Statistical Inference in the Presence of Missing Data

by

Malgorzata Winiszewska, B.Sc., M.Sc.

A thesis submitted to

the Faculty of Graduate Studies and Research

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

at

School of Mathematics and Statistics

Ottawa-Carleton Institute for Mathematics and Statistics

Carleton University

Ottawa, Ontario

August, 2014

Abstract

In this thesis, we study statistical inference in the presence of missing data. In Chapters 2-4, we obtain asymptotically valid imputed estimators for the population mean, distribution function and correlation coefficient, and propose adjustments to Shao and Sitter (1996) bootstrap confidence intervals under imputation for missing data. We show that the adjusted bootstrap estimators should be used with bootstrap data obtained by imitating the process of imputing the original data set.

In Chapter 5, we establish a goodness-of-fit test that can be applied to the case of longitudinal data with missing at random (MAR) observations, by combining the concepts of weighted generalized estimating equations (Robins et al., 1995) and score test statistic for goodness-of-fit (Hosmer and Lemeshow, 1980; Horton et al., 1999). We show that the proposed goodness-of-fit method that incorporates the missingness process should be used when dealing with intermittent missingness.

In Chapter 6, we study a conditional model for a mixture of correlated, discrete and continuous, outcomes and apply the likelihood method to MAR data. We conduct a simulation study to compare the performance of estimators resulting from the joint model with estimators based on separate models for binary and continuous outcomes. We show that when all data are observed, adopting the mixed model does not lead to notable improvements; on the contrary, under a scenario with binary MAR data, the joint model performs significantly better.

Acknowledgements

I wish to express my appreciation for my supervisors Dr. Jon Rao and Dr. Sanjoy Sinha. It has been a true privilege to work with them and a great honour to have the opportunity to learn from their vast knowledge and experience. Thank you both for your patience, guidance and support.

Professor Rao has influenced me not only as a student, but as a professional and a person. I have been continuously inspired by his passion for research, unparalleled expertise, commitment to values, and most of all his sincere dedication to teaching. Thank you for being my mentor.

I would like to thank Professor Sinha for his guidance, hands-on programming advice and his distinctive kindness throughout the years in the program.

I would like to thank Professor Yongsong Qin and Professor Kalyan Das for their constructive suggestions on Chapters 2-4 and Chapter 5 respectively.

I would like to thank the members of my thesis committee, Professors David Haziza, Mayer Alvo, Roland Thomas and Patrick Farrell, for their time and constructive feedback.

To all my family and friends: thank you for your encouragement. Particularly, I wish to thank my husband for years of understanding and unwavering succour, my parents, for their immeasurable support, and my sister, for always being there for me. This work could not have been completed without you.

And last, I want to thank my 3-year-old son Luca, for brightening every day of my life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Missing Data

Missing observations are commonly encountered in real data including sample surveys, epidemiological studies or clinical trials, and can be due to non-response, missing measurements or study withdrawals. A common approach is to ignore the missing observations; this however, can cause bias or lead to inefficient estimators.

Suppose that $\{(Y_i, \delta_i); i = 1, ..., n\}$ are independent identically distributed (i.i.d.) samples of incomplete data generated from a random vector (Y, δ), where Y is the variable of interest and the response indicator $\delta_i = 1$ when Y_i is observed and $\delta_i = 0$ if Y_i is missing. Let $Y = (Y_{obs}, Y_{miss})$ where $Y_{obs} = \{Y_i | \delta_i = 1, i = 1, .., n\}$ denotes the observed values, and $Y_{miss} = \{Y_i | \delta_i = 0, i = 1, .., n\}$ denotes the missing values. The framework for missing data, introduced by Rubin (1976), consists of three missing data mechanisms that describe the relationships between measured variables and the probability of missing data: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). In particular,

- Data are MCAR when missingness does not depend on the values of the data, missing or observed. That is, $P(\delta_i = 1|Y_i) = P(\delta_i = 1)$.
- Data are MAR when missingness depends only on observed, known responses and possibly on auxiliary information X, but is unrelated to the the sets of unobserved responses. That is $P(\delta_i|Y, X) = P(\delta_i|Y_{obs}, X)$.
- Data are MNAR when missingness depends on unobserved, unknown data. That is, probability of response depends on the variable of interest Y.

MCAR is a strong, and often unrealistic, assumption, but it can be relaxed in practice by considering imputation classes as shown in Chapter 3. The less restrictive MAR mechanism is more plausible. Both MAR, and MCAR as a special case of MAR, are often referred to as ignorable response mechanisms; this means that $P(\delta_i|Y, X)$ can be ignored and valid likelihood analysis can be obtained if the data model is correctly specified (Fitzmaurice et al., 2008). When the response mechanism is nonignorable, the data are MNAR; in this case, a model for missingness must be considered in the analysis to prevent bias.

Common statistical methods that address missingness include imputation, weighting approaches and the likelihood. Imputation techniques are sample-based and replace missing values by, one or more, plausible values. We consider imputation in Chapters 2-4. In the weighting techniques, missing data are handled indirectly through simultaneous models for the data and for the missingness process. We adapt the weighted generalized estimating equations (WGEE) approach (Robins et al., 1995) to longitudinal MAR data in Chapter 5. Finally, the likelihood approach uses all of the available data to obtain consistent and asymptotically efficient estimators (Alison, 2012). We apply the likelihood method to MAR data in Chapter 6.

1.2 Longitudinal Data

In longitudinal studies, subjects are followed over time and data consist of repeated measures over series of time points for all subjects. Examples of longitudinal studies include clinical trials, biomedical research, observational and experimental studies. Analysis of longitudinal data require methods that can properly account for the intra-subject correlation of response measurements. Diggle et al. (2002) provide a comprehensive overview of various models and methods for the analysis of longitudinal data.

Missing data are very common in longitudinal studies. In addition to Rubin's nomenclature, we distinguish between two patterns for missingness in longitudinal data: monotone and intermittent. Monotone missingness, or so-called dropout, means that if an observation is missing, then all subsequent observations are also missing for a given individual; on the contrary, intermittent pattern means that either missing, or observed, response may be present at any time for a given individual in the data file.

1.3 Empirical Likelihood Confidence Intervals Under Full Response

Likelihood-based methods have been shown to generate efficient estimators and short confidence intervals under various settings. The methods are flexible and apply to most models and to different types of data. They can be used in various data settings, including incomplete or sampled with a bias data, and can incorporate auxiliary information in the form of constraints on the domain of the likelihood function. The empirical likelihood (EL) method, proposed by Owen (1988), is a nonparametric counterpart to parametric likelihood-based tests and confidence regions. The EL methods have been proposed for many parameters of interest such as population mean, distribution function and estimating equations (Qin and Lawless, 1994; Owen, 2001). Owen (1988) proved that, in the context of independent identically distributed random variables, the empirical likelihood ratio statistic has an asymptotic chi-square distribution and can be used to form confidence intervals. The EL confidence intervals are range preserving and transformation invariant, their shape is determined by the data and they do not require evaluation of standard errors of estimators and provide well balanced tail error rates. Historically, Hartley and Rao (1968, 1969) originated the concept of the empirical likelihood method in survey sampling constructing so-called scale-load estimators for the population mean under simple random sampling and unequal probability sampling with replacement.

We study bootstrap EL confidence intervals in the presence of missing data in Chapters 2-4. We now outline the concept of the empirical likelihood with fully observed data, following the theory presented in Owen (2001) and Qin and Lawless (1994).

