normality can be viewed as perturbing the original T_K with a $K^{-1/2}$-term.

Appendix III shows the details for deriving the Edgeworth expansion for T_K,

which is shown as follows:

$$P(T_K \leq x) = \boldsymbol{\Phi}(x) + \boldsymbol{\varphi}(x)\left\{K^{-1/2}q_1(x) + K^{-1}q_2(x) + O(K^{-3/2})\right\}, \qquad (5.2.3)$$

where

$$\begin{aligned} q_1(x) &= \frac{1}{6}\boldsymbol{\gamma}\,(2x^2+1); \\ q_2(x) &= x\left\{\frac{1}{12}\boldsymbol{\kappa}\,(x^2-3) - \frac{1}{18}\boldsymbol{\gamma}^2\,(x^4+2x^2-3) + \frac{1}{4}\,(x^2+3)\right\}. \end{aligned}$$

The skewness $\boldsymbol{\gamma}$ and kurtosis $\boldsymbol{\kappa}$ for U_i's, are defined as $EU^3/\boldsymbol{\sigma}^3$ and $EU^4/\boldsymbol{\sigma}^4 - 3$ respectively. This expansion (5.2.3) is valid if *Cramér's condition* is satisfied (particularly, U has an absolutely continuous distribution), and $E(|U|^5) < \infty$ (Hall 1987).

5.2.1 First-term correction

The attempt to remove the first-term in (5.2.3) to improve the normal approximation in higher order is also called correction for skewness, since the skewness is the key factor in the first term. However it cannot be done uniformly in x since the first term is a function of x. When the value of x is fixed, for example as normal quantile, the corrections for skewness can be done for confidence interval constructions. In other words, the first-term corrections are usually done at fixed x. Fortunately enough, in Abramovitch and Singh (1985), a uniform skewness and higher order correction based on Edgeworth expansion is proposed for pivotal statistics and bootstrap methods are applied for the corrected statistics. Now we introduce the important theorem proved by Abramovitch and Singh (1985).

Theorem 5.1 *(Theorem 1 in Abramovitch and Singh, 1985):*

Suppose that T admits an Edgeworth expansion

$$P(T \leq x) = \mathbf{\Phi}(x) + n^{-1/2}\, p\,(F, x) + o\,(n^{-1/2})$$

uniformly in x, where $p\,(F, x)$ is a polynomial in x whose coefficients possibly depend upon F (F is the population from which the sample of size n is obtained, annotated by the author of this thesis). Let $\widehat{p_n}$ be an estimator of $p\,(F, T)$ which satisfies the condition that

$$\textit{for all } \boldsymbol{\epsilon} > 0, P(\,|\,\widehat{p_n} - p\,(F, T)| > \boldsymbol{\epsilon}\,) = o\,(n^{-1/2}) \tag{5.2.4}$$

as $n \to \infty$. Then T_1 defined as

$$T_1 = T + n^{-1/2}\,\widehat{p_n}$$

has the following Edgeworth expansion:

$$P(T_1 \leq x) = \mathbf{\Phi}(x) + o\,(n^{-1/2}) \tag{5.2.5}$$

uniformly in x.

Abramovitch and Singh (1985) suggested to use $p\,(F_n, T)$ as $\widehat{p_n}$, where F_n is the empirical c.d.f. from the sample, if $p\,(F, x)$ is a polynomial in x, its coefficients depend on F only through its first several moments and those moments satisfy certain bounded conditions. The authors gave a modified t-statistic based on this theory and showed the distribution obtained by bootstrapping T_1 beats the normal approximation.

Our settings at the beginning of this section fit the conditions of this theory quite well, therefore we can similarly define:

$$T'_K = T_K + K^{-1/2}\,\widehat{q_1}(T_K), \tag{5.2.6}$$

in which

$$\widehat{q}_1(T_K) = \frac{1}{6}\,\widehat{\boldsymbol{\gamma}}\,(2\,T_K^2 + 1),$$

where $\widehat{\boldsymbol{\gamma}} = \dfrac{1}{K}\sum_{i=1}^{K}(U_i - \overline{U})^3/\widetilde{\boldsymbol{\sigma}}^3$, and $\widetilde{\boldsymbol{\sigma}}^2 = \dfrac{1}{K-1}\sum_{i=1}^{K}(U_i - \overline{U})^2$ (the choice of sample skewness and kurtosis estimators can be found in Joanes and Gill 1998). And it follows from the theorem introduced above that

$$P(T'_K \leq x) = \boldsymbol{\Phi}(x) + O(K^{-1}) \tag{5.2.7}$$

uniformly in x. The K^{-1}-term in the Edgeworth expansion can be further removed in a similar way. Theoretically this way of corrections can be carried out up to j-terms as long as the expansion is valid. However, it may not be worthwhile going beyond a two-term correction since higher order correction involves estimation of higher moments which are generally hard to achieve reasonable accuracy. Even a second-term correction does not seem to be necessary, since for moderate sample size the significant improvement is expected to be seen when the first-term is removed from the Edgeworth expansion, but when further removing the second-term the improvement may not be so significant any more. The simulations in Abramovitch and Singh (1985) showed that the fist-term correction is enough for the approximation of distribution. On the other hand, since the Edgeworth expansion is an asymptotic result based on large number theory as well, the improvement in the sense of approximation of distributions can only be seen in case of moderate sample sizes. In other words, when the sample size is large enough, removing the first-term may not matter much; while when the sample size is too small, this approach also suffers from small sample problems.

5.2.2 Simple perturbation methods for parameter estimates

Thus, we have a higher order normal approximation for T'_K in form of (5.2.6) which has the approximate distribution (5.2.7). A naive interpretation for T'_K is that, it can be viewed as a perturbation to T_K by a $K^{-1/2}$-term. For example, such perturbation could be $T_K - \frac{1}{\sqrt{K}}\,\widehat{q}_1(T_K)$, or $T_K - \frac{1}{\sqrt{K}}\,\widehat{q}_1(T_K) + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is a random variate from standard normal distribution. Some simple perturbation procedures are proposed to generate bootstrap copies of regression estimates:

$$T_K = \boldsymbol{\delta} \quad \text{denoted as } \textit{NM1}; \tag{5.2.8}$$

$$T'_K = \boldsymbol{\delta} \quad \text{denoted as } \textit{NM2}; \tag{5.2.9}$$

$$T_K = -\frac{1}{\sqrt{K}}\,\widehat{q}_1(T_K) + \boldsymbol{\delta} \quad \text{denoted as } \textit{NM3}; \tag{5.2.10}$$

$$T_K = T^*_K \quad \text{denoted as } \textit{EFB}; \tag{5.2.11}$$

$$T'_K = T'^*_K \quad \text{denoted as } \textit{FTCB}; \tag{5.2.12}$$

$$T_K = T'^*_K \quad \text{denoted as } \textit{FTP}; \tag{5.2.13}$$

where $\boldsymbol{\delta}$ is a random variate that can be generated from standard normal distribution repeatedly; T^*_K and T'^*_K are bootstrap copies of T_K and T'_K. The first three perturbation methods (5.2.8), (5.2.9) and (5.2.10) rely on the normal approximation of the studentized EF statistic and the first-term corrected version. The following three (5.2.11), (5.2.12) and (5.2.13) are empirical bootstrap methods like in Hu and Kalbfleisch (2000). We repeatedly solve the perturbed estimating equations and compare the estimation results in terms of variance estimation, MSE of parameter estimates and the confidence intervals. We simulate 1000 balanced normal and Poisson data with sample size $K = 40$ and $n = 4$ using the settings in Chapter 4. The bootstrap time is 1000.

Table 5.1 Parameter estimation and confidence interval construction by simple perturbation methods

	$NM1$	$NM2$	$NM3$	EFB	$FTCB$	FTP	LZ	NV	RR	MR
					normal responses					
(SD ratio-1)* 100	**-0.041**	-0.141	**-0.051**	-3.818	-3.280	-3.140	-0.666	-0.762	0.5167	-0.3058
MSE (e-05)	**4.043**	4.058	**4.051**	4.361	4.315	4.304	4.049	4.142	4.274	4.239
80%	78.6	78.7	78.1	79.3	**79.6**	**79.5**	78.0	78.4	78.5	78.6
CI cover. 90%	**90.1**	**90.1**	89.9	91.3	91.2	91.2	88.9	89.3	-	-
95%	95.6	**95.4**	**95.4**	96.2	95.7	96.4	95.9	96.1	94.6	94.4
					Poisson responses					
(SD ratio-1)* 100	2.740	2.721	2.614	-1.110	**-0.443**	**0.178**	-3.739	-2.399	-0.811	-0.814
MSE * 100	**0.282**	**0.280**	0.282	0.303	0.298	0.298	0.289	0.285	0.290	0.289
80%	79.6	**79.8**	79.3	**80.0**	79.6	79.1	78.0	78.5	78.6	79.6
CI cover. 90%	**89.8**	**90.4**	89.0	88.9	88.9	88.7	87.7	88.4	-	-
95%	**94.8**	94.6	94.2	**94.8**	94.3	94.1	93.5	94.1	94.2	94.3

The simulation results are presented in Table 5.1 compared with the Liang and Zeger (1986) sandwich estimator (denoted as LZ) and the naive estimator (denoted as NV). We also include the results of RR-norm and MR-norm procedure from Chapter 4 for comparison. For each row of the table, the best two results are shown in bold. The first row for normal responses (or Poisson, in the lower part of the table) shows the relative efficiency of standard deviation estimates to the true standard deviation of $\hat{\beta}$ obtained from 1000 simulations. The second row is MSE of the parameter estimates. The following three rows show the 80%, 90% and 95% confidence interval coverage probabilities (in %). For the first six perturbation methods and the last two smooth bootstrap methods, quantile methods are used; for LZ and NV the confidence intervals are constructed as SD-type.

For normal responses, the perturbation methods using normal approximation give the closest variance estimates to the true variance of parameter estimates and their parameter estimates have the smallest MSE. The two smooth bootstrap

procedures also yield satisfactory SD estimates. The perturbation methods using normal approximation give the best coverage for 90% and 95% confidence interval; while for 80% confidence interval, FTCB and FTP methods give the best coverage. Confidence intervals by LZ and NV estimators are less accurate than those by the simple perturbation methods and the smooth bootstrap methods. For Poisson responses, the perturbation methods FTCB and FTP produce the closest variance estimates to the true values but the perturbation methods using normal approximation still yield parameter estimates with smallest MSE. As for the confidence interval coverage, the first two perturbation methods using normal approximation seem to give the best coverage for most of the time. The coverage probabilities by EFB method are very close to the nominal values as well. RR and MR procedures give better coverage than LZ and NV but less accurate than the simple perturbation methods based on the studentized estimating function statistics. The results have shown that the simple perturbation methods based on the normal approximation can generally yield parameter estimates with smaller MSE. All the six simple perturbation methods generally have better confidence interval coverage than the common methods by LZ and NV, and even better than the smooth bootstrap methods which we have observed outperform some classical resampling methods.

Remark 1. The methods discussed above can be easily extended to multidimensional parameter cases, i.e. $p > 1$. Statistical inferences such as the variance estimation and confidence interval construction for the components of the parameter estimates can be done based on the bootstrap copies obtained from repeatedly solving perturbed estimating equations (5.2.8) - (5.2.13). He and Hu (2002) suggested a Markov Chain Marginal Bootstrap (MCMB, available in SAS version

9.1) for the bootstrap procedures for multidimensional parameter problems. The MCMB method solves one component of the parameter by fixing all the other components at each bootstrap time.

Remark 2. We have tried to integrate the idea of smooth bootstrap introduced in Chapter 3 with the bootstrap methods based on studentized estimating functions discussed in this chapter. That is, we can bootstrap the weighted version of T_K or T'_K, for example,

$$T_{K,w} = \frac{\sum_{i=1}^{K} w_i U_i}{\sqrt{\sum_{i=1}^{K} (w_i U_i - \overline{U}_w)^2}},$$

where $\overline{U}_w = K^{-1} \sum_{i=1}^{K} w_i U_i$. Empirical bootstrap methods such as (5.2.11) have been investigated for the estimation of parameters and the corresponding confidence intervals. Unfortunately, we find that except that the point estimates of the parameter are consistent, other results are not satisfactory: the estimated variance is far less than the true value; the confidence interval coverage probabilities are less than the nominal values. The most likely reason for the poor performance of the smoothed estimating function bootstrap is that such perturbation seems to be so liberal that it results in great increase of the variance of the studentized estimating functions. Another possible explanation is that there should be more appropriate choice for the weights, for example data-dependent weights. This direction would be worthy of further investigation.

5.3 Methods for confidence interval construction

In this section we suggest easier ways particularly for the construction of confidence intervals: one relies on the better normal approximation after the first-term

correction; the other one, bootstrap with much less computation. Basically, for both of the two methods, there are two steps to construct confidence intervals for parameter estimates: the fist step is to obtain the confidence intervals for the estimating function $\sum_{i=1}^{K} U_i$ or the studentized estimating function T'_K based on normal approximation or from bootstrap distribution; second, invert the confidence interval of the estimating function or the pivotal statistic to that of the regression parameter estimates $\widehat{\boldsymbol{\beta}}$. By this way, only small number of estimating equations are to be solved and computational time could be reduced greatly. However, if one needs to estimate the variance of the parameter estimates using the inference about T'_K, the simple perturbation methods proposed in the previous section could be applied.

5.3.1 First-term corrected C.I. for EF

Given the improved normal approximation, first-term corrected confidence interval can be constructed for $\sum_{i=1}^{K} U_i$. Then the confidence intervals of regression parameter estimates can be calculated by solving one or two equations, setting $\sum_{i=1}^{K} U_i$ at the the endpoints of its confidence interval. Let

$$a_1 = z_{\alpha/2} - K^{-1/2}\,\widehat{q}_1(z_{\alpha/2}), \tag{5.3.1}$$

$$a_2 = -z_{\alpha/2} - K^{-1/2}\,\widehat{q}_1(-z_{\alpha/2}), \tag{5.3.2}$$

where $z_{\alpha/2} = -\boldsymbol{\Phi}^{-1}(\alpha/2)$. Then the corollary below follows from Theorem 2 in Abramovitch and Singh (1985):

Corollary 5.1 *If T_K admits a one-term Edgeworth expansion and $\widehat{q}_1\,(\,F_K, T_K)$*

is a valid estimate for $q_1(F, T_K)$ *required in Theorem 5.1, then,*

$$P(\sum_{i=1}^{K} U_i < \sum_{i=1}^{K} \widehat{U}_i - a_1 \widehat{V}) = P(T'_K > z_{\alpha/2}) + O(K^{-1});$$

and

$$P(\sum_{i=1}^{K} U_i > \sum_{i=1}^{K} \widehat{U}_i - a_2 \widehat{V}) = P(T'_K < -z_{\alpha/2}) + O(K^{-1}).$$

Therefore the interval

$$\left[\sum_{i=1}^{K} \widehat{U}_i - a_1 \widehat{V}, \sum_{i=1}^{K} \widehat{U}_i - a_2 \widehat{V}\right] \tag{5.3.3}$$

is a confidence interval for $\sum_{i=1}^{K} U_i$ leaving out $\alpha/2 + O(K^{-1})$ probability in either tail. Note that the function $q_1(x)$ is a quadratic function, which results in $\widehat{q}_1(z_{\alpha/2}) = \widehat{q}_1(-z_{\alpha/2})$. Hence the length of this interval (5.3.3) is $2\, z_{\alpha/2}\, \widehat{V}$, the same as the interval $\left[\sum_{i=1}^{K} \widehat{U}_i - z_{\alpha/2}\, \widehat{V}, \sum_{i=1}^{K} \widehat{U}_i + z_{\alpha/2}\, \widehat{V}\right]$ based on the central limit theorem; however the latter leaves lower order probability in both tails. Via this method, a two-sided confidence interval for the parameter estimates, say $\widehat{\boldsymbol{\beta}}$, can be obtained by solving only two equations, namely

$$\sum_{i=1}^{K} U_i = \sum_{i=1}^{K} U_i(\widehat{\boldsymbol{\theta}}) - \widehat{a}_j \widehat{V(\widehat{\boldsymbol{\theta}})}, \quad j = 1, 2, \tag{5.3.4}$$

where "$\widehat{a}_j$" are those a_j defined in (5.3.1) and (5.3.2) evaluated at $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}})$ in the GEE estimation; $U_i(\widehat{\boldsymbol{\theta}})$ and $\widehat{V(\widehat{\boldsymbol{\theta}})}$ are the values of U_i and $\widehat{V}$ evaluated at $\widehat{\boldsymbol{\theta}}$ respectively. The resulting confidence intervals for $\widehat{\boldsymbol{\beta}}$ can also be used for hypothesis testing of the parameter estimates, i.e. reject the null hypothesis when the confidence interval does not intersect the null hypothesis.

5.3.2 Bootstrapping first-term corrected EF statistic

Hu and Kalbfleisch (2000) suggested to approximate the distribution of T_K by bootstrapping. Here we can also apply their EFB idea to the GEE settings, bootstrapping the first-term corrected statistic, T'_K. We suggest to approximate the distribution of the studentized EF statistic T'_K by bootstrapping T'_K. Such approximation is expected to be more accurate than bootstrapping T_K in terms of distribution approximation, such as confidence interval construction. The rationale lying behind is that: the preference for bootstrapping studentized estimating function rather than bootstrapping $\sum_{i=1}^{K} U_i$ is because the studentized EF is an "asymptotic" pivot, and our effort on correcting the first-term in the Edgeworth expansion is to make this "asymptotic" more accurate, namely of higher order. More specifically, the bootstrap procedure is carried out as follows:

- **Step 1:** Obtain the bootstrap distribution of T'_K;
- **Step 2:** Find the quantiles ${T'_K}^*{}_{\alpha/2}$ from T'_K's bootstrap distribution;
- **Step 3:** Solve $T'_K = {T'_K}^*{}_{\alpha/2}$ for the endpoints of confidence intervals for $\widehat{\boldsymbol{\beta}}$.

In the first step, the bootstrap copies of T_K, T_K^* are first obtained in the same way as in Hu and Kalbfleisch (2000),

$$T_K^* = \frac{\sum_{i=1}^{K} \widehat{U}_i^*}{\sqrt{\sum_{i=1}^{K} (\widehat{U}_i^* - \overline{\widehat{U}}^*)^2}}, \tag{5.3.5}$$

where $\widehat{U}_i^*$ are bootstrap samples from $U_i(\widehat{\boldsymbol{\theta}})_{i=1}^K$, and $\overline{\widehat{U}}^*$ is the mean of each bootstrap sample. Then the bootstrap copies of T'_K, ${T'_K}^*$ are obtained using the definition (5.2.6) and the moments are also calculated using $\widehat{U}_i^*$'s. This procedure

is referred to as "First-term corrected EF bootstrap". We also include another similar procedure in the later simulation studies, that is, in the third step, we solve $T_K = {T'_K}^*{}_{\alpha/2}$ instead. This can be viewed as a perturbation to the studentized EF statistic, which is supposed to be asymptotically equivalent to the one by solving $T'_K = {T'_K}^*{}_{\alpha/2}$. But since T_K is of simpler form than T'_K, faster computation by solving $T_K = {T'_K}^*{}_{\alpha/2}$ might be expected. In both bootstrapping procedures, again only two equations are to be solved for constructing a two-sided confidence interval for the regression parameter estimates.

Remark 1. There is one important assumption for the validity of the EF bootstrap methods discussed above. That is the *monotonicity* of the studentized EF function with regard to the parameter estimates. Only when this condition is satisfied, we can invert the quantiles of the bootstrap distribution of studentized EF statistic to those of the parameter estimates. However in GEE models, most of the cases the estimating function is a nonlinear function of the regression parameter, hence it is not convenient to verify this condition analytically. To check this condition, we suggest a practical way which is easy to implement in any model inferences, that is, to observe the behavior of the estimated studentized EF statistic in a neighborhood of the parameter estimates, for example three estimated standard deviations away from the parameter estimates. The parameter estimates together with the estimated standard deviations must be consistent indeed. Then, evaluate the studentized EF statistic at the values in this range, draw a plot of the estimated studentized EF statistic versus those parameter values and observe whether the monotonicity is held in this range. If it shows monotone relationship of the studentized EF and the parameter, the inversion of quantiles from the bootstrap distribution of the studentized EF to the endpoints of confidence

interval for parameter estimated is valid and the bootstrapping methods based on the studentized estimating function discussed above can be carried out.

Remark 2. Consider the two methods discussed in Section 5.3.1 and 5.3.2. The first approach relies on the improved normal distribution; while the second one applies the idea of bootstrapping. The normal approximation to the distribution of T'_K is valid with a remainder of order $o(K^{-1/2})$ (typically $O(K^{-1})$), whereas the bootstrap distribution of T'_K approximates its true distribution with an error term $o(K^{-1})$, which means that the bootstrap gives "an extra step of accuracy". Abramovitch and Singh (1985) argued that this is because the bootstrap distribution of T'_K has the Edgeworth expansion:

$$\mathbf{\Phi}(x) + K^{-1}q_2(F_K, x)\boldsymbol{\phi}(x) + o(K^{-1}) = \mathbf{\Phi}(x) + K^{-1}q_2(F, x)\boldsymbol{\phi}(x) + o(K^{-1}), \text{a.s.},$$

and the validity of this expansion for the bootstrap distribution does not require additional assumptions because all the requirements on F_K are guaranteed for large K by the same requirements on F. Therefore, bootstrapping T'_K basically amounts to a second-order correction, which is the same asymptotic order as the approximation by the second-term correction. In other words, bootstrapping T'_K can achieve the same asymptotic accuracy as a second-term correction of normal approximation, without having to derive the complicated form of the second term in the Edgeworth expansion.

Remark 3. The methods for confidence interval construction stated above focus on scalar parameter cases. For multidimensional parameter cases, confidence regions for the parameter could be obtained using the inversion idea as in Hu and Kalbfleisch (2000). But the component-wise confidence intervals would be more convenient to explain than the confidence regions because the latter are not easy

to describe. Therefore, for multidimensional parameter, the simple perturbation methods discussed in the previous section could be used for variance estimation as well as confidence interval construction.

5.3.3 Simulation studies

In this section, we carry out simulation studies to investigate the two types of methods for the construction of confidence intervals for scalar parameters. We use similar settings which have been used in Chapter 4 for the investigation of smooth bootstrap methods. 1000 Normal and Poisson multivariate correlated responses of different sample sizes 10, 20, 30, 40 and 50 are generated respectively for the confidence interval estimation.

We mainly compare the performances of six procedures for constructing confidence interval for the slope parameter:

- SD-type confidence intervals by Liang and Zeger (1985) sandwich estimator, denoted as LZ;
- SD-type confidence intervals by the naive estimator, denoted as NV;
- first-term corrected EF confidence interval, denoted as $FTCI$;
- the EF bootstrap, i.e. $T_K = T_K^*$, denoted as EFB;
- first-term corrected EF bootstrap, $T_K' = T_K'^*$, denoted as $FTCB$;
- perturbation to the studentized EF statistic, $T_K = T_K'^*$, denoted as FTP.

We also consider different nominal values of confidence intervals, namely 80%, 90% and 95%. The coverage probabilities, the average confidence interval lengths and the standard error of the lengths are tabulated. The tabulated results are from the simulation using correct working correlation which is AR1; in case of misspecification of correlation structure, the results are similar except the poor performance of NV in case of strong correlation.

Before carrying out the last three bootstrap methods, we checked the monotonicity condition following the simple way suggested in Remark 1 in Section 5.3.2. Figure 5.1 shows the plots for sample size 40. The top two plots are for normal responses, where the left one is for the studentized EF statistic T_K and the right one is for the first-term corrected studentized EF statistic T'_K; the bottom two plots are the two studentized EF statistics for Poisson responses. It is very obvious that in such a neighborhood of the parameter estimates those EF statistics are monotone. Therefore we can carry out simulations for the last three bootstrap methods above. The bootstrap time is 1000.