Suppose we observe independent and identically distributed (i.i.d.) data $y_1, ..., y_n$ where each y_i, $i = 1, ..., n$ is distributed according to an unknown distribution $F(\cdot) \subset \mathcal{F}$. Let p_i be a probability mass assigned to sample point y_i, that is $p_i = F(y_i) - F(y_i^-)$ and $p_i > 0$, $\sum_{i=1}^n p_i = 1$, $i = 1, ..., n$. The empirical likelihood function is defined by

$$L(F) = \prod_{i=1}^n dF(y_i) = \prod_{i=1}^n p_i,$$

with F being unspecified and therefore, treated nonparametrically. The empirical likelihood $L(F)$ is maximized by the empirical distribution function

$$F_n(y) = n^{-1} \sum_{i=1}^n I\left(Y_i \leq y\right).$$

The empirical likelihood ratio is

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^n np_i.$$

The profile likelihood ratio function is defined by $\mathcal{R}(\theta) = \sup(R(F)|\theta(F), F \in \mathcal{F}\}$ and the corresponding empirical likelihood confidence regions are of the form $\{\theta|\mathcal{R}(\theta) \geq r_0\}$.

For example, for the mean μ of F,

$$\mathcal{R}(\mu) = \max\left\{\prod_{i=1}^n np_i \,\middle|\, \sum_{i=1}^n p_i y_i = \mu, \sum_{i=1}^n p_i = 1, p_i \geq 0\right\}, \tag{1.1}$$

and

$$\{\mu|\mathcal{R}(\mu) \geq r_0\} = \left\{ \sum_{i=1}^{n} p_i Y_i = \mu \,\middle|\, \prod_{i=1}^{n} np_i \geq r_0, \sum_{i=1}^{n} p_i = 1, p_i \geq 0 \right\}. \tag{1.2}$$

We define the empirical likelihood ratio for μ as

$$l(\mu) = -2\log(\mathcal{R}(\mu)).$$

Using Lagrange multiplier method, it can be shown that $p_i = n^{-1}\left\{1 + \lambda'(y_i - \mu)\right\}^{-1}$ where λ is the solution to

$$\sum_{i=1}^{n} \frac{(y_i - \mu)}{1 + \lambda'(y_i - \mu)} = 0. \tag{1.3}$$

It follows that

$$l(\mu) = 2\sum_{i=1}^{n} \log\left\{1 + \lambda'(y_i - \mu)\right\}, \tag{1.4}$$

and (1.4) converges in distribution to χ_1^2 as $n \to \infty$.

Owen's univariate empirical likelihood theorem (Owen, 2001) is stated below.

Theorem 1.1 *Let $Y_1, ..., Y_n$ be independent random variables with common distribution F. Let $\mu = E(y_i)$ and assume that $0 < Var(Y_i) < \infty$. Then as $n \to \infty$,*

$$l(\mu) \xrightarrow{d} \chi_1^2. \tag{1.5}$$

Based on this result, an approximate α-level EL confidence interval on μ may be obtained as $\{\mu|l(\mu) \leq c_\alpha\}$, where c_α is such that $P\left(\chi_1^2(\alpha) \leq c_\alpha\right) = \alpha$.

Finding a solution to (1.3) is a major computational problem, and can be done using the modified Newton-Raphson algorithm proposed by Wu (2005) and based on the procedure presented in Chen et al. (2002).

1.4 Empirical Likelihood and Estimating Equations

Estimating equations serve as a means for expressing the association among parameters, their corresponding statistics, and nuisance parameters. In order to specify the correlation coefficient ρ, Owen (2001) considered a parameter vector $\theta = (E(y), E(z), \sigma_y, \sigma_z, \rho)$ and formed five estimating equations:

$$0 = E(y - E(y)),$$

$$0 = E(z - E(z)),$$

$$0 = E((y - E(y))^2 - \sigma_y^2),$$

$$0 = E((z - E(z))^2 - \sigma_z^2),$$

$$0 = E((y - E(y))(z - E(z)) - \rho\sigma_y\sigma_z).$$

Since we are interested in making inference about ρ, we treat the other four parameters in θ as nuisance parameters. Qin and Lawless (1994) formed empirical likelihood ratio test statistic for obtaining confidence limits for fully observed i.i.d. data through linking estimating equations and empirical likelihood. They defined the empirical log likelihood ratio in terms of r functionally independent unbiased estimating functions $g_j(x, \theta)$, such that $E(g_j(x, \theta)) = 0$, $j = 1, ..., r$, as follows:

$$l_E(\theta) = \sum \log(1 + t'(\theta)g(x, \theta)), \tag{1.6}$$

with $g(x,\theta) = (g_1(x,\theta),...,g_r(x,\theta))'$. In particular, they considered the case with the number of estimating equations r greater than the number of parameters p, and proposed the following corollary.

Corollary 1.2 (Qin and Lawless, 1994) *Let $\theta' = (\theta_1, \theta_2)'$, where θ_1 is a $q \times 1$ vector and θ_2 is $(p-q) \times 1$. For $H_0 : \theta_1 = \theta_1^0$, the profile likelihood ratio test statistic is*

$$W_E = 2l_E(\theta_1^0, \tilde{\theta}_2^0) - 2l_E(\tilde{\theta}_1, \tilde{\theta}_2), \tag{1.7}$$

where l_E is defined in (1.6) $\tilde{\theta}$ minimizes $l_E(\theta)$ with respect to θ and $\tilde{\theta}_2^0$ minimizes $l_E(\theta_1^0, \theta_2)$ with respect to θ_2. Under H_0, $W_E \to \chi_q^2$ as $n \to \infty$.

1.5 Imputation

Complete case analysis that discards units with incomplete information may lead to biased survey estimates unless the data are MCAR. In addition, the resulting estimators could have larger variance compared to the variance of estimators under a full response scenario.

Imputation is a common method to compensate for unit nonresponse in sample surveys. Imputation techniques replace missing responses by one or more plausible values with the main objective to reduce the bias in survey estimates that could be due to simply ignoring units with item nonresponse.

A variety of imputation methods have been developed and can be divided into two classes: deterministic or random, respectively depending on whether the imputed data is fixed, given the sample, or random (Kalton and Kasprzyk, 1986). The mean imputation, under which the sample mean of respondents is used to replace all missing data, is a popular deterministic imputation method that produces a fixed imputed value given the sample. In random hot-deck imputation, a simple random sample with replacement is selected from the set of respondents and the associated values are used as donors for non-respondents. Random imputation methods may result in different imputed values, given the sample, if the process is repeated.

There are advantages and disadvantages to both classes of methods. For example, deterministic methods do not preserve the distribution of the imputed variables, while random methods do. Chen, et al. (2000) indicated that deterministic imputation leads to an inconsistent estimator of the distribution function, unlike random imputation that yields a consistent estimator. On the other hand, random imputation induces imputation variance due to random selection of imputed values, which is not the case for deterministic imputation.

1.5.1 Multiple Imputation

We also distinguish between single and multiple imputation methods. Single imputation uses a single imputed value to fill in the missing item and results in one complete data file. Multiple imputation on the other hand, refers to the procedure of replacing each missing value by $M \geq 2$ imputed values and results in M complete data sets. The analyses are carried on each imputation set and combined together (Rubin, 1978).

1.5.2 Fractional Imputation

Fractional imputation method, proposed by Kalton and Kish (1984), is an alternative to multiple imputation. Under this technique, $J \geq 1$ imputed values are randomly selected for each missing observation and a weight equal to a fraction J^{-1} of the original survey weight of each donor is assigned to each imputed value. It should be noted that J is a fixed number and does not depend on the sample size. The disadvantage of this method is that all $J \geq 1$ imputed values have to be stored in the data file for each missing observation, therefore in practice, J is usually small.

As a compromise to deterministic and random imputation methods, fractional imputation was designed to reduce the imputation variance and yet preserve the distribution as in hot-deck imputation. It can be shown that as J increases, the imputation variance decreases and that the method leads to consistent imputed estimators of the mean as well as the distribution function.