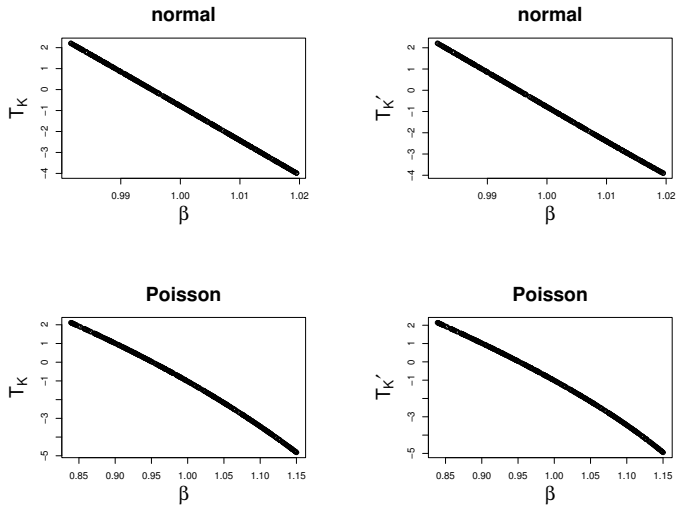

Figure 5.1: Checking the monotonicity of the studentized EF statistics in the regression parameter

Table 5.2 Confidence interval coverage probabilities and average confidence interval lengths (with standard errors), for *normal* responses (in %)

	$K = 50$			$K = 40$			$K = 30$			$K = 20$			$K = 10$		
	80%	90%	95%	80%	90%	95%	80%	90%	95%	80%	90%	95%	80%	90%	95%
LZ	79.2	88.2	92.6	77.4	89.8	94.4	73.8	85.5	92.4	75.6	86.2	92.6	74.7	84.1	89.0
length	*1.46*	*1.88*	*2.24*	*1.62*	*2.08*	*2.48*	*1.85*	*2.37*	*2.83*	*2.25*	*2.88*	*3.43*	*3.03*	*3.88*	*4.63*
(*s.e.*)	(0.146)	(0.187)	(0.223)	(0.187)	(0.240)	(0.287)	(0.250)	(0.322)	(0.383)	(0.365)	(0.468)	(0.558)	(0.749)	(0.962)	(1.114)
NV	80.6	89.0	93.0	78.8	89.6	94.6	75.6	87.3	92.4	77.2	88.7	93.7	76.4	86.8	92.3
length	*1.48*	*1.90*	*2.27*	*1.64*	*2.11*	*2.51*	*1.89*	*2.42*	*2.89*	*2.31*	*2.97*	*3.53*	*3.22*	*4.13*	*4.92*
(*s.e.*)	(0.117)	(0.151)	(0.180)	(0.152)	(0.190)	(0.232)	(0.201)	(0.258)	(0.308)	(0.289)	(0.371)	(0.441)	(0.580)	(0.745)	(0.888)
FTCI	79.8	88.4	93.2	**79.0**	**90.2**	**95.0**	74.3	87.2	92.5	76.2	87.4	93.6	75.8	86.9	91.6
length	*1.48*	*1.91*	*2.09*	*1.65*	*2.12*	*2.33*	*1.88*	*2.43*	*2.67*	*2.31*	*2.99*	*3.29*	*3.24*	*4.20*	*4.63*
(*s.e.*)	(0.149)	(0.192)	(0.211)	(0.190)	(0.244)	(0.268)	(0.256)	(0.330)	(0.362)	(0.378)	(0.487)	(0.537)	(0.835)	(1.048)	(1.114)
EFB	81.4	89.4	93.4	78.6	90.6	95.6	76.6	88.0	93.7	78.2	89.8	94.7	81.0	89.5	93.9
length	*1.51*	*1.95*	*2.15*	*1.68*	*2.18*	*2.39*	*1.95*	*2.53*	*2.79*	*2.44*	*3.20*	*3.54*	*3.71*	*4.97*	*5.47*
(*s.e.*)	(0.158)	(0.208)	(0.230)	(0.200)	(0.263)	(0.291)	(0.274)	(0.357)	(0.395)	(0.416)	(0.547)	(0.614)	(1.019)	(1.342)	(1.441)
FTCB	81.2	89.0	93.2	79.2	90.4	95.2	76.4	87.6	93.2	77.8	89.2	94.3	78.5	88.3	92.8
length	*1.50*	*1.94*	*2.13*	*1.67*	*2.16*	*2.37*	*1.93*	*2.50*	*2.76*	*2.41*	*3.14*	*3.46*	*3.58*	*4.67*	*5.07*
(*s.e.*)	(0.158)	(0.206)	(0.229)	(0.199)	(0.264)	(0.291)	(0.274)	(0.360)	(0.399)	(0.424)	(0.568)	(0.635)	(1.070)	(1.346)	(1.413)
FTP	81.0	88.8	93.0	78.4	90.6	95.8	76.0	87.9	93.6	77.6	89.6	94.5	78.6	88.9	93.7
length	*1.50*	*1.94*	*2.14*	*1.67*	*2.16*	*2.37*	*1.92*	*2.50*	*2.74*	*2.39*	*3.10*	*3.43*	*3.50*	*4.62*	*5.07*
(*s.e.*)	(0.158)	(0.206)	(0.227)	(0.199)	(0.263)	(0.291)	(0.272)	(0.356)	(0.394)	(0.415)	(0.546)	(0.620)	(1.010)	(1.341)	(1.433)

Table 5.2 shows the simulation results for normal responses. The first row entries are the coverage probabilities of the SD-type confidence intervals by the sandwich estimator for different sample sizes and at different level of significance (in %); the second row entries are the average lengths of this type of confidence intervals (in %); and the third row are the standard errors for the lengths (in %). The same table layout is used for the rest five groups of confidence intervals. As expected, as the sample size decreases, the coverage probabilities are further away from the nominal values.

For the normal data, the most accurate coverage of the confidence intervals of the parameter estimates occurs at the moderate sample size 40, where FTCI gives the coverage closest to the nominal values, followed by FTCB and NV. EFB and FTP give similar coverage which is not so good as FTCI and FTCB. LZ gives the worst coverage which can be expected because it underestimates the variance of the parameter estimates. At a larger sample size, the three bootstrap methods and the FTCI together with NV perform similarly, since given a large enough sample size, the difference caused by first-term correction is small. But when the sample size decreases, the coverage by NV and LZ decreases, because both of these SD-type confidence intervals relies on normal approximation of the parameter estimates which needs large sample size to have good performance. The same situation goes with FTCI, since it also relies on a normal approximation, but of the studentized estimating function this time. However, the other three bootstrap methods can still provide satisfactory coverage in case of small sample sizes. EFB seems to be the best among the three, followed by FTP and then FTCB. Higher coverage pays the price of wider length. In almost all the cases, the three bootstrap methods give wider confidence intervals than FTCI, NV and

LZ, where the lengths of FTCI and NV confidence intervals are similar and LZ gives the narrowest again due to the underestimation of the variance. As for the standard errors of the lengths of those confidence intervals, NV is usually the smallest and then LZ and FTCI. The three bootstrap methods have similar larger variation in the length of confidence intervals due to the resampling nature.

Most of the comments for normal responses can also be given exactly for the results in Table 5.3, for Poisson responses. There are some differences worth attention. The naive confidence interval outperforms FTCI, which might mean that for this Poisson data set the normal approximation works well for the parameter estimates but may not work well for the studentized estimating function. But the three bootstrap methods work satisfactorily and better than NV especially in the cases of small sample sizes. When the sample size is 40, FTCB gives the most accurate coverage and also for the sample size 30. When the sample size is 20 or smaller, EFB performs slightly better. But the lengths of FTCB confidence intervals are always narrower than those of EFB's. Therefore, we conclude the bootstrap methods based on studentized EF statistic T_K or first-term corrected statistic T'_K can provide much better confidence intervals than the NV and LZ methods. When the sample size is moderate, bootstrap methods based on T'_K can further improve the confidence interval coverage and always give narrower confidence intervals with satisfactory coverage when compared with bootstrap methods based on the statistic without first-term correction.

Table 5.3 Confidence interval coverage probabilities and average confidence interval lengths (with standard errors), for *Poisson* responses (in %)

	$K = 50$			$K = 40$			$K = 30$			$K = 20$			$K = 10$		
	80%	90%	95%	80%	90%	95%	80%	90%	95%	80%	90%	95%	80%	90%	95%
LZ	79.8	91.4	96.4	78.8	88.4	93.6	77.1	87.5	93.7	76.1	85.9	91.7	73.4	83.6	89.1
length	*12.13*	*15.57*	*18.55*	*13.44*	*17.25*	*20.55*	*15.27*	*19.60*	*23.36*	*31.72*	*23.73*	*28.28*	*24.73*	*31.74*	*37.82*
(*s.e.*)	(1.38)	(1.78)	(2.12)	(1.70)	(2.18)	(2.60)	(2.21)	(2.84)	(3.39)	(3.17)	(4.07)	(4.85)	(6.01)	(7.71)	(9.19)
NV	80.8	91.8	97.0	77.6	88.6	95.0	78.5	88.9	94.9	77.3	87.8	92.6	76.6	86.8	93.1
length	*12.20*	*15.66*	*18.66*	*13.66*	*17.54*	*20.90*	*15.70*	*20.15*	*24.02*	*28.04*	*24.57*	*29.28*	*26.46*	*33.96*	*40.46*
(*s.e.*)	(1.16)	(1.50)	(1.78)	(1.40)	(1.80)	(2.14)	(1.83)	(2.35)	(2.80)	(2.80)	(3.60)	(4.29)	(5.08)	(6.52)	(7.77)
FTCI	80.0	91.4	95.8	79.2	88.8	93.0	77.5	87.4	92.9	76.4	86.6	92.0	74.5	84.4	90.7
length	*12.35*	*15.93*	*17.15*	*13.72*	*17.73*	*19.06*	*15.71*	*20.34*	*21.84*	*34.65*	*24.98*	*26.75*	*26.81*	*35.11*	*37.46*
(*s.e.*)	(1.44)	(1.87)	(1.96)	(1.77)	(2.32)	(2.41)	(2.36)	(3.11)	(3.22)	(3.46)	(4.59)	(4.73)	(7.10)	(9.45)	(9.65)
EFB	82.2	92.6	97.6	80.4	89.2	94.8	79.3	89.4	95.2	79.4	88.8	93.9	81.5	90.6	94.2
length	*12.54*	*16.22*	*17.88*	*14.03*	*18.17*	*20.06*	*16.20*	*21.06*	*23.32*	*37.04*	*26.52*	*29.62*	*30.56*	*40.98*	*45.57*
(*s.e.*)	(1.49)	(1.95)	(2.15)	(1.90)	(2.48)	(2.69)	(2.46)	(3.22)	(3.54)	(3.70)	(4.87)	(5.37)	(8.43)	(10.96)	(12.03)
FTCB	81.6	92.8	97.0	**79.8**	**89.8**	**94.9**	79.2	89.3	94.9	78.8	87.9	93.7	78.9	89.3	94.0
length	*12.50*	*16.17*	*17.77*	*13.98*	*18.09*	*19.89*	*16.11*	*20.92*	*23.04*	*31.19*	*25.11*	*28.86*	*29.38*	*38.46*	*42.34*
(*s.e.*)	(1.52)	(2.05)	(2.22)	(2.05)	(2.64)	(2.80)	(2.70)	(3.49)	(3.72)	(4.11)	(5.36)	(5.70)	(9.18)	(11.09)	(11.86)
FTP	80.8	92.4	96.4	80.6	89.4	94.4	79.4	89.0	94.9	78.5	87.6	93.5	79.0	88.9	93.7
length	*12.48*	*16.13*	*17.90*	*13.95*	*18.03*	*20.03*	*16.03*	*20.78*	*23.14*	*37.62*	*25.84*	*28.91*	*28.98*	*38.32*	*42.81*
(*s.e.*)	(1.49)	(1.98)	(2.23)	(1.92)	(2.50)	(2.79)	(2.47)	(3.25)	(3.66)	(3.76)	(4.99)	(5.60)	(8.63)	(11.27)	(12.38)

5.4 Direct generalization to non-i.i.d. cases

In the discussions above we assume the components of the estimating function are independent and identically distributed. For longitudinal data, the assumption of independence of the components is generally valid, but the components may not have equal variances. Liu (1988) discussed the bootstrap procedures under some non-i.i.d. models and justified the validity of bootstrapping non-i.i.d. data via Edgeworth expansions. Based on this result, the bootstrap methods proposed in this Chapter are valid in non-i.i.d. cases in GEE settings. Thus, such methods can be easily generalized to general longitudinal data analysis. This robustness to heterogeneity has already been verified in some of the simulation studies shown in the previous sections. Liu (1988) obtained the results via similar arguments using Edgeworth expansion. Below we give out some important results and the corresponding conditions needed for the validity of those bootstrap methods for non-i.i.d. cases.

Theorem 5.2 *(Theorem 1 in Liu 1988.)*

Let $U_1, \ldots, U_n$ *be a set of independent random observations and let* $\boldsymbol{\mu}_i$ *and* $\boldsymbol{\sigma}_i^2$ *denote respectively, the mean and the variance of* U_i, $i = 1, \ldots, n$. *Also let* $\nu_n^2 = \sum_{i=1}^{n} \boldsymbol{\sigma}_i^2/n$, *and* $V_n^2 = \sum_{i=1}^{n}(U_i - \bar{U}_n)^2/n$. *If (i)* $\lim_{n\to\infty} \sum_{i=1}^{n}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_n)^2/n = 0$, *(ii)* $\liminf_{n\to\infty} \nu_n^2 > 0$ *and (iii)* $E|U_i|^{2+\delta} \leq M < \infty$, *for some* $\delta > 0$ *and for all* i, *then*

$$\lim_{n\to\infty} ||P^*(\sqrt{n}(\bar{U}_n^* - \bar{U}_n) \leq x) - P(\sqrt{n}(\bar{U}_n - \bar{\boldsymbol{\mu}}_n) \leq x)||_\infty = 0, \quad a.s.$$

(P^* stands for the bootstrap probability and $||\,||_\infty$ stands for the sup-norm over x.)