1.5.3 Imputation Classes

In practice, observations are often divided into homogenous groups, called imputation classes, such that the missing values can be imputed independently, using separate imputation procedures, within each class. The sample is divided into classes according to auxiliary variables that are associated with the variable to be imputed (Brick and Kalton, 1996). The concept of forming imputation classes is related to stratification in survey sampling (however, their goals are different) and stratification techniques can be used to form imputation classes. Haziza and Beaumont (2007) compare different methods for constructing imputation classes.

1.6 Bootstrap Confidence Intervals for Imputed Data

Bootstrap is a useful, computer intensive, method that can be used to estimate sampling distributions of estimators. The concept was first introduced by Efron (1979) for i.i.d. samples and extended by Rao and Wu (1988) to complex sampling designs. Shao and Tu (1996) provide an overview of the bootstrap theory and applications in the i.i.d. case. Shao and Sitter (1996) showed that under imputation for missing data, the usual bootstrap method leads to invalid results. That is, if the imputed values were treated as if they were the true observations, and proportion of missing data was considerable, the variance of the imputed estimator would be underestimated as any inflation in variance caused by imputation and missing data would be ignored. Instead, Shao and Sitter (1996) proposed that the bootstrap data set should be imputed in the same way as the original data set to generate asymptotically valid variance and estimators of population parameters. In this thesis, we apply the Shao and Sitter procedure under different imputation methods in Chapters 2-4.

1.7 Outline of the Thesis

The thesis is organized as follows.

In Chapter 2, we establish the asymptotic normality of the imputed estimators of the mean and the distribution function under fractional imputation. We then construct adjusted bootstrap percentile (BP) confidence intervals based on the bootstrap data obtained by imitating the process of imputing the original data set in bootstrap resampling. We establish the limiting distributions of the empirical likelihood ratio statistics and study bootstrap calibrated EL confidence intervals.

In Chapter 3, we extend the theory introduced in Chapter 2 to imputation classes. We establish the asymptotic normality of fractionally imputed estimators of the mean and the distribution function with imputation classes, and construct asymptotically valid bootstrap percentile and empirical likelihood confidence intervals.

In Chapter 4, we construct confidence intervals on the correlation coefficient under joint regression imputation. We investigate asymptotic properties of the estimators and construct bootstrap percentile and empirical-likelihood confidence intervals on the correlation coefficient after applying joint regression imputation to the data.

In Chapter 5, we review Liang and Zeger's (1986) concept of generalized estimating equations and outline the weighted generalized estimating equations (WGEE) method to analyse MAR data (Robins et al., 1995). We propose a goodness-of-fit test that can be applied to the case of longitudinal data with MAR observations and draw a comparison between the proposed goodness-of-fit method, which incorporates estimation of the missingness model parameters, and the ordinary method that ignores the missingness process.

In Chapter 6, we study properties of the conditional mixed, discrete and continuous, outcomes model and apply the likelihood method to MAR data. Specifically, we compare the performance of estimation based on a joint model for the mixed outcomes with estimation based on modeling the binary and continuous outcomes separately when all data are observed, and under a scenario with binary data missing at random.

Simulation studies are conducted to assess performance of the proposed methods and results are presented at the end of each chapter.

Conclusions and suggestions for future research are discussed in Chapter 7.

Chapter 2

Confidence Intervals for Population Mean and Distribution Function Under Fractional Imputation

2.1 Introduction

Missing observations are commonly encountered in data from sample surveys due to nonresponse and imputation is used to compensate for nonresponse. Shao and Sitter (1996) proposed a bootstrap approach for handling imputed data by imputing the bootstrap samples in the same way as the original data set. In this chapter, we construct bootstrap percentile and bootstrap empirical likelihood confidence intervals on the mean $\mu = E(Y)$ and the distribution function $\theta := F(y) = P(Y \leq y)$, $y \in R$, and propose an adjustment to the Shao and Sitter's (1996) bootstrap confidence intervals under fractional imputation (Kim and Fuller, 2004). Qin et al. (2008) obtained asymptotically correct, normal approximaton (NA) and empirical likelihood (EL), confidence intervals for marginal parameters under mean, random hot-deck and adjusted random hot-deck imputation methods. In this chapter, we consider fractional imputation, with $J \geq 1$ imputed values, and form bootstrap confidence intervals.

The chapter is organised as follows. In Section 2.2, we establish the asymptotic normality of the imputed estimators of the mean $\mu = E(Y)$ and the distribution function $\theta = F(y)$ under fractional imputation. We then construct adjusted bootstrap percentile (BP) confidence intervals based on the bootstrap data obtained by imitating the process of imputing the original data set in bootstrap resampling. In Section 2.3, we establish limiting distributions of the empirical likelihood (EL) ratio statistics and study bootstrap calibrated EL confidence intervals. We conduct a simulation study on performance of the proposed bootstrap intervals and present the results in Section 2.4. Additional theorems and proofs are shown in the appendix (Section 2.6).

2.1.1 Framework

In this chapter, we focus on inference about the mean $\mu = E(Y)$, and show corresponding results for the distribution function $\theta = F(y)$, for given y, in the presence of missing values. Particularly, we consider the case of independent identically distributed (i.i.d.) samples of incomplete data $\{(Y_i, \delta_i); i = 1, 2, ..., n\}$ generated from random vector (Y, δ), where $\delta_i = 0$ if Y_i is missing and $\delta_i = 1$ otherwise. We assume no parametric structure on the distribution of Y except that $0 < var(Y) = \sigma^2 < \infty$. Further, we assume that Y is missing completely at random (MCAR) with

$$P(\delta_i = 1|Y_i) = P(\delta_i = 1), \tag{2.1}$$

and denote the probability of response by p, that is

$$p = P(\delta_i = 1),\ 0 < p \leq 1. \tag{2.2}$$

Note that assumption (2.1) is relaxed in Chapter 3.

2.1.2 Fractional Imputation

Imputation is the process of determining and assigning replacement values for missing data. Both random and deterministic imputation methods have their advantages and disadvantages. For example, random imputation results in consistent estimators of the mean and distribution functions of Y (Chen et al., 2000); however, it induces imputation variance due to random selection of imputed values. Imputation variance can be a significant component to total variance especially when non-response is high. Deterministic imputation, on the other hand, eliminates imputation variance but the distribution of item values is not preserved and so it leads to an inconsistent imputed estimator of the distribution function.

In this chapter, we use fractional imputation to deal with missing data. The fractional imputation method, proposed by Kalton and Kish (1984) and studied further by Kim and Fuller (2004), is an alternative to multiple imputation. It replaces each missing value with $J \geq 1$ randomly selected imputed values and assigns a fraction J^{-1} to each imputed value. It can be shown that as J increases, the imputation variance decreases and that the method leads to consistent imputed estimators of the mean $\mu = E(Y)$ as well as the distribution function $F(y)$. The disadvantage of this method is that all $J \geq 1$ imputed values have to be stored in the data file for each missing Y_i. Also confidence intervals require identification flags on the imputed values present in data file, which in practice may be difficult to obtain due to confidentiality reasons (Qin et al., 2008).

Let $r = \sum_{i=1}^{n} \delta_i$ be the number of respondents and $m = n - r$ represent the number of missing units. Denote the sets of respondents as s_r, and let s_m be the non-respondents, in the sample s ($s = s_r \cup s_m$). Under fractional imputation, for each missing Y_i, $i \in s_m$, we generate J imputed values

$$Y_{ij} = \bar{Y}_r + \epsilon^*_{ij}, \;\; j = 1, ..., J, \tag{2.3}$$

where

$$\bar{Y}_r = \frac{1}{r} \sum_{i \in s_r} Y_i \tag{2.4}$$

is the mean of respondents, and $\left\{\epsilon^*_{ij}, i = 1, ..., n, \;\; j = 1, ..., J\right\}$ are drawn by simple random sampling with replacement from the donor residuals $\left\{\hat{\epsilon}_i = Y_i - \bar{Y}_r, i \in s_r\right\}$. After fractional imputation, the imputed data file consists of $\{(\widetilde{Y}_i, \delta_i); i = 1, 2, ..., n\}$, where $\widetilde{Y}_i = Y_i$ if $\delta_i = 1$ or $\widetilde{Y}_i = (Y_{i1} ... Y_{iJ})$ if $\delta_i = 0$ and the fraction J^{-1} is attached to each imputed value. Random imputation is a special case of fractional imputation with $J = 1$.

2.2 Normal Approximation

The fractionally-imputed estimators of the mean μ and the distribution function θ are respectively given by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i Y_i + (1-\delta_i)\frac{1}{J}\sum_{j=1}^{J} Y_{ij}\right\}, \tag{2.5}$$

and

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i I(Y_i \leq y) + (1-\delta_i)\frac{1}{J}\sum_{j=1}^{J} I(Y_{ij} \leq y)\right\}. \tag{2.6}$$

Let E^* denote the expectation with respect to randomness in the imputation procedure. Note that since $E^*[Y_{ij}] = r^{-1}\sum_{i\in s_r} Y_i = r^{-1}\sum_{i=1}^{n}\delta_i Y_i = \bar{Y}_r$, we have

$$\begin{aligned} E^*[\hat{\mu}] &= \frac{1}{n}\sum_{i=1}^{n}\delta_i Y_i + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{1}{J}\sum_{j=1}^{J}\bar{Y}_r \\ &= \bar{Y}_r\frac{r}{n} + \bar{Y}_r\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i) = \bar{Y}_r\left(\frac{r}{n} + \frac{n-r}{n}\right) = \bar{Y}_r. \end{aligned}$$

Similarly for the distribution function, $E^*[I(Y_{ij} \leq y)] = r^{-1}\sum_{i=1}^{n}\delta_i I(Y_i \leq y) = \bar{\theta}_r$, so that

$$E^*[\hat{\theta}] = \frac{1}{n}\sum_{i=1}^{n}\delta_i I(Y_i \leq y) + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{1}{J}\sum_{j=1}^{J}\bar{\theta}_r = \bar{\theta}_r, \tag{2.7}$$

where

$$\bar{\theta}_r = r^{-1}\sum_{i\in s_r} I(Y_i \leq y). \tag{2.8}$$

2.2.1 Ordinary Confidence Intervals

Theorem 2.1 states the results on asymptotic normality of the fractionally-imputed estimators $\hat{\mu}$ and $\hat{\theta}$.

Theorem 2.1 *Assume that* $0 < p = P(\delta_i = 1) \leq 1$, $0 < \sigma^2 = Var(Y_i) < \infty$ *and that there exists an* $\alpha_0 > 0$ *such that* $E|Y_i|^{2+\alpha_0} < \infty$. *Then, as* $n \to \infty$,

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N\left(0, \sigma_\mu^2\right), \tag{2.9}$$

and

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, \sigma_d^2\right), \tag{2.10}$$

where $\sigma_\mu^2 = \left(p^{-1} + J^{-1}(1-p)\right)\sigma^2$ *and* $\sigma_d^2 = \left(p^{-1} + J^{-1}(1-p)\right)\theta\left(1-\theta\right).$

Using this result, the corresponding ordinary normal approximation confidence intervals for μ and θ are respectively given by

$$\mu \in \left(\hat{\mu} - z_{\alpha/2}\hat{\sigma}_\mu/\sqrt{n},\ \hat{\mu} + z_{\alpha/2}\hat{\sigma}_\mu/\sqrt{n}\right), \tag{2.11}$$

and

$$\theta \in \left(\hat{\theta} - z_{\alpha/2}\hat{\sigma}_d/\sqrt{n},\ \hat{\theta} + z_{\alpha/2}\hat{\sigma}_d/\sqrt{n}\right), \tag{2.12}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile from standard normal distribution and

$$\hat{\sigma}_\mu^2 = \left(\hat{p}^{-1} + J^{-1}(1-\hat{p})\right)(r-1)^{-1}\sum_{i \in s_r}(Y_i - \bar{Y}_r)^2, \tag{2.13}$$

$$\hat{\sigma}_d^2 = \left(\hat{p}^{-1} + J^{-1}(1-\hat{p})\right)(r-1)^{-1}\sum_{i\in s_r}\left(I(Y_i \le y) - \bar{\theta}_r\right)^2, \tag{2.14}$$

with $\hat{p} = r/n$, $\bar{Y}_r$ and $\bar{\theta}_r$ defined respectively by (2.4) and (2.8).

Note that $\hat{\theta}$ is a proportion and $s^2 = (r-1)^{-1}\sum_{i\in s_r}\left(I(Y_i \le y) - \bar{\theta}_r\right)^2$ is a consistent estimator of $Var(I(Y_i \le y)) = \theta\{1-\theta\}$. It can also be shown that $s^2 = r(r-1)^{-1}\left(\bar{\theta}_r - \bar{\theta}_r^2\right) = r(r-1)^{-1}\left\{\left(\frac{n}{r}\hat{\theta} - \left(\frac{n}{r}\hat{\theta}\right)^2\right) - \left(\frac{m}{r}\bar{\theta}_m - \left(\frac{m}{r}\bar{\theta}_m\right)^2\right)\right\}$, where $\bar{\theta}_m = m^{-1}\sum_{i\in s_m}(1-\delta_i)J^{-1}\sum_{j=1}^{J} I(Y_{ij} \le y)$.

2.2.2 Bootstrap Confidence Intervals

We employ the method proposed by Shao and Sitter (1996) to approximate the asymptotic distributions of $\sqrt{n}(\hat{\mu} - \mu)$ and $\sqrt{n}(\hat{\theta} - \theta)$ under fractional imputation. The steps of the procedure are as follows:

1. Set $b = 1$.

2. Draw a simple random sample $D^* = \{(Y_{b,i}, \delta_{b,i}), i = 1, ..., n\}$ with replacement from the imputed data set $D = \{(\tilde{Y}_i, \delta_i), i = 1, ..., n\}$.

3. When $\delta_{b,i} = 0$, apply the same imputation procedure that was used on the original data set using the subsample of bootstrap respondents as donors. That is, under fractional imputation, we generate $J \ge 1$ imputed values $Y_{b,ij} = \bar{Y}_{b,r} + \epsilon^*_{b,ij}$, where $\left\{\epsilon^*_{b,ij},\ j = 1, ..., J\right\}$ are drawn by simple random sampling with replacement from the donor residuals $\left\{\hat{\epsilon}_{b,l} = Y_{b,l} - \bar{Y}_{b,r},\ l \in s_{b,r}\right\}$ where $\bar{Y}_{b,r} = \sum_{i=1}^{n}\delta_{b,i}Y_{b,i}/\sum_{i=1}^{n}\delta_{b,i}$ and $s_{b,r} = \{i : \delta_{b,i} = 1\}$.