Theorem 5.2 states the validity of bootstrap methods for non-i.i.d. random series, such as U_i's in GEE situations in terms of "first order limit". Liu (1988) further investigated the second order asymptotic properties in case of the sample mean, which can be directly applied in the generalized estimating functions. Such properties are established via the Edgeworth expansion for the studentized statistics.

$$P(\sqrt{n}\,\frac{\bar{U}_n - \bar{\boldsymbol{\mu}}_n}{V_n} \le x) = \boldsymbol{\Phi}(x) + \frac{\bar{\boldsymbol{\mu}}_{3,n}}{6\nu_n^3\sqrt{n}}(2x^2+1)\boldsymbol{\phi}(x) + o(n^{-1/2}); \quad (5.4.1)$$

$$P^*(\sqrt{n}\,\frac{\bar{U}_n^* - \bar{U}_n}{V_n^*} \le x) = \boldsymbol{\Phi}(x) + \frac{\hat{\kappa}_3}{6V_n^3\sqrt{n}}(2x^2+1)\boldsymbol{\phi}(x) + o(n^{-1/2}); \quad (5.4.2)$$

where $V_n^{*2} = \sum_{i=1}^{n}(U_i^* - \bar{U}_n^*)^2/n$, $\bar{\boldsymbol{\mu}}_{3,n} = \sum_{i=1}^{n} E(U_i - \boldsymbol{\mu}_i)^3/n$, $\hat{\kappa}_3 = \sum_{i=1}^{n}(U_i - \bar{U}_n)^3/n$. If $\kappa_3 - \bar{\boldsymbol{\mu}}_{3,n} \to 0$, *a.s.* as $n \to \infty$, "there is a total $n^{-1/2}$-term correction by the bootstrap in the studentized case." The existence of those Edgeworth expansions for the studentized statistics needs some conditions. The second part of Theorem 2 in Liu (1988) stated those conditions.

Theorem 5.3 *(The second part of Theorem 2 in Liu 1988)*

The expansion (5.4.1) and (5.4.2) will hold under conditions (a), (b) and (c):

(a) There exists a nonlattice distribution H with mean 0 and variance 1, and a sequence k_n with $k_n/log(n) \to \infty$, such that k_n of the population G_i's are of the form $G_i(x) = H((x - \boldsymbol{\mu}_i)/\boldsymbol{\sigma}_i)$ with the $\boldsymbol{\sigma}_i$'s bounded away from 0, where G_i is the distribution of U_i.

(b) $E|U_i|^{6+\delta} \le M_2 < \infty$ for some $\delta > 0$.

(c) $\liminf_{n\to\infty} \nu_n^2 > 0$ and $\sum_{i=1}^{n}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_n)^2/n = o(n^{-1/2})$.

It is easy to verify that the conditions required in Theorem 5.3 is valid in GEE settings (as long as the estimating functions are continuous). Therefore, the bootstrap methods proposed in Section 5.2 and 5.3 for the parameter estimation are valid in general GEE settings.

Chapter 6

Discussions

This chapter contains an outline of the results discussed in the thesis, the problems that we encountered and the topics for further research.

6.1 Concluding remarks

Longitudinal data are becoming increasingly common in many research areas. Consequently, the number of applications of GEE methods will continue to increase due to its flexibility and robustness. The resampling methods will definitely add greater versatility to the GEE models because of the increasing computational ability.

In this thesis, basically two types of resampling methods are proposed for the inference of regression parameters in GEE models. The first approach is "smooth bootstrap," a random perturbation to the estimating algorithms that provides a simple way to yield bootstrapped copies of parameter estimates. Two versions of

the smooth bootstrap method are proposed and investigated analytically and via Monte Carlo simulation in Chapter 3 and 4. One is robust to the misspecification of the within-subject correlation structure and the other one is model-based which is more efficient when the covariance model is correctly specified. Smooth bootstrap helps to give more accurate variance estimates and confidence intervals for different types of data and sample sizes compared with the commonly used "sandwich" estimators and some classical resampling methods applied to longitudinal data.

The second type of approach proposed in this thesis applies the resampling scheme to the estimating function instead of the parameter estimates. Chapter 5 proposes several simple perturbation methods to the estimating functions for the parameter estimation. Particularly, one bootstrap scheme and a first-term corrected confidence interval method are proposed for confidence interval construction. In the bootstrapping scheme, the first-term corrected studentized EF is the subject to be bootstrapped and only two equations need to be solved for the construction of confidence intervals for parameter estimates. The first-term corrected studentized estimating function statistic is obtained via the Edgeworth expansion. Bootstrapping this first-term corrected statistic will give higher order distribution approximation when compared with bootstrapping the studentized EF without first-term correction and lead to improved confidence intervals.

6.2 Topics for further research

As with any research project, there are usually more unanswered questions at the end of the study than at the beginning. This is certainly true here. Some of the major interests for future work are summarized as follows.

First, the smooth bootstrap methods presented here capture the the first order and second order moments. The choice of such random weights may depend on other factors not considered here. This type of data independent weights may not be the optimal choices to accommodate the characteristics of the longitudinal data. Further research should be undertaken to develop the use of data dependent (or data-driven) weights. Moreover, the application of such resampling methods in the bias estimation for the GEE parameters is also a subject for further investigation.

Second, in multidimensional parameter cases, the higher order approximation for the distribution of the studentized estimating functions could be reached under methods similar to those discussed and such investigation might involve multivariate Edgeworth expansions (Chamber 1967). Other types of estimating functions could be considered as well. Li (1998) considered an optimal estimating equation based on the first three cumulants for scalar parameter. Li added an extra term of order $n^{-1/2}$ to the quasilikelihood function by including the information of the skewness of observations when there is departure from exponential family. The resampling methods proposed in this thesis could be applied to this method and generalization to multidimensional parameters would be interesting.

Third, the smooth bootstrap methods and the estimating function bootstrap can also be applied to the estimation of nuisance parameters. There is actually

much room for the improvement of nuisance parameter estimation since we have also observe inefficient estimation for the scale and correlation parameter in our simulation studies especially for non-normal data. By improving the accuracy of estimating the nuisance parameters, the efficiency of the estimation of regression parameters will also be improved.

As an alternative to the mean regression model, the application of quantile regression model to longitudinal data has received extensive attention in recent years. Due to the non-smooth nature of the estimating function in the quantile regression, the methods available for mean regression model cannot be directly applied. Parzen *et al.* (1994) and Tian *et al.* (2004) discussed some resampling methods for such non-smooth estimating functions. The application of resampling methods like the smooth bootstrap proposed in this thesis may be possible for direct application, but the choices of weights will have a greater effect on the consistency of the resampling procedures given the non-smooth nature of the estimating functions. The idea of estimating function bootstrap may also be applied to such estimation procedure but the conditions for its validity need more careful investigation. As for the first-term correction, the Edgeworth expansion does not exist for non-smooth function. Therefore one may try Bahadur representation in such direction.

Last but not least, there is a need for studying mechanisms for generating clustered correlated data. We introduced some procedures for generating correlated Poisson data and Lognormal data in Chapter 4. In addition, more general mechanisms to generate other types of correlated data is an interesting subject for further study. These mechanisms should correspond to phenomena observed

in practice so that the conclusions from the simulation studies would be more adaptable to the real data sets.

Appendix I

I.1 About sandwich estimators

In this part, we give analytical argument for the variability of $\mathbf{V}_{MD}$, $\mathbf{V}_{KC}$, $\mathbf{V}_P$ and $\mathbf{V}_{LZ}$ following the way in Pan (2001). As in Pan (2001), we treat the $\mathbf{D}_i$, $\mathbf{V}_i$ and $\mathbf{A}_i$ as fixed, and compare the variabilities of those four covariance estimators by comparing their middle factors since they have the same two outside factors:

$$\begin{aligned}
M_{LZ} &= \sum_{i=1}^{K} D_i^T V_i^{-1} \widehat{\boldsymbol{\epsilon}}_i \widehat{\boldsymbol{\epsilon}}_i^T V_i^{-1} D_i \,; \\
M_{MD} &= \sum_{i=1}^{K} D_i^T V_i^{-1} (I_i - H_{ii})^{-1} \widehat{\boldsymbol{\epsilon}}_i \widehat{\boldsymbol{\epsilon}}_i^T (I_i - H_{ii}^T)^{-1} V_i^{-1} D_i \,; \\
M_{KC} &= \sum_{i=1}^{K} D_i^T V_i^{-1} (I_i - H_{ii})^{-1/2} \widehat{\boldsymbol{\epsilon}}_i \widehat{\boldsymbol{\epsilon}}_i^T (I_i - H_{ii}^T)^{-1/2} V_i^{-1} D_i \,; \\
M_P &= \sum_{i=1}^{K} D_i^T V_i^{-1} A_i^{1/2} \left(\sum_{j=1}^{K} A_j^{-1/2} \widehat{\boldsymbol{\epsilon}}_j \widehat{\boldsymbol{\epsilon}}_j^T A_j^{-1/2} / K \right) A_i^{1/2} V_i^{-1} D_i \,.
\end{aligned}$$

Pan (2001) has already proved that under mild regularity conditions, asymptotically $cov\big\{\mathrm{vec}(M_{LZ})\big\} \geq cov\big\{\mathrm{vec}(M_P)\big\}$. We further show that:

$$cov\big\{\mathrm{vec}(M_P)\big\} \leq cov\big\{\mathrm{vec}(M_{LZ})\big\} \leq cov\big\{\mathrm{vec}(M_{KC})\big\} \leq cov\big\{\mathrm{vec}(M_{MD})\big\}$$

Proof. Using the Kronecker product $\otimes$ and $\mathrm{vec}(\cdot)$ operations, we have:

$$\begin{aligned}
\mathrm{vec}(M_{LZ}) &= \sum_{i=1}^{K} C_i \ \mathrm{vec}(\epsilon_i \epsilon_i^T) \\
\mathrm{vec}(M_{MD}) &= \sum_{i=1}^{K} C_i \ \mathrm{vec}\left((I_i - H_{ii})^{-1} \epsilon_i \epsilon_i^T (I_i - H_{ii}^T)^{-1}\right) \\
\mathrm{vec}(M_{KC}) &= \sum_{i=1}^{K} C_i \ \mathrm{vec}\left((I_i - H_{ii})^{-1/2} \epsilon_i \epsilon_i^T (I_i - H_{ii}^T)^{-1/2}\right) \\
\mathrm{vec}(M_P) &= \sum_{i=1}^{K} C_i \ \mathrm{vec}\left(A_i^{1/2} \sum_{j=1}^{K} A_j^{-1/2} \widehat{\boldsymbol{\epsilon}}_j \widehat{\boldsymbol{\epsilon}}_j^T \ A_j^{-1/2}/K \ A_i^{1/2}\right)
\end{aligned}$$

and then,

$$\begin{aligned}
\mathrm{cov}\big\{\mathrm{vec}(M_{LZ})\big\} &= \sum_{i=1}^{K} C_i \, \Omega_i \, C_i^T \\
\mathrm{cov}\big\{\mathrm{vec}(M_{MD})\big\} &= \sum_{i=1}^{K} C_i \, G_i^2 \ \Omega_i \, G_i^{2\,T} \ C_i^T \\
\mathrm{cov}\big\{\mathrm{vec}(M_{KC})\big\} &= \sum_{i=1}^{K} C_i \, G_i \ \Omega_i \, G_i^T \ C_i^T \\
\mathrm{cov}\big\{\mathrm{vec}(M_P)\big\} &= \sum_{i=1}^{K} C_i \left\{ F_i \sum_{j=1}^{K} \left(\frac{1}{K^2} F_j^{-1} \Omega_j F_j^{-1} \right) F_i \right\} C_i^T
\end{aligned}$$

where $C_i = (D_i^T V_i^{-1}) \otimes (D_i^T V_i^{-1})$, $F_i = A_i^{1/2} \otimes A_i^{1/2}$, $G_i = (I_i - H_{ii})^{-1/2} \otimes (I_i - H_{ii})^{-1/2}$ and $\Omega_i = \mathrm{cov}\big\{\mathrm{vec}(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T)\big\}$.

Hence,

$$\begin{aligned}
\mathrm{cov}\big\{\mathrm{vec}(M_{LZ})\big\} - \mathrm{cov}\big\{\mathrm{vec}(M_{MD})\big\} &= \sum_{i=1}^{K} C_i \left(\Omega_i - G_i \ \Omega_i \ G_i^T\right) C_i^T \leq 0, \\
\mathrm{cov}\big\{\mathrm{vec}(M_{LZ})\big\} - \mathrm{cov}\big\{\mathrm{vec}(M_{KC})\big\} &= \sum_{i=1}^{K} C_i \left(\Omega_i - G_i^2 \, \Omega_i \, G_i^{2\,T}\right) C_i^T \leq 0, \\
\mathrm{cov}\big\{\mathrm{vec}(M_{KC})\big\} - \mathrm{cov}\big\{\mathrm{vec}(M_{MD})\big\} &= \sum_{i=1}^{K} C_i \ G_i \left(\Omega_i - G_i \, \Omega_i \, G_i^T\right) G_i^T C_i^T \leq 0, \\
\mathrm{cov}\big\{\mathrm{vec}(M_P)\big\} - \mathrm{cov}\big\{\mathrm{vec}(M_{LZ})\big\} &= \sum_{i=1}^{K} C_i \left(F_i \frac{\sum_{j=1}^{K} F_j^{-1} \Omega_j F_j^{-1}}{K^2} F_i - \Omega_i \right) C_i^T \leq 0,
\end{aligned}$$

since the elements of H_{ii} are very close to 0. The last expression is from Pan(2001).