4. Compute the imputed bootstrap estimators of μ and θ from the fractionally-

imputed bootstrap data:

$$\hat{\mu}_b = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{b,i}Y_{b,i} + (1-\delta_{b,i})\frac{1}{J}\sum_{j=1}^{J}Y_{b,ij}\right\},$$

and

$$\hat{\theta}_b = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{b,i}I(Y_{b,i}\le y) + (1-\delta_{b,i})\frac{1}{J}\sum_{j=1}^{J}I(Y_{b,ij}\le y)\right\}.$$

5. Repeat steps 2-4, for $b = 2, ..., B$, where B is a large number.

The usual bootstrap analogues of $\hat{\mu}-\mu$ and $\hat{\theta}-\theta$ are respectively given by $\hat{\mu}_b-\hat{\mu}$ and $\hat{\theta}_b-\hat{\theta}$. Theorem 2.2 states that, under fractional imputation, the distributions of $\sqrt{n}(\hat{\mu}-\mu)$ and $\sqrt{n}(\hat{\theta}-\theta)$ can be respectively approximated by the modified bootstrap versions $\sqrt{n}(\hat{\mu}_b-\overline{Y}_r)$ and $\sqrt{n}(\hat{\theta}_b-\bar{\theta}_r)$.

Theorem 2.2 *Suppose that the conditions in Theorem 2.1 are satisfied, then as* $n\to\infty$,

$$\sup_{x\in R}\left|P_b\left\{\sqrt{n}(\hat{\mu}_b-\overline{Y}_r)\le x\right\} - P\left\{\sqrt{n}(\hat{\mu}-\mu)\le x\right\}\right| \xrightarrow{P} 0, \tag{2.15}$$

and

$$\sup_{x\in R}\left|P_b\left\{\sqrt{n}(\hat{\theta}_b-\bar{\theta}_r)\le x\right\} - P\left\{\sqrt{n}(\hat{\theta}-\theta)\le x\right\}\right| \xrightarrow{P} 0. \tag{2.16}$$

where P_b *denotes the conditional probability given* D.

That is, the proposed adjustments to Shao and Sitters's (1996) statistics are given by

$$\mu_{nb} = \overline{Y}_r - \hat{\mu} \text{ and } F_{nb} = \bar{\theta}_r - \hat{\theta}. \tag{2.17}$$

Note that μ_{nb} and F_{nb} are asymptotically normal, that is

$$\sqrt{n}\mu_{nb} \xrightarrow{d} N\left(0, J^{-1}(1-p)\sigma^2\right) \text{ and } \sqrt{n}F_{nb} \xrightarrow{d} N\left(0, J^{-1}(1-p)\theta\left(1-\theta\right)\right). \quad (2.18)$$

Theorem 2.2 shows that, in the presence of missing data, the usual bootstrap statistic proposed by Shao and Sitter (1996) does not approximate the original pivotal. However, when J is large enough, the adjustment factor μ_{nb} becomes negligible, that is $\sqrt{n}\mu_{nb} = o_p(1)$ as $J \to \infty$. We also note that $\mu_{nb} = 0$ under deterministic imputation when $J = 1$ and $Y_{ij} = \overline{Y}_r$ for all $i \in s_m$. This means that no adjustment is needed for the deterministic imputation. However, as mentioned before, deterministic imputation leads to an inconsistent estimator of the distribution function of Y.

We construct the adjusted bootstrap percentile confidence intervals for μ as follows. We repeat the bootstrap process independently B times to obtain $\hat{\mu}_1, ..., \hat{\mu}_B$ and select the $100(1-\alpha/2)$ and $100(\alpha/2)$ sample quantiles of $\{\hat{\mu}_b, 1 \leq b \leq B\}$. The $(1-\alpha)$-level adjusted bootstrap percentile intervals on μ are given by $(\hat{\mu} - (\hat{\mu}_{b,1-\alpha/2} - \bar{Y}_r), \hat{\mu} - (\hat{\mu}_{b,\alpha/2} - \bar{Y}_r))$. The adjusted bootstrap percentile confidence intervals for θ can be formed in the similar way.

2.3 Empirical Likelihood

2.3.1 Ordinary Confidence Intervals

Empirical likelihood methods for constructing confidence regions under full response were explored by Owen (1988, 1990). It is known that empirical likelihood confidence regions respect the range of the parameter space, they are invariant under transformations, and their shapes (symmetry) are determined by the data. We form the empirical likelihood ratios for μ and θ, under fractional imputation, following the theory presented in Qin et al. (2008). Define

$$Z_{i,m}(\mu) = \delta_i Y_i + (1-\delta_i)J^{-1}\sum_{j=1}^{J} Y_{ij} - \mu, \tag{2.19}$$

and

$$Z_{i,d}(\theta) = \delta_i I(Y_i \leq y) + (1-\delta_i)J^{-1}\sum_{j=1}^{J} I(Y_{ij} \leq y) - \theta, \tag{2.20}$$

where $i = 1, ..., n$.

The respective empirical log-likelihood ratios are given by

$$l_{n,m}(\mu) = -2\max_{p_1 \ldots p_n}\left\{\sum_{i=1}^{n}\log(np_i)\middle|\ \sum_{i=1}^{n} p_i Z_{i,m}(\mu) = 0,\ \sum_{i=1}^{n} p_i = 1\right\}, \tag{2.21}$$

and

$$l_{n,d}(\theta) = -2\max_{p_1 \ldots p_n}\left\{\sum_{i=1}^{n}\log(np_i)\middle|\sum_{i=1}^{n}\ p_i Z_{id}(\theta) = 0,\ \sum_{i=1}^{n} p_i = 1\right\}. \tag{2.22}$$

It can be shown, using the Lagrange multiplier method, that

$$l_{n,m}(\mu) = 2\sum_{i=1}^{n} \log\left\{1 + \lambda_{n,m} Z_{i,m}(\mu)\right\},$$

and

$$l_{n,d}(\theta) = 2\sum_{i=1}^{n} \log\left\{1 + \lambda_{n,d} Z_{i,d}(\theta)\right\},$$

where $\lambda_{n,m}$ and $\lambda_{n,d}$ are respectively solutions to equations

$$\frac{1}{n}\sum_{i=1}^{n} \frac{Z_{i,m}(\mu)}{1 + \lambda_{n,m} Z_{i,m}(\mu)} = 0 \text{ and } \frac{1}{n}\sum_{i=1}^{n} \frac{Z_{i,d}(\theta)}{1 + \lambda_{n,d} Z_{i,d}(\theta)} = 0.$$

The asymptotic distributions of $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$ are established in Theorem 2.3.

Theorem 2.3 *Suppose that the conditions in Theorem 2.1 are satisfied. Then as* $n \to \infty$,

$$l_{n,m}(\mu) \xrightarrow{d} c_m \chi_1^2 \text{ and } l_{n,d}(\theta) \xrightarrow{d} c_d \chi_1^2,$$

where the scaling factors are given by $c_m = \sigma_\mu^2/\sigma_1^2$, $c_d = \sigma_d^2/\sigma_2^2$, *with* σ_μ^2 *and* σ_d^2 *defined in Theorem 2.1,* $\sigma_1^2 = \sigma^2 - (J-1)\,J^{-1}(1-p)\sigma^2$,*and* $\sigma_2^2 = \theta(1-\theta) - (J-1)\,J^{-1}(1-p)\theta(1-\theta)$.

Note that the empirical likelihood ratio under imputation is asymptotically distributed as a scaled chi-square variable, unlike the original result under full response (Owen, 2001), which is due to dependent data after imputation (Qin et al., 2008). To construct confidence intervals, consistent estimators of the scaling factors are required. Using Theorem 2.3, a (1-α)- level confidence interval on μ ,with asymptotically correct coverage probability, can be constructed as

$$\left\{\mu | \left(\hat{\sigma}_2^2/\hat{\sigma}_\mu^2\right) l_{n,m}(\mu) \leq \chi_\alpha^2(1)\right\}, \tag{2.23}$$

where $\chi_\alpha^2(1)$ is the upper α quantile of the χ^2 distribution with one degree of freedom, $\hat{\sigma}_\mu^2$ is given by (2.13) and

$$\hat{\sigma}_2^2 = \left((1-\hat{p}) J^{-1} + \hat{p}\right)(r-1)^{-1} \sum_{i \in s_r} (Y_i - \bar{Y}_r)^2, \tag{2.24}$$

with $\hat{p} = r/n$ and $\bar{Y}_r = r^{-1} \sum_{i \in s_r} Y_i$. The (1-$\alpha$)- level confidence interval on θ with asymptotically correct coverage probability can be constructed similarly.

2.3.2 Bootstrap Calibrated Confidence Intervals

We now use the adjusted bootstrap method to approximate the asymptotic distributions of $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$ given by (2.21) and (2.22) and construct bootstrap calibrated empirical likelihood confidence intervals. Let

$$Z_{b,i,m}(\hat{\mu}) = \delta_{b,i} Y_{b,i} + (1-\delta_{b,i}) J^{-1} \sum_{j=1}^{J} Y_{b,ij} - \overline{Y}_r, \tag{2.25}$$

and

$$Z_{b,i,d}(\hat{\theta}) = \delta_{b,i} I(Y_{b,i} \leq y) + (1-\delta_{b,i}) J^{-1} \sum_{j=1}^{J} I(Y_{b,ij} \leq y) - \bar{\theta}_r. \tag{2.26}$$

The proposed adjusted bootstrap analogs of $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$ are respectively given by

$$l_{b,n,m}(\hat{\mu}) = -2 \max_{p_1 \dots p_n} \left\{ \sum_{i=1}^{n} \log(np_i) \middle| \sum_{i=1}^{n} p_i Z_{b,i,m}(\hat{\mu}) = 0, \sum_{i=1}^{n} p_i = 1 \right\},$$

and

$$l_{b,n,d}(\hat{\theta}) = -2 \max_{p_1 \dots p_n} \left\{ \sum_{i=1}^{n} \log(np_i) \middle| \sum_{i=1}^{n} p_i Z_{b,i,d}(\hat{\theta}) = 0, \sum_{i=1}^{n} p_i = 1 \right\}.$$

It can be shown, using the Lagrange multiplier method, that

$$l_{b,n,m}(\hat{\mu}) = 2 \sum_{i=1}^{n} \log \{1 + \lambda_{b,n,m} Z_{b,i,m}(\hat{\mu})\},$$

and

$$l_{b,n,d}(\hat{\theta}) = 2 \sum_{i=1}^{n} \log \left\{1 + \lambda_{b,n,d} Z_{b,i,d}(\hat{\theta})\right\},$$

where, respectively, $\lambda_{b,n,m}$ and $\lambda_{b,n,d}$ are the solutions to

$$\frac{1}{n} \sum_{i=1}^{n} \frac{Z_{b,i,m}(\hat{\mu})}{1 + \lambda_{b,n,m} Z_{b,i,m}(\hat{\mu})} = 0 \text{ and } \frac{1}{n} \sum_{i=1}^{n} \frac{Z_{b,i,d}(\hat{\theta})}{1 + \lambda_{n,d} Z_{b,i,d}(\hat{\theta})} = 0.$$

Note that the ordinary bootstrap analogs of $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$, say $\tilde{l}_{n,m}(\hat{\mu})$ and $\tilde{l}_{n,d}(\hat{\theta})$, would be based on $\tilde{Z}_{b,i,m}(\hat{\mu}) = \delta_{b,i} Y_{b,i} + (1 - \delta_{b,i}) J^{-1} \sum_{j=1}^{J} Y_{b,ij} - \hat{\mu}$ and $\tilde{Z}_{b,i,d}(\hat{\theta}) = \delta_{b,i} I(Y_{b,i} \leq y) + (1 - \delta_{b,i}) J^{-1} \sum_{j=1}^{J} I(Y_{b,ij} \leq y) - \hat{\theta}$. However, for fixed J, the asymptotic distribution of $l_{n,m}(\mu)$ cannot be approximated by the asymptotic distribution of $\tilde{l}_{n,m}(\hat{\mu})$ especially in the case of random hot deck imputation; only when $J \to \infty$,

$\tilde{l}_{n,m}(\hat{\mu})$ could be used. Also, no adjustment would be needed under deterministic imputation, however deterministic imputation would result in inconsistent estimators of the distribution function.

Theorem 2.4 below states that $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$ can be respectively approximated by the proposed adjusted bootstrap analogs $l_{b,n,m}(\hat{\mu})$ and $l_{b,n,d}(\hat{\theta})$.

Theorem 2.4 *Suppose that the conditions in Theorem 2.1 are satisfied. Then as* $n \to \infty$,

$$\sup_{x \in R} \left| P_b \left\{ l_{b,n,m}(\hat{\mu}) \leq x \right\} - P \left\{ l_{n,m}(\mu) \leq x \right\} \right| \xrightarrow{P} 0, \tag{2.27}$$

and

$$\sup_{x \in R} \left| P_b \left\{ l_{b,n,d}(\hat{\theta}) \leq x \right\} - P \left\{ l_{n,d}(\theta) \leq x \right\} \right| \xrightarrow{P} 0. \tag{2.28}$$

Based on this result, the adjusted bootstrap EL confidence intervals on μ and θ can be constructed as follows. We repeat the bootstrap process independently B times and obtain $l^1_{b,n,m}(\hat{\mu}), ..., l^B_{b,n,m}(\hat{\mu})$ and $l^1_{b,n,d}(\hat{\theta}), ..., l^B_{b,n,d}(\hat{\theta})$. Let $l_{1-\alpha,m}$ and $l_{1-\alpha,d}$ be respectively the $100(1-\alpha)\%$ sample quantiles of $\left\{ l^k_{b,n,m}(\hat{\mu}), 1 \leq k \leq B \right\}$ and $\left\{ l^k_{b,n,d}(\hat{\theta}), 1 \leq k \leq B \right\}$. The $(1-\alpha)$-level adjusted bootstrap EL intervals on μ and θ are respectively given by

$$\left\{ \mu | \, l_{n,m}(\mu) \leq l_{1-\alpha,m} \right\} \text{ and } \left\{ \theta | \, l_{n,d}(\theta) \leq l_{1-\alpha,d} \right\}. \tag{2.29}$$

2.4 Simulation Study

We conducted a small simulation study to investigate the performance of the proposed bootstrap percentile (BP) and bootstrap empirical likelihood (EL) confidence intervals on the population mean and the distribution function. In particular, we compared the performance of the proposed adjusted bootstrap 95% confidence intervals versus their ordinary (unadjusted) counterparts based on two methods: the bootstrap percentile (BP) and the empirical likelihood (EL). Confidence intervals were examined in terms of their coverage probabilities and their average lengths. In our simulations, the precision of comparisons among the same test procedures at different settings was achieved by re-using the values of input random numbers, in the sense that the results were correlated by having common observations for each of the simulation runs.

We considered several scenarios with different $J-n-p-B$ combinations for the fractional imputation parameter (J), sample size (n), response probability (p) and number of bootstrap repetitions (B). The results were based on 2000 simulations programmed in R/S-PLUS. Note that the standard error for simulated coverage of the 95% confidence intervals was approximately 0.01 with 2000 simulation runs.

2.4.1 Data Frame

The population Y was generated from the standard exponential distribution. We assumed that Y_i is MCAR, that is $P(\delta_i = 1 \mid Y_i) = P(\delta_i = 1) = p$, $0 < p \leq 1$, and generated $\delta_i^{\prime}s$ as i.i.d. Bernoulli(p) random variables.