I.2 Proof of Proposition 3.1

Proof. Fix $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$, apply Taylor expansion to the two estimating equations at the true value of regression estimator, $\boldsymbol{\beta}_0$,

$$\begin{aligned} g_1(\widehat{\boldsymbol{\beta}}) &= \sum_{i=1}^{K} C_i(\boldsymbol{\beta}_0)\,\boldsymbol{\epsilon}_{i0} + \sum_{i=1}^{K} \left.\frac{\partial(C_i\,\boldsymbol{\epsilon}_i)}{\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p\,(K^{1/2}); \\ g_2(\widetilde{\boldsymbol{\beta}}) &= \sum_{i=1}^{K} C_i(\bar{\boldsymbol{\beta}})\,\boldsymbol{\epsilon}_{i0} + \sum_{i=1}^{K} \left\{-C_i(\bar{\boldsymbol{\beta}})\,D_i(\boldsymbol{\beta}_0)\right\} (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p\,(K^{1/2}). \end{aligned}$$

Denote $\cdot_{i0} = \cdot_i(\boldsymbol{\beta}_0)$, for example the $p\times p$ matrix $M_{i0} = D_i^T(\boldsymbol{\beta}_0)\,V_i^{-1}(\boldsymbol{\beta}_0)\,D_i(\boldsymbol{\beta}_0)$, the $p \times n_i$ matrix $C_{i0} = C_i(\boldsymbol{\beta}_0)$, and $\dot{C_{i0}} = \left.\dfrac{\partial C_i}{\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$. Since C_{i0} is $p \times n_i$, the form of $\dfrac{\partial C_i}{\partial\boldsymbol{\beta}^T}$ is expressed as

$$\begin{pmatrix} \dfrac{\partial C_{i01}}{\partial\boldsymbol{\beta}^T} \\ \vdots \\ \dfrac{\partial C_{i0j}}{\partial\boldsymbol{\beta}^T} \\ \vdots \\ \dfrac{\partial C_{i0n_i}}{\partial\boldsymbol{\beta}^T} \end{pmatrix},$$

where C_{i0j} is of $p\times 1$, the jth column of C_{i0}, $1 \le j \le n_i$. Hence $\dfrac{\partial C_{i0j}}{\partial\boldsymbol{\beta}^T}$ of dimension $p \times p$ and $\dfrac{\partial C_i}{\partial\boldsymbol{\beta}^T}$ is of dimension $(p \times n_i) \times p$, or equivalently the derivative can be expressed as another way

$$\begin{pmatrix} \dfrac{\partial C_{i01}^T}{\partial\boldsymbol{\beta}^T} \\ \vdots \\ \dfrac{\partial C_{i0l}^T}{\partial\boldsymbol{\beta}^T} \\ \vdots \\ \dfrac{\partial C_{i0p}^T}{\partial\boldsymbol{\beta}^T} \end{pmatrix},$$

where C_{i0l} is of $n_i \times 1$, the lth row of C_{i0}, $1 \le l \le p$. Hence $\dfrac{\partial C_{i01}^T}{\partial\boldsymbol{\beta}^T}$ is of dimension $n_i \times p$ and there are p blocks of these matrices.

Therefore, $\dfrac{\partial(C_i\,\boldsymbol{\epsilon}_i)}{\partial\boldsymbol{\beta}^T} = M_{i0} + \boldsymbol{\epsilon}_{i0}^T\,\dot{C_{i0}}$, where the second term of the right-hand

side is of the form

$$\begin{pmatrix} \boldsymbol{\epsilon}_{i0}^T \dfrac{\partial C_{i01}^T}{\partial \boldsymbol{\beta}^T} \\ \vdots \\ \boldsymbol{\epsilon}_{i0}^T \dfrac{\partial C_{i0l}^T}{\partial \boldsymbol{\beta}^T} \\ \vdots \\ \boldsymbol{\epsilon}_{i0}^T \dfrac{\partial C_{i0p}^T}{\partial \boldsymbol{\beta}^T} \end{pmatrix},$$

resulting in a $p \times p$ matrix.

Since $\bar{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + o_p(1)$, let $\bar{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is a vector such that $||\boldsymbol{\delta}|| = o_p(1)$. Hence $C_i(\bar{\boldsymbol{\beta}}) = C_i(\boldsymbol{\beta}_0 + \boldsymbol{\delta}) = C_{i0} + \dot{C_{i0}}' \boldsymbol{\delta} + o_p(1)$. The expression of $\dot{C_{i0}}' \boldsymbol{\delta}$ is as follows

$$\begin{pmatrix} \dfrac{\partial C_{i01}}{\partial \boldsymbol{\beta}^T} \boldsymbol{\delta} & \cdots & \dfrac{\partial C_{i0j}}{\partial \boldsymbol{\beta}^T} \boldsymbol{\delta} & \cdots & \dfrac{\partial C_{i0n_i}}{\partial \boldsymbol{\beta}^T} \boldsymbol{\delta} \end{pmatrix},$$

which gives a $p \times n_i$ matrix, the same dimension as C_{i0}.

Therefore one can have the expression for $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \left\{ \sum_{i=1}^K M_{i0} - \sum_{i=1}^K \boldsymbol{\epsilon}_{i0}^T \dot{C_{i0}} \right\}^{-1} \left\{ \sum_{i=1}^K C_{i0}\, \boldsymbol{\epsilon}_{i0} \right\} + o_p(K^{-1/2}) \quad (6.2.1)$$

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= \boldsymbol{\beta}_0 + \left\{ \sum_{i=1}^K C_i(\bar{\boldsymbol{\beta}})\, D_i(\boldsymbol{\beta}_0) \right\}^{-1} \left\{ \sum_{i=1}^K C_i(\bar{\boldsymbol{\beta}})\, \boldsymbol{\epsilon}_{i0} \right\} + o_p(K^{-1/2}) \\ &\approx \boldsymbol{\beta}_0 + \left\{ \sum_{i=1}^K M_{i0} + \sum_{i=1}^K \dot{C_{i0}}' \boldsymbol{\delta} D_{i0} \right\}^{-1} \left\{ \sum_{i=1}^K C_{i0}\, \boldsymbol{\epsilon}_{i0} \right\} + o_p(K^{-1/2}) \quad (6.2.2) \end{aligned}$$

Now compare the first part of the second term in (6.2.1) and (6.2.2). Note that $\boldsymbol{\epsilon}_i = \boldsymbol{\epsilon}_{i0} - D_{i0}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ and $E(\boldsymbol{\epsilon}_i) = 0$. In a close neighborhood of $\boldsymbol{\beta}_0$, say $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| = \boldsymbol{\delta}$, $\boldsymbol{\epsilon}_{i0} = D_{i0}\boldsymbol{\delta} + o_p(1)$. Then the conclusion follows. □

Appendix II

In this appendix, we give the justification for the multivariate correlated data generation in the simulations and the relevant R code.

III.1 Justification of correlated data generation

Method 1): AR-1 Poisson with $\mathbf{Var}(\boldsymbol{\mu}) = \boldsymbol{\mu} + \boldsymbol{\lambda}\boldsymbol{\mu}^2$

• Input: N, ni, $\beta = c(0,1)^T$, $\boldsymbol{\rho}$, "corstr", $x_j = \log(\text{runif(ni)}+1\text{:ni})$, and $\boldsymbol{\mu}_j = \exp(x_j^T\beta)$.

• Generate $y_i \leftarrow \text{rpois}(\boldsymbol{\mu}_i * \boldsymbol{\xi}_i)$, where $\boldsymbol{\xi}_i$ is a gamma random variable with mean 1 and variance $\boldsymbol{\lambda}$. To generate correlated $\boldsymbol{\xi}$, first

1) Calculate

$$\text{Cov}(y) = \text{diag}(\sqrt{\text{Var}(\boldsymbol{\mu})}\ )\ \times \text{AR1}(\boldsymbol{\rho}, \text{ni})\ \times \text{diag}(\sqrt{\text{Var}(\boldsymbol{\mu})}\ ),$$

and correlation matrix of $\boldsymbol{\xi}$

$$R_0 = \text{Cor}(\boldsymbol{\xi}) = \sqrt{\text{Cov}(\boldsymbol{\xi})/\boldsymbol{\lambda}} = \sqrt{\{\text{diag}(1/\boldsymbol{\mu})\ \times (\text{Cov}(y) - \text{diag}(\boldsymbol{\mu}))\ \times \text{diag}(1/\boldsymbol{\mu})\}/\boldsymbol{\lambda}};$$

2) With $d = 2/\boldsymbol{\lambda}$, generate $\boldsymbol{\xi}_i = \sum_{j=1}^{d} z_{ji}^2/d$, where each of the four ni-variate $(z_{j.})$ is independently sampled from $N(0, R_0)$. That is,

$$\begin{array}{cccc} z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \\ z_{31} & z_{32} & z_{33} & z_{34} \\ z_{41} & z_{42} & z_{43} & z_{44} \\ \vdots & \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \boldsymbol{\xi}_{i1} & \boldsymbol{\xi}_{i2} & \boldsymbol{\xi}_{i3} & \boldsymbol{\xi}_{i4} \end{array} \quad \begin{array}{l} \leftarrow N(0, R_0) \\ \leftarrow N(0, R_0) \\ \leftarrow N(0, R_0) \\ \leftarrow N(0, R_0) \\ \\ \\ \rightarrow \boldsymbol{\xi}_i \end{array}$$

Method 2): AR-1 Poisson with $\mathbf{Var}(\boldsymbol{\mu}) = \boldsymbol{\mu}/(1-\boldsymbol{\rho}^2)$

• Input: N, ni, $\boldsymbol{\rho}$, and $\boldsymbol{\mu}_j = (\boldsymbol{\mu}_{j1}, \cdots, \boldsymbol{\mu}_{j,ni})^{\mathrm{T}}, j = 1, \ldots, N$

• Generate $y_j = (y_{j1}, \ldots, y_{j,ni})^T$ for subject j (omit j for convenience):

1) $y_0 \leftarrow \text{rpois}(\boldsymbol{\mu}_1)$

$\mathbf{E}(y_0) = \boldsymbol{\mu}_1$, $\text{Var}(y_0) = \boldsymbol{\mu}_1$;

2) $y_1 \leftarrow \text{rpois}\left(\boldsymbol{\mu}_1 + \dfrac{\boldsymbol{\rho}}{\sqrt{1-\boldsymbol{\rho}^2}}(y_0 - \boldsymbol{\mu}_1)\right)$

$\mathbf{E}(y_1|y_0) = \text{Var}(y_1|y_0) = \boldsymbol{\mu}_1 + \dfrac{\boldsymbol{\rho}}{\sqrt{1-\boldsymbol{\rho}^2}}(y_0 - \boldsymbol{\mu}_1)$;

$\mathbf{E}(y_1) = \boldsymbol{\mu}_1$,

$\text{Var}(y_1) = \mathbf{E}(V(y_1|y_0)) + V(E(y_1|y_0)) = \boldsymbol{\mu}_1 + \dfrac{\boldsymbol{\rho}^2}{1-\boldsymbol{\rho}^2}\boldsymbol{\mu}_1 = \dfrac{1}{1-\boldsymbol{\rho}^2}\boldsymbol{\mu}_1$;

3) $y_2 \leftarrow \text{rpois}\left(\boldsymbol{\mu}_2 + \boldsymbol{\rho}\sqrt{\dfrac{\boldsymbol{\mu}_2}{\boldsymbol{\mu}_1}}(y_1 - \boldsymbol{\mu}_1)\right)$

$\mathbf{E}(y_2) = \mathbf{E}(\boldsymbol{\mu}_2 + \boldsymbol{\rho}\sqrt{\dfrac{\boldsymbol{\mu}_2}{\boldsymbol{\mu}_1}}(y_1 - \boldsymbol{\mu}_1)) = \boldsymbol{\mu}_2$,

$\text{Var}(y_2) = \boldsymbol{\mu}_2 + \boldsymbol{\rho}^2\dfrac{\boldsymbol{\mu}_2}{\boldsymbol{\mu}_1}\text{Var}(y_1) = \dfrac{1}{1-\boldsymbol{\rho}^2}\boldsymbol{\mu}_2$,