We formed confidence intervals for the population mean: $\mu = E(Y) = 1$ and the distribution function $F(y)$ for the following values of y: $F1 := F(0.2877) = 0.25$, $F2 := F(0.6932) = 0.50$, and $F3 := F(1.3863) = 0.75$

2.4.2 Confidence Intervals

The ordinary bootstrap versions of the confidence intervals were obtained by ignoring the proposed adjustments μ_{nb} and F_{nb} defined by (2.17). In particular, the 95% BP confidence interval for μ for the adjusted method was $(\hat{\mu} - P_{0.975},\ \hat{\mu} - P_{0.025})$, where P_α was the 100α percentile of the sampling distribution of $(\hat{\mu}_b - \bar{Y}_r)$. While for the ordinary method, P_α was based on the sampling distribution of $(\hat{\mu}_b - \hat{\mu})$. Similarly, for θ.Note that the lengths of the BP confidence intervals were the same under the ordinary and adjusted methods as the proposed adjustment, present in the upper and lower bounds, cancelled each other out in the calculation of the interval length.

The bisection method proposed by Wu (2005) was used to obtain $\lambda_{n,m,b}$ ($\lambda_{n,d,b}$) and find lower and upper bounds of the (1-α)% empirical likelihood confidence intervals using bootstrap sample percentiles of $\lambda_{n,m,b}$ ($\lambda_{n,d,b}$) as cut-off values of the χ^2_1 distribution. Note that the ordinary analogs of $Z_{b,i,m}(\hat{\mu})$ (2.25) and $Z_{b,i,d}(\hat{\theta})$ (2.26) after setting the adjustments (2.17) to zero are

$$\tilde{Z}_{b,i,m}(\hat{\mu}) = \delta_{b,i} Y_{b,i} + (1 - \delta_{b,i}) J^{-1} \sum_{j=1}^{J} Y_{b,ij} - \hat{\mu}, \tag{2.30}$$

and

$$\tilde{Z}_{b,i,d}(\hat{\theta}) = \delta_{b,i} I(Y_{b,i} \leq y) + (1 - \delta_{b,i}) J^{-1} \sum_{j=1}^{J} I(Y_{b,ij} \leq y) - \hat{\theta}. \tag{2.31}$$

2.4.3 Results

Table 2.1 displays the coverage probabilities and average lengths of the 95% confidence intervals for the population mean $\mu = E(Y)$. Tables 2.2, 2.3 and 2.4 display the coverage probabilities and average lengths of the 95% confidence intervals for the distribution function for $F1 = F(0.29) = 0.25$, $F2 = F(0.69) = 0.5$ and $F3 = F(1.39) = 0.75$ respectively. Coverage probabilities under the four methods with $J = 1$ are presented graphically as box plots in Figure 2.1.

Generally, in terms of coverage probabilities, the proposed adjusted BP and bootstrap EL methods resulted in smaller departures from the nominal level relative to their ordinary counterparts. We observe that the ordinary-BP method led to severe undercoverage of the confidence intervals for the mean under random imputation ($J = 1$). However, the BP interval coverage improved notably when the adjustment was considered. For example, with $p = 0.7$ and $n = 200$ the coverage was 88% under the ordinary method, versus 94% for the adjusted method. Under fractional imputation, the ordinary-BP method led to reasonable coverage probabilities, and again the adjusted-BP method resulted in coverage probabilities that were closer to nominal 95%. Using the same example scenario, now with $J = 5$, the coverage probabilities under the ordinary and adjusted BP methods were 92% and 93% respectively. Similar trends were observed for the distribution functions $F1$, $F2$ and $F3$, that is, the adjusted-BP confidence intervals had better coverage compared with the ordinary-BP intervals.

On the other hand, the ordinary-EL method led to slight overcoverage, especially for the cases under random imputation ($J = 1$). This overcoverage tendency was corrected by the proposed adjusted-EL method which led to coverage that was very close to the nominal 95% for most cases. For example, for the ordinary and adjusted EL methods the same scenario of $p = 0.7$ and $n = 200$ gave coverage probabilities of 97% and 95% respectively. For $J = 5$, the ordinary-EL method led to coverage close to 95%. The coverage of the ordinary-BP method was evidently poorer for scenarios with low response rates and $J = 1$ (it improved when the adjusted-BP method was used); in contrast, the EL methods performed well overall, even for scenarios with low response. Generally, in terms of confidence interval coverage probabilities, the adjusted-EL method outperformed all the other three methods. In terms of the average lengths of the confidence intervals, especially for the mean, the ordinary-EL method generated longer intervals compared with the BP method. However, the adjusted-EL method resulted in shorter intervals compared with the ordinary EL. For the distribution functions $F1$, $F2$ and $F3$, the ordinary-EL confidence intervals were only slightly longer than the corresponding BP intervals, while those based on the adjusted-EL method had similar lengths as the BP intervals.

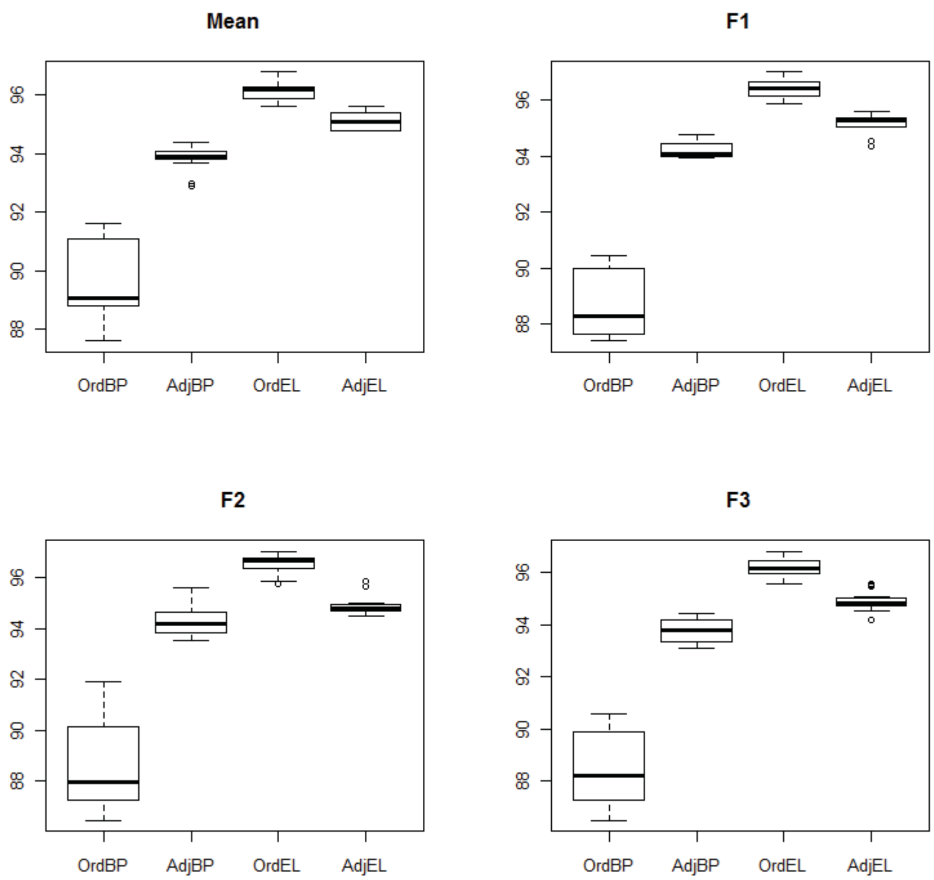

Figure 2.1: Box plot of bootstrap confidence interval coverage probabilities for all $n-p-B$ scenarios with J=1.

Table 2.1: Bootstrap confidence interval coverage probability and average interval length for the mean μ under random (J=1) and fractional (J=5) imputations with sample size n, response probability p, $B = 1000k$ bootstrap repetitions.