$\text{Cov}(y_2, y_1) = \mathbf{E}(\boldsymbol{\mu}_2 y_1 + \boldsymbol{\rho}\sqrt{\dfrac{\boldsymbol{\mu}_2}{\boldsymbol{\mu}_1}}y_1^2 - \boldsymbol{\rho}\sqrt{\boldsymbol{\mu}_1\boldsymbol{\mu}_2}y_1) - \boldsymbol{\mu}_2\boldsymbol{\mu}_1 = \dfrac{\boldsymbol{\rho}}{1-\boldsymbol{\rho}^2}\sqrt{\boldsymbol{\mu}_2\boldsymbol{\mu}_1}$,

$\text{Cor}(y_2, y_1) = \boldsymbol{\rho}$;

4) $y_3 \leftarrow \text{rpois}\left(\boldsymbol{\mu}_3 + \boldsymbol{\rho}\sqrt{\dfrac{\boldsymbol{\mu}_3}{\boldsymbol{\mu}_2}}(y_2 - \boldsymbol{\mu}_2)\right)$

$\mathbf{E}(y_3) = \boldsymbol{\mu}_3$, $\text{Var}(y_2) = \dfrac{1}{1-\boldsymbol{\rho}^2}\boldsymbol{\mu}_3$,

$\text{Cov}(y_3, y_2) = \dfrac{\boldsymbol{\rho}}{1-\boldsymbol{\rho}^2}\sqrt{\boldsymbol{\mu}_3\boldsymbol{\mu}_2}$, $\text{Cor}(y_3, y_2) = \boldsymbol{\rho}$,

$\text{Cov}(y_3, y_1) = \mathbf{E}[\mathbf{E}(\boldsymbol{\mu}_3 y_1 + \boldsymbol{\rho}\sqrt{\dfrac{\boldsymbol{\mu}_3}{\boldsymbol{\mu}_2}}y_1(y_2 - \boldsymbol{\mu}_2))|y_1] - \boldsymbol{\mu}_3\boldsymbol{\mu}_1 = \dfrac{\boldsymbol{\rho}^2}{1-\boldsymbol{\rho}^2}\sqrt{\boldsymbol{\mu}_3\boldsymbol{\mu}_1}$,

$\text{Cor}(y_3, y_1) = \boldsymbol{\rho}^2$;

5) repeatedly generate $y_{ni} \leftarrow \text{rpois}\left(\boldsymbol{\mu}_{ni-1} + \boldsymbol{\rho}\sqrt{\dfrac{\boldsymbol{\mu}_{ni}}{\boldsymbol{\mu}_{ni-1}}}(y_{ni-1} - \boldsymbol{\mu}_{ni-1})\right)$

$$\mathbf{E}(y_{ni}) = \boldsymbol{\mu}_{ni},\ \mathrm{Var}(y_{ni}) = \frac{1}{1-\boldsymbol{\rho}^2}\boldsymbol{\mu}_{ni},\ \mathrm{Cor}(y_r, y_s) = \boldsymbol{\rho}^{|r-s|}.$$

Method 3): AR-1 Poisson with $\mathbf{Var}(\boldsymbol{\mu}) = a\boldsymbol{\mu} + b\boldsymbol{\mu}^2$

• Input: N, ni, $\beta = c(0,1)^T$, $\boldsymbol{\rho}$, $x_j = \log(\text{runif(ni)}+1{:}\text{ni})$, and $\boldsymbol{\mu}_j = \exp(x_j^T\beta)$.

• Generate $y_j = (y_{j1}, \ldots, y_{j,ni})^T$ for subject j (omit j for convenience):

1) $y_1 \leftarrow \text{rpois}(\boldsymbol{\mu}_1 * t_1)$, where $t_1 \leftarrow \text{rgamma}(a_1, a_1)$,

$$a_1 = \frac{\boldsymbol{\mu}_1}{(a-1) + b * \boldsymbol{\mu}_1}.$$

$\mathbf{E}(y_1) = E(\boldsymbol{\mu}_1 t_1) = \boldsymbol{\mu}_1$,

$\mathrm{Var}(y_1) = E(\boldsymbol{\mu}_1 t_1) + V(\boldsymbol{\mu}_1 t_1) = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^2/a_1 = a\boldsymbol{\mu}_1 + b\boldsymbol{\mu}_1^2$;

2) $y_2 \leftarrow \text{rpois}(\eta_2 * t_2)$, where $t_2 \leftarrow \text{rgamma}(a_2, a_2)$,

$$\eta_2 = \boldsymbol{\mu}_2 + \boldsymbol{\rho}\sqrt{\frac{a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2}{a\boldsymbol{\mu}_1 + b\boldsymbol{\mu}_1^2}}(y_1 - \boldsymbol{\mu}_1)$$

and

$$a_2 = \frac{\boldsymbol{\mu}_2^2 + \boldsymbol{\rho}^2(a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2)}{(a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2)(1-\boldsymbol{\rho}^2) - \boldsymbol{\mu}_2}.$$

$$\begin{aligned}
\mathbf{E}(y_2) &= E(\eta_2)E(t_2) = \boldsymbol{\mu}_2, \\
\mathrm{Var}(y_2) &= E(\eta_2 t_2) + V(\eta_2 t_2) \\
&= E(\eta_2)E(t_2) + E(\eta_2^2)E(t_2^2) - (E(\eta_2))^2(E(t_2))^2 \\
&= \boldsymbol{\mu}_2 + \left(\boldsymbol{\mu}_2^2 + \boldsymbol{\rho}^2\frac{a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2}{a\boldsymbol{\mu}_1 + b\boldsymbol{\mu}_1^2}(a\boldsymbol{\mu}_1 + b\boldsymbol{\mu}_1^2)\right)\left(1 + \frac{1}{a_2}\right) - \boldsymbol{\mu}_2^2 \\
&= a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2, \\
\mathrm{Cov}(y_2, y_1) &= E(y_2 y_1) - \boldsymbol{\mu}_2\boldsymbol{\mu}_1 \\
&= \left(E(y_2)E(y_1)E(t_2) + \boldsymbol{\rho}\sqrt{\frac{a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2}{a\boldsymbol{\mu}_1 + b\boldsymbol{\mu}_1^2}}E(t_2)(E(y_1^2) - \boldsymbol{\mu}_1 E(y_1))\right) - \boldsymbol{\mu}_2\boldsymbol{\mu}_1 \\
&= \boldsymbol{\rho}\sqrt{(a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2)(a\boldsymbol{\mu}_1 + b\boldsymbol{\mu}_1^2)}, \\
\mathrm{Cor}(y_2, y_1) &= \boldsymbol{\rho}.
\end{aligned}$$

3) $y_3 \leftarrow \text{rpois}(\eta_3 * \text{rgamma}(a_3, a_3)$, where

$$\eta_3 = \boldsymbol{\mu}_3 + \boldsymbol{\rho}\sqrt{\frac{a\boldsymbol{\mu}_3 + b\boldsymbol{\mu}_3^2}{a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2}}(y_2 - \boldsymbol{\mu}_2)$$

and

$$a_3 = \frac{\boldsymbol{\mu}_3^2 + \boldsymbol{\rho}^2(a\boldsymbol{\mu}_3 + b\boldsymbol{\mu}_3^2)}{(a\boldsymbol{\mu}_3 + b\boldsymbol{\mu}_3^2)(1-\boldsymbol{\rho}^2) - \boldsymbol{\mu}_3}.$$

$$\begin{aligned}
\mathbf{E}(y_3) &= \boldsymbol{\mu}_3, \\
\mathrm{Var}(y_3) &= a\boldsymbol{\mu}_3 + b\boldsymbol{\mu}_3^2, \\
\mathrm{Cov}(y_3, y_2) &= \boldsymbol{\rho}\sqrt{(a\boldsymbol{\mu}_3 + b\boldsymbol{\mu}_3^2)(a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2)}, \\
\mathrm{Cor}(y_3, y_2) &= \boldsymbol{\rho}, \\
\mathrm{Cov}(y_3, y_1) &= E(y_3 y_1) - \boldsymbol{\mu}_3\boldsymbol{\mu}_1 \\
&= \boldsymbol{\mu}_3\boldsymbol{\mu}_1 + \boldsymbol{\rho}\sqrt{\frac{a\boldsymbol{\mu}_3 + b\boldsymbol{\mu}_3^2}{a\boldsymbol{\mu}_2 + b\boldsymbol{\mu}_2^2}}(E(y_1 y_2) - \boldsymbol{\mu}_2\boldsymbol{\mu}_1) - \boldsymbol{\mu}_3\boldsymbol{\mu}_1 \\
&= \boldsymbol{\rho}^2\sqrt{(a\boldsymbol{\mu}_3 + b\boldsymbol{\mu}_3^2)(a\boldsymbol{\mu}_1 + b\boldsymbol{\mu}_1^2)}, \\
\mathrm{Cor}(y_3, y_1) &= \boldsymbol{\rho}^2.
\end{aligned}$$

4) repeatedly generate $y_j \leftarrow \mathrm{rpois}(\eta_j * \mathrm{rgamma}(a_j, a_j))$, where

$$\eta_j = \boldsymbol{\mu}_j + \boldsymbol{\rho}\sqrt{\frac{a\boldsymbol{\mu}_j + b\boldsymbol{\mu}_j^2}{a\boldsymbol{\mu}_{j-1} + b\boldsymbol{\mu}_{j-1}^2}}(y_{j-1} - \boldsymbol{\mu}_{j-1})$$

and

$$a_j = \frac{\boldsymbol{\mu}_j^2 + \boldsymbol{\rho}^2(a\boldsymbol{\mu}_j + b\boldsymbol{\mu}_j^2)}{(a\boldsymbol{\mu}_j + b\boldsymbol{\mu}_j^2)(1-\boldsymbol{\rho}^2) - \boldsymbol{\mu}_j},$$

$\mathbf{E}(y_i) = \boldsymbol{\mu}_i$, $\mathrm{Var}(y_i) = a\boldsymbol{\mu}_i + b\boldsymbol{\mu}_i^2$,

$\mathrm{Cov}(y_r, y_s) = \boldsymbol{\rho}^{|r-s|}\sqrt{(a\boldsymbol{\mu}_r + \boldsymbol{\mu}_r^2)(a\boldsymbol{\mu}_s + b\boldsymbol{\mu}_s^2)}$, $\mathrm{Cor}(y_r, y_s) = \boldsymbol{\rho}^{|r-s|}$.

III.2 R code for correlated multivariate generation

In this appendix, we list the R code for generating multivariate correlated data for reference. Other R code for the simulation studies are available upon request from the author.

1. Correlated normal data

```
gennorm<-function(K, n, b, cv) {   y<-matrix(NA,K*n,4)
    dimnames(y)<-list(NULL,c("id","yi","xi","time"))
```

```
    for (i in 1:K) {
    xi<- runif(n)+c(1:n)
    id<-rep(i,n)
    time<-1:n
    mu<-b[1]+b[2]*xi
    err<-t(cv)%*%rnorm(n)
    yi<-mu+sigma*err
    y[((i-1)*n+1):(i*n),] <-cbind(id,yi,xi,time)
    }
y }
```

2. Correlated normal data

```
#simulate a single correlated lognormal error vector
mlnormal<-function(n, mu, V) { y<-rep(NA,n)
  a<-log(1+diag(V)/mu^2)
  m<-log(mu)-a/2
  mm<-outer(m,m,FUN="+")
  aa<-outer(a,a,FUN="+")
  xmu<-mm+aa/2
  A<-log(V+exp(xmu))-xmu
  y<-m+t(chol(A))%*%rnorm(n)
  return(exp(y))
} #simulation for a multivariate correlated lognormal data
Mlnormal<-function(K, n, btrue, Rtrue) # mu is mean of yi, V is
variance of yi {  y<-matrix(NA,K*n,4)
   dimnames(y)<-list(NULL,c("id","yi","xi","time"))
   for (i in 1:K)
     { time<-1:n
       xi<-log(time+runif(n))
       id<-rep(i,n)
       eta<-btrue[1]+btrue[2]*xi
       mui<-exp(eta)          # mean
       Atrue<-diag(mui)             # s.e., variance is mu^2
       Vtrue<-Atrue%*%Rtrue%*%Atrue   # covariance
       yi<-mlnormal(n,mui,Vtrue)
       y[((i-1)*n+1):(i*n),] <-cbind(id,yi,xi,time)
      }
y }
```

3. Correlated overdispersed Poisson data

```
# generate overdispered Poisson, mean=exp(beta*x),var=mu+lamda*mu^2
```

```
rcnorm<-function(mu, Sigma)  y<-mu+t(chol(Sigma))%*%rnorm(length(mu))

varf1<-function(mean) var<-mean+lambda*mean^2

mvrpois<-function(K, n, b, rho) {   y<-matrix(NA,K*n,4)
    dimnames(y)<-list(NULL,c("id","yi","xi","time"))
    for (i in 1:K)    {
        id<-rep(i,n)
        time<-1:n
        xi<-log(runif(n)+1:n)
        mui<-exp(b[1]+b[2]*xi)
        var<-varf1(mui)     #overdispersion
        covy<-diag(sqrt(var))%*%ar1(rho,n)%*%diag(sqrt(var))# cov. of y
        coverr<-diag(1/mui)%*%(covy-diag(mui))%*%diag(1/mui)# cov. of error
        R0<-sqrt(coverr/lambda)
        #eigen(R0)
        z<-matrix(0,n,n)
        err<-yi<-rep(0,n)
        d=2/lambda
        for(j in 1:d) z[,j]<-rcnorm(rep(0,n), R0)
        #for(j in 1:d) z[,j]<-mvrnorm(n,rep(0,n),R0)
        for(j in 1:n) {
            err[j]=sum(z[j,]^2) / 4
            yi[j]<-rpois(1,mui[j]*err[j])
            }
        y[((i-1)*n+1):(i*n),] <-cbind(id,yi,xi,time)
        }
y }


## generate overdispersed Posisson data with AR(1) correlation
structure with var = mu/(1-rho^2)

varf2<-function(mu) mu/(1-rho^2)

genarpoi2 <- function(N, ni, rho, muvec) {   y0 <-rpois(N,muvec[,1])
    out <- matrix(NA, nr = N, nc = ni)
    out[,1] <- rpois(N, muvec[,1]+ rho*sqrt(1/(1-rho^2))*(y0-muvec[,1]))
    for(j in 2:ni)
    out[,j] <- rpois(N, muvec[,j]+ rho*sqrt(muvec[,j]
    /muvec[,j-1])*(out[,j-1]-muvec[,j-1]))
    out
}
```

```
AR1poisson<-function(K, n, b, rho) {   y<-matrix(NA,K,n)
    x<-matrix(log(rep(1:n,rep(K,n))+runif(K*n)), nc=n) # the covariate matrix
    #x <- t(matrix((1:n)*1 + 0*rnorm(K*n),nr=n))         # covariate
    mu <- exp(b[1]+b[2]*x)        # mean matrix
    y<-genarpoi2(K, n, rho, mu)
    dd<- as.data.frame(cbind(y=as.numeric(t(y)),x=as.numeric(t(x))))
    dd$id<-as.vector(t(matrix(rep(c(1:K),n),nc=n)))
    dd
 }

## A more general approach to generate AR1 Poisson data using gamma
frailty model;

varf3<-function(mu) g1*mu+g2*mu^2

genarpoi <- function(K, n, rho,g1,g2)
 {  y<-x<-mu<-id<-matrix(NA,K,n)
    id<-rep(1:K, rep(n,K))
    x1<-log(1+runif(K))
    x[,1]<-x1
    mu1<-exp(x1)  # mu=exp(b0+b1*x), b0=0, b1=1
    mu[,1]<-mu1
    a1<-mu1/(g1-1+g2*mu1)
    y1<-rpois(K,mu1*rgamma(K,a1,a1))
    y[,1]<-y1
    for( j in 2:n){
        x2<-log(j+runif(K))
        x[,j]<-x2
        mu2<-exp(x2)
        mu[,j]<-mu2
        eta<-mu[,j]+rho*sqrt((g1*mu[,j]+g2*(mu[,j])^2)
        /(g1*mu[,j-1]+g2*(mu[,j-1])^2))*(y[,j-1]-mu[,j-1])
        a2<-((mu[,j])^2+rho^2*(g1*mu[,j]+g2*(mu[,j])^2))
        /((g1*mu[,j]+g2*(mu[,j])^2)*(1-rho^2)-mu[,j])
        y[,j]<-rpois(K,eta*rgamma(K,a2,a2))
    }
    yy<-(as.numeric(t(y)))
    xx<-(as.numeric(t(x)))
    df<-matrix(NA,K*n,3)
    dimnames(df)<-list(NULL,c("id","yi","xi"))
    as.data.frame(cbind(id,yy,xx))
}
```

Appendix III

Edgeworth expansion for T_K

For the derivation of the Edgeworth expansion for the statistic

$$T_K = \frac{\sum_{i=1}^{K} U_i}{\sqrt{\sum_{i=1}^{K} U_i^2}} = \frac{\sum_{i=1}^{K} U_i}{\widehat{V}},$$

we begin with its characteristic function:

$$\boldsymbol{\phi}_K(t) = E e^{itT_K} = \exp\left\{ it\boldsymbol{\gamma}_{1,K} + \frac{(it)^2}{2}\boldsymbol{\gamma}_{2,K} + \frac{(it)^3}{6}\boldsymbol{\gamma}_{3,K} + \dots \right\}$$

First prepare some results for the moments of U ($\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ are respectively the skewness and kurtosis of U). The right row for the raw moments of average $\overline{U}$ is obtained from the left row and the following relationships: $E(U_i U_j U_k) = 0$ and $E(U_i U_j U_k U_l) = 0$ if one index differs from the rest; $E(U_i^2 U_j^2) = \boldsymbol{\sigma}^4$.

$$\begin{array}{ll} E(U) = 0 & E(\overline{U}) = 0 \\ E(U^2) = \boldsymbol{\sigma}^2 & E(\overline{U}^{\,2}) = K^{-1}\,\boldsymbol{\sigma}^2 \\ E(U^3) = \boldsymbol{\gamma}\,\boldsymbol{\sigma}^3 & E(\overline{U}^{\,3}) = K^{-2}\,\boldsymbol{\gamma}\,\boldsymbol{\sigma}^3 \\ E(U^4) = (\boldsymbol{\kappa} + 3)\,\boldsymbol{\sigma}^4 & E(\overline{U}^{\,4}) = K^{-3}\,\boldsymbol{\sigma}^4\,(\boldsymbol{\kappa} + 3K) \end{array}$$

Some other useful preliminary results:

- $E(\sum_{i=1}^{K} U_i^2) = K\boldsymbol{\sigma}^2 = V^2$;
- $E[(\sum_{i=1}^{K} U_i^2)^2] = K(\boldsymbol{\kappa} + 3)\,\boldsymbol{\sigma}^4 + K(K-1)\boldsymbol{\sigma}^4 = K\boldsymbol{\sigma}^4\,(\boldsymbol{\kappa} + 2 + K)$;
- $E[\sum_{i=1}^{K} U_i\,(\sum_{i=1}^{K} U_i^2)] = E[\sum_{i=1}^{K} U_i\,(\sum_{i=1}^{K} U_i)^2] = K\,E(U^3) = K\boldsymbol{\gamma}\,\boldsymbol{\sigma}^3$;

- $E[\sum_{i=1}^{K} U_i^2 \, (\sum_{i=1}^{K} U_i)] = K\, E(U^3) = K\boldsymbol{\gamma}\, \boldsymbol{\sigma}^3$;

- $E[\sum_{i=1}^{K} U_i^2 \, (\sum_{i=1}^{K} U_i^2)] = E[\sum_{i=1}^{K} U_i^2 \, (\sum_{i=1}^{K} U_i)^2] = K\, \boldsymbol{\sigma}^4 \, (\boldsymbol{\kappa} + 2 + K)$.

- For the variance estimator $\widehat{V}^2 = \dfrac{K}{K-1} \sum_{i=1}^{K} (U_i - \overline{U})^2$:

$$E(\widehat{V}^2) = K\boldsymbol{\sigma}^2 = V^2\,;$$
$$E\left[\sum_{i=1}^{K} U_i\, \widehat{V}^2\right] = \boldsymbol{\gamma}\, \boldsymbol{\sigma}^3\,;$$
$$E\left[\left(\sum U_i\right)^2 \, \widehat{V}^2\right] = K\, \boldsymbol{\sigma}^4 \, (\boldsymbol{\kappa} + 2 + K)\,.$$

We write $\sum_{i=1}^{K}$ as $\sum$ later on to make the presentation clearer. Express T_K as

$$\frac{\sum U_i}{V} \left(1 + \frac{\widehat{V}^2 - V^2}{V^2}\right)^{-1/2},$$

to apply Taylor expansion, then the first four raw moments of T_K can be computed as follows:

$$\begin{aligned}
\boldsymbol{\mu}'_{1,\,K} = E(T_K) &= E\,\frac{\sum U_i}{V}\left[1 - \frac{1}{2}\left(\frac{\widehat{V}^2 - V^2}{V^2}\right) + O(K^{-3/2})\right] \\
&= E\left(\frac{\sum U_i}{V}\right) - \frac{1}{2V}\, E\left[\sum U_i \left(\frac{\widehat{V}^2 - V^2}{V^2}\right)\right] + O(K^{-3/2}) \\
&= -\frac{1}{2}\, K^{-1/2}\, \boldsymbol{\gamma} + O(K^{-3/2});
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\mu}'_{2,\,K} = E(T_K^2) &= E\,\frac{(\sum U_i)^2}{V^2}\left[\left(1 + \frac{\widehat{V}^2 - V^2}{V^2}\right)^{-1}\right] \\
&= E\,\frac{\sum U_i^2}{V^2} - \frac{1}{V^4}\, E\left[\sum U_i^2\, (\widehat{V}^2 - V^2)\right] \\
&\qquad + \frac{1}{V^6}\, E\left[(\sum U_i)^2\, (\widehat{V}^2 - V^2)^2\right] + O(K^{-2}) \\
&= 1 - K^{-1}(\boldsymbol{\kappa} + 2) + K^{-1}(3 + 2\boldsymbol{\gamma}^2 + \boldsymbol{\kappa} + 2) + O(K^{-2}) \\
&= 1 + K^{-1}\,(2\,\boldsymbol{\gamma}^2 + 3) + O(K^{-2});
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\mu}'_{3,K} = E(T_K^3) &= E\,\frac{(\sum U_i)^3}{V^3}\left[\left(1+\frac{\widehat{V}^2-V^2}{V^2}\right)^{-3/2}\right] \\
&= E\,\frac{(\sum U_i)^3}{V^3}\left[1-\frac{3}{2}\frac{\widehat{V}^2-V^2}{V^2}+O(K^{-3/2})\right] \\
&= -\frac{7}{2}\,K^{-1/2}\boldsymbol{\gamma}+O(K^{-3/2});
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\mu}'_{4,K} = E(T_K^4) &= E\,\frac{(\sum U_i)^4}{V^4}\left[\left(1+\frac{\widehat{V}^2-V^2}{V^2}\right)^{-2}\right] \\
&= E\,\frac{(\sum U_i)^4}{V^4}\left[1-2\left(\frac{\widehat{V}^2-V^2}{V^2}\right)+3\left(\frac{\widehat{V}^2-V^2}{V^2}\right)^2+O(K^{-2})\right] \\
&= (3+K^{-1}\boldsymbol{\kappa})-2\,K^{-1}[4\,\boldsymbol{\gamma}^2+6\,(\boldsymbol{\kappa}+2)]+30K^{-1} \\
&\qquad +3\,K^{-1}[12\,\boldsymbol{\gamma}^2+3\,(\boldsymbol{\kappa}+2)]+\ldots \\
&= 3+K^{-1}(28\,\boldsymbol{\gamma}^2-2\,\boldsymbol{\kappa}+24)+O(K^{-2}),
\end{aligned}$$

where $\boldsymbol{\gamma}=EU^3/\boldsymbol{\sigma}^3$ and $\boldsymbol{\kappa}=EU^4/\boldsymbol{\sigma}^4-3$ are the third and fourth cumulants of U.

Therefore by the relationship between cumulants and raw moments, one can obtain:

$$\begin{aligned}
\boldsymbol{\gamma}_{1,K} &= \boldsymbol{\mu}'_{1,K}=\frac{1}{2}\,K^{-1/2}\boldsymbol{\gamma}+\,O(K^{-3/2});\\[2ex]
\boldsymbol{\gamma}_{2,K} &= \boldsymbol{\mu}'_{2,K}-\boldsymbol{\mu}'^{\,2}_{1,K}=1+\frac{1}{4}\,K^{-1}\,(7\,\boldsymbol{\gamma}^2+12)+\,O(K^{-2});\\[2ex]
\boldsymbol{\gamma}_{3,K} &= \boldsymbol{\mu}'_{3,K}-3\boldsymbol{\mu}'_{2,K}\boldsymbol{\mu}'_{1,K}+2\boldsymbol{\mu}'^{\,3}_{1,K}=-2K^{-1/2}\boldsymbol{\gamma}+\,O(K^{-3/2});\\[2ex]
\boldsymbol{\gamma}_{4,K} &= \boldsymbol{\mu}'_{4,K}-4\boldsymbol{\mu}'_{1,K}\boldsymbol{\mu}'_{3,K}-3\boldsymbol{\mu}'^{\,2}_{2,K}+12\boldsymbol{\mu}'^{\,2}_{1,K}\boldsymbol{\mu}'_{2,K}-6\boldsymbol{\mu}'^{\,4}_{1,K}\\
&= K^{-1}\,(12\,\boldsymbol{\gamma}^2-2\,\boldsymbol{\kappa}+6)+\,O(K^{-2})
\end{aligned}$$

Substituting the above cumulants in the characteristic function of T_K, it follows by

Taylor expansion of exponential function that:

$$\begin{aligned}
\boldsymbol{\phi}_K(t) &= \exp\left\{-\frac{t^2}{2} + K^{-1/2}\left[-\frac{\boldsymbol{\gamma}}{2}it - \frac{\boldsymbol{\gamma}}{3}(it)^3\right]\right. \\
&\qquad \left. + K^{-1}\left[\frac{(it)^2}{8}\left(7\boldsymbol{\gamma}^2 + 12\right) + \frac{(it)^4}{4!}(12\,\boldsymbol{\gamma}^3 - 2\,\boldsymbol{\kappa} + 6)\right] + \ldots\right\} \\
&= e^{-t^2/2}\left\{1 + K^{-1/2}\left[-\frac{\boldsymbol{\gamma}}{2}it - \frac{\boldsymbol{\gamma}}{3}(it)^3\right]\right. \\
&\qquad + K^{-1}\left[\frac{(it)^2}{8}\left(7\boldsymbol{\gamma}^2 + 12\right) + \frac{(it)^4}{4!}(12\,\boldsymbol{\gamma}^3 - 2\,\boldsymbol{\kappa} + 6)\right. \\
&\qquad\qquad \left.\left. + \frac{1}{2}\left(-\frac{\boldsymbol{\gamma}}{2}it - \frac{\boldsymbol{\gamma}}{3}(it)^3\right)^2\right] + \ldots\right\} \\
&= e^{-t^2/2}\left\{K^{-1/2}\left[-\frac{\boldsymbol{\gamma}}{2}it - \frac{\boldsymbol{\gamma}}{3}(it)^3\right]\right. \\
&\qquad \left. + K^{-1}\left[(it)^2\left(\boldsymbol{\gamma}^2 + \frac{3}{2}\right) + (it)^4\left(-\frac{1}{12}\,\boldsymbol{\kappa} + \frac{2}{3}\,\boldsymbol{\gamma}^2 + \frac{1}{4}\right) + (it)^6\,\frac{\boldsymbol{\gamma}^2}{18}\right] + \ldots\right\}.
\end{aligned}$$

By Fourier transform, it can be obtained that:

$$\begin{aligned}
P(T_K \le x) &= \boldsymbol{\Phi}(x) + \boldsymbol{\varphi}(x)\left\{K^{-1/2}\left[\frac{\boldsymbol{\gamma}}{2}H_0(x) + \frac{\boldsymbol{\gamma}}{3}H_2(x)\right] + K^{-1}\left[-\left(\boldsymbol{\gamma}^2 + \frac{3}{2}\right)H_1(x)\right.\right. \\
&\qquad \left.\left. + \left(\frac{1}{12}\,\boldsymbol{\kappa} - \frac{2}{3}\,\boldsymbol{\gamma}^2 - \frac{1}{4}\right)H_3(x) - \frac{\boldsymbol{\gamma}^2}{18}\,H_5(x)\right] + \ldots\right\} \\
&= \boldsymbol{\Phi}(x) + \boldsymbol{\varphi}(x)\left\{K^{-1/2}q_1(x) + K^{-1}q_2(x) + O(K^{-3/2})\right\},
\end{aligned}$$

where $H_i(x)$ are Hermite Polynomials defined by

$$\boldsymbol{\phi}(x)H_j(x) = (-1)^j\left[\frac{d^j}{dx^j}\boldsymbol{\phi}(x)\right],$$

and

$$\begin{aligned}
q_1(x) &= \frac{1}{6}\,\boldsymbol{\gamma}\,(2x^2 + 1); \\
q_2(x) &= x\left\{\frac{1}{12}\,\boldsymbol{\kappa}\,(x^2 - 3) - \frac{1}{18}\,\boldsymbol{\gamma}^2\,(x^4 + 2x^2 - 3) + \frac{1}{4}\,(x^2 + 3)\right\}.
\end{aligned}$$

This expansion of remainder $O(K^{-j/2})$ is valid if U has an absolutely continuous distribution and $E(|U|^{j+2}) < \infty$ (Hall 1987).

Bibliography

[1] Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Annals of Statistics* **13,** 116-132.

[2] Chambers, J. M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika* **54,** 367-383.

[3] Chebyshev, P. L. (1890). Sur deux théorèmes relatifs aux probabilités. *Acta Mathematica* **14,** 305-315.

[4] Claeskens, G., Aerts, M. and Molenberghs, G. (2003). A quadratic bootstrap method and improved estimation in logistic regression. *Statistics & Probability Letters* **61,** 383-394.

[5] Crowder, M. J. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measurements. *Biometrika* **82,** 407-410.

[6] Davis, C. S. (2001). *Statistical Methods for the Analysis of Repeated Measurements.* Springer-Verlag, New York.

[7] Davis, R. A., Dunsmuir, W. T. M. and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* **87,** 491-505.s

[8] Diggle, P. J., Heagerty, P., Liang, K. - Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data.* Oxford University Press, Oxford.

[9] Dobson, A. J. (1990). *Introduction to Generalized Linear Models.* Oxford University Press, Oxford.

[10] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7,** 1-26.

[11] Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics.

[12] Efron, B. (1992). Six questions raised by the bootstrap. *Exploring the limits of bootstrap*, Lepage, R. and Billard, L. (Ed.). Wiley, pp. 99-126.

[13] Efron, B. and Tibshirani, R. J. (1994). *An Introduciton to the Bootstrap.* Chapman and Hall, New York.

[14] Emond, M. J., Ritz, J. and Oakes, D. (1997). Bias in GEE estimates from misspsecified models for longitudinal data. *Communications in Statistics, Theory and Methods* **26,** 15-32.

[15] Fisher, R. A. (1935). *The Design of Experiments.* Oliver & Boyd, Edinburgh.

[16] Fitzmaurice, G., Laird, N. and Ware, J. (2004). *Applied Longitudinal Analysis.* John Wiley and Sons, New York.

[17] Good, P. I. (2001). *Resampling Methods: a practical guide to data analysis..* Birkhauser.

[18] Grizzle, J. E., Starmer, C. F. and Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics* **25,** 489-504.

[19] Hall, D. B. (2001). On the application of extended quasi-likelihood to the clustered data case. *Canadian Journal of Statistics* **29,** 77-97.

[20] Hall, D. B. and Severini, T. A. (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association* **93,** 1365-1375.

[21] Hall, P. (1987). Edgeworth expansion for Student's t statistic under minimal moment conditions. *Annals of Probability* **15,** 920-931.

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

[23] Hardin, J. W. and Hilbe, J. M. (2002). *Generalized Estimating Equations.* Chapman and Hall / CRC.

[24] He, X. and Hu, F. (2002). Markov Chain Marginal Bootstrap. *Journal of the American Statistical Association* **97,** 783-795.

[25] Henderson, R. and Shimakura, S. (2003). A serially correlated gamma frailty model for longitudinal count data. *Biometrika* **90**, 355-366.

[26] Hu, F. (2001). Efficiency and Robustness of a Resampling M-Estimator in the Linear Model. *Journal of Multivariate Analysis* **78,** 252-271.

[27] Hu, F. and Kalbfleisch, J. D. (2000). The estimating function bootstrap (with discussions). *Canadian Journal of Statistics* **28,** 449-499.

[28] Hu, F. and Zidek, J. V. (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika* **82,** 263-275.

[29] Jiang, W. and Kalbfleisch J. D. (2004). Resampling methods for estimating functions with U-statistic structure. *Working Paper Series. The University of Michigan Department of Biostatistics.* Working Paper 33. URL: http://www.bepress.com/umichbiostat/paper33.

[30] Jiang, J. and Zhang, W. (2001). Robust estimating in generalised linear mixed models. *Biometrika* **88,** 753-765.

[31] Jin, Z., Ying, Z. and Wei, L. J. (2001). A resampling method by perturbing the minimand. *Biometrika* **88,** 381-390.

[32] Joanes, D. N. and Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician* **47,** 183-189.

[33] Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* **96,** 1387-1396.

[34] Koch, G. G., Landis, J. R., Freeman, J. L, Freeman, Jr., D. H. and Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics* **33,** 133-158.

[35] Kurtz, A. K. (1948). A research test of Rorschach test. *Personnel Psychology* **1,** 41-53.

[36] Lahiri., S. N. (2003). *Resampling Methods for Dependent Data*. Springer.

[37] Lele, S. (1991a). Jackknifing linear estimating equations: asymptotic theory and applications in stochastic processes. *Journal of the Royal Statistical Society, Series B* **53,** 253-267.

[38] Lele, S. (1991b). Resampling using estimating equations. *Estimating Functions*, Godambe V.P. (Ed.). Oxford University Press, pp. 295-304.

[39] Li, B. (1998). An optimal estimating equation based on the first three cumulants. *Biometrika* **85,** 103-114.

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

[41] Liang, K. - Y., Zeger, S. L. and Qaqish, B. (1992). Multivariate regression analyses for categorical data (with discussion). *Journal of the Royal Statistical Society, Series B* **54,** 3-40.

[42] Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1990a). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika* **78,** 153-160.

[43] Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1990b). Using the Jackknife to Estimate the Variance of Regression Estimators from Repeated Measures Studies. *Communications in Statistics, Theory and Methods* **19,** 821-845.

[44] Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Annals of Statistics* **16,** 1696-1708.

[45] Liu, R. Y. and Singh, K. (1992). Efficiency and Robustness in Resampling. *Annals of Statistics* **20,** 370-384.

[46] Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics* **57,** 126-134.

[47] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Chapman and Hall, London.

[48] Moulton, L. H. and Zeger, S. L. (1989). Analyzing repeated measures on generalized linear models via the bootstrap. *Biometrics* **45,** 381-394.

[49] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* **135,** 370-384.

[50] Paik, M. C. (1988). Repeated measurement analysis for nonnormal data in small samples. *Communications in Statistics, Simulation and Computation* **17,** 1155-1171.

[51] Pan, W. (2001). On the robust variance estimator in generalised estimating equations. *Biometrika* **88,** 901-906.

[52] Pan, W., Louis, T. A. and Connett, J. E. (2000). A note on marginal linear regression with correlated response data. *The American Statistician* **54,** 191-195.

[53] Park, T., Davis, C. S. and Li, N. (1998). Alternative GEE estimation procedures for discrete longitudinal data. *Computational Statistics and Data Analysis* **28,** 243-256.

[54] Parzen, M.I., Wei, L.J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81,** 341-350.

[55] Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulation and Computation* **23,** 939-951.

[56] Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44,** 1033-1048.

[57] Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47,** 825-839.

[58] Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B* **11,** 18-84.

[59] Schildscrout, J. S. and Heagerty, P, J. (2005). Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostatistics* **6,** 633-652.

[60] Sherman, M. and le Cessie, S. (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. *Communications in Statistics, Simulation and Computation* **26,** 901-925.

[61] Tian, L., Liu, J., Zhao, Y. and Wei, L. J. (2004). Statistical inference based on non-smooth estimating functions. *Biometrika***91,** 943-954.

[62] Timm, N. H. (1980). Multivariate analysis of variance of repeated measurements. In Krishnaiah, P. R. and Sen, P. K. editors, *Handbook of Statistics, Volume 1: Analysis of Variance,* pages 41-87. Elsevier Science Publishers, Amsterdam.

[63] Tukey, J. W. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics* **29,** 614.

[64] Wang, Y. - G. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika* **90,** 29-41.

[65] Wang, Y. - G. and Carey, V. (2004). Unbiased estimating equations from working correlation models for irregularly timed repeated measures. *Journal of the American Statistical Association* **99,** 845-853.

[66] Wang, Y. - G. and Li, Y. (2005). Smooth bootstrap methods for analysis of longitudinal data. submitted.

[67] Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician* **39,** 95-101.

[68] Wedderburn, R. W. M. (1974). Quasilikelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61,** 439-447.

[69] Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics* **14,** 1261-95.

[70] Yan, J. and Fine, J. (2004). Estimating equations for association structures. *Statistics in Medicine* **23,** 859-874.

[71] Zeger, S. L. and Liang, K. - Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine* **11,** 1825-1839.

[72] Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77,** 642-648.

[73] Ziegler, A., Kastner, C., Brunner, D. and Blettner, M. (2000). Familial association of lipid profile: a generalized estimating equations approach. *Statistics in Medicine* **19,** 3345-3357.