J	n	p	k	Coverage (%)				Average Length			
				OrdBP	AdjBP	OrdEL	AdjEL	OrdBP	AdjBP	OrdEL	AdjEL
1	120	0.7	1	88.8	93.8	96.3	95.6	0.462	0.462	0.522	0.485
			2	88.8	94.0	96.3	95.4	0.462	0.462	0.522	0.486
			3	89.0	93.9	96.3	95.3	0.462	0.462	0.522	0.486
		0.8	1	88.7	92.9	95.6	94.8	0.425	0.425	0.469	0.443
			2	88.9	93.0	95.9	95.0	0.425	0.425	0.470	0.444
			3	89.0	93.0	95.9	94.8	0.425	0.425	0.470	0.444
		0.9	1	91.1	93.7	96.0	95.2	0.388	0.388	0.420	0.405
			2	91.1	93.9	95.9	95.2	0.389	0.389	0.421	0.405
			3	90.9	93.8	95.9	95.2	0.388	0.388	0.421	0.405
	200	0.7	1	87.6	94.1	96.4	94.8	0.359	0.359	0.398	0.369
			2	87.7	93.9	96.6	94.9	0.359	0.359	0.399	0.370
			3	87.9	93.9	96.8	94.9	0.359	0.359	0.399	0.370
		0.8	1	89.2	94.2	96.3	94.9	0.330	0.330	0.360	0.339
			2	89.1	94.3	96.1	94.8	0.330	0.330	0.361	0.340
			3	89.3	94.4	96.0	94.8	0.330	0.330	0.361	0.340
		0.9	1	91.5	93.8	96.2	95.5	0.301	0.301	0.319	0.309
			2	91.6	94.1	96.2	95.5	0.301	0.301	0.320	0.309
			3	91.6	94.1	96.3	95.5	0.301	0.301	0.320	0.309
5	120	0.7	1	92.4	93.7	95.2	94.8	0.429	0.429	0.470	0.459
			2	92.7	93.7	95.5	95.0	0.430	0.430	0.470	0.460
			3	92.7	93.6	95.5	95.0	0.430	0.430	0.470	0.460
		0.8	1	92.0	92.8	95.0	94.6	0.401	0.401	0.430	0.424
			2	92.1	92.7	94.9	94.7	0.401	0.401	0.431	0.424
			3	91.9	92.7	95.1	94.5	0.401	0.401	0.431	0.424
		0.9	1	93.4	93.3	95.6	95.3	0.375	0.375	0.396	0.393
			2	93.4	93.3	95.7	95.5	0.375	0.375	0.397	0.393
			3	93.4	93.4	95.8	95.7	0.375	0.375	0.397	0.393
	200	0.7	1	91.7	93.2	95.2	94.4	0.334	0.334	0.355	0.348
			2	92.0	93.3	95.4	94.6	0.334	0.334	0.355	0.348
			3	92.1	93.4	95.5	94.6	0.334	0.334	0.356	0.348
		0.8	1	93.1	94.1	95.3	94.9	0.311	0.311	0.327	0.322
			2	93.1	94.4	95.2	94.8	0.312	0.312	0.327	0.322
			3	92.9	94.3	95.3	95.0	0.312	0.312	0.327	0.322
		0.9	1	93.9	94.0	95.5	95.3	0.291	0.291	0.302	0.300
			2	93.8	94.0	95.6	95.3	0.291	0.291	0.302	0.300
			3	94.0	94.1	95.5	95.3	0.291	0.291	0.302	0.300

Table 2.2: Bootstrap confidence interval coverage probability and average interval length for the mean $F1 = 0.25$ under random (J=1) and fractional (J=5) imputations with sample size n, response probability p, $B = 1000k$ bootstrap repetitions.

J	n	p	k	Coverage (%)				Average Length			
				OrdBP	AdjBP	OrdEL	AdjEL	OrdBP	AdjBP	OrdEL	AdjEL
1	120	0.7	1	87.5	94.1	96.7	95.2	0.201	0.201	0.213	0.200
			2	87.5	94.0	96.7	95.3	0.202	0.202	0.213	0.200
			3	87.6	94.0	96.8	95.4	0.202	0.202	0.214	0.200
		0.8	1	87.4	94.0	96.4	94.6	0.184	0.184	0.194	0.183
			2	87.8	94.0	96.3	94.6	0.185	0.185	0.194	0.184
			3	87.7	94.1	96.4	94.4	0.185	0.185	0.194	0.184
		0.9	1	90.3	94.2	95.9	95.3	0.169	0.169	0.174	0.168
			2	90.4	94.0	95.9	95.3	0.169	0.169	0.174	0.168
			3	90.5	94.1	95.9	95.3	0.169	0.169	0.174	0.168
	200	0.7	1	87.7	94.4	96.8	95.5	0.157	0.157	0.166	0.156
			2	87.7	94.7	96.9	95.6	0.157	0.157	0.167	0.156
			3	87.8	94.7	97.0	95.6	0.157	0.157	0.167	0.156
		0.8	1	88.8	94.5	96.5	95.1	0.143	0.143	0.151	0.143
			2	88.9	94.7	96.6	95.2	0.144	0.144	0.151	0.143
			3	88.9	94.8	96.6	95.1	0.144	0.144	0.151	0.143
		0.9	1	89.9	94.0	95.9	95.1	0.131	0.131	0.135	0.131
			2	90.1	94.1	96.2	95.4	0.131	0.131	0.135	0.131
			3	90.0	94.0	96.3	95.4	0.131	0.131	0.135	0.131
5	120	0.7	1	92.8	93.7	95.9	95.4	0.188	0.188	0.189	0.186
			2	92.6	93.9	95.8	95.3	0.188	0.188	0.189	0.186
			3	92.7	93.8	95.7	95.4	0.188	0.188	0.189	0.186
		0.8	1	92.8	94.1	95.0	94.6	0.175	0.175	0.176	0.173
			2	92.7	94.0	95.0	94.7	0.175	0.175	0.176	0.174
			3	93.1	94.1	95.1	94.5	0.175	0.175	0.176	0.174
		0.9	1	93.0	93.8	94.8	94.8	0.163	0.163	0.164	0.162
			2	93.2	93.9	94.9	94.8	0.164	0.164	0.164	0.163
			3	93.4	93.7	95.0	94.9	0.164	0.164	0.164	0.163
	200	0.7	1	92.0	93.8	95.7	95.3	0.146	0.146	0.147	0.145
			2	92.2	93.9	95.8	95.1	0.146	0.146	0.148	0.145
			3	92.2	94.0	95.8	95.2	0.146	0.146	0.148	0.145
		0.8	1	93.5	94.5	95.7	95.2	0.135	0.135	0.137	0.135
			2	93.2	94.5	95.7	95.4	0.136	0.136	0.137	0.135
			3	93.4	94.7	95.9	95.4	0.136	0.136	0.137	0.135
		0.9	1	93.4	94.3	95.3	95.2	0.127	0.127	0.128	0.127
			2	93.5	94.2	95.4	95.4	0.127	0.127	0.128	0.127
			3	93.3	94.2	95.5	95.3	0.127	0.127	0.128	0.127

Table 2.3: Bootstrap confidence interval coverage probability and average interval length for the mean $F2 = 0.5$ under random (J=1) and fractional (J=5) imputations with sample size n, response probability p, $B = 1000k$ bootstrap repetitions

J	n	p	k	Coverage (%)				Average Length			
				OrdBP	AdjBP	OrdEL	AdjEL	OrdBP	AdjBP	OrdEL	AdjEL
1	120	0.7	1	86.8	93.8	96.6	94.8	0.234	0.234	0.247	0.231
			2	86.6	93.9	96.8	94.7	0.234	0.234	0.247	0.231
			3	86.5	94.0	96.8	94.8	0.234	0.234	0.247	0.231
		0.8	1	87.3	93.6	96.2	94.7	0.214	0.214	0.224	0.211
			2	87.4	93.6	96.4	94.7	0.214	0.214	0.224	0.212
			3	87.3	93.7	96.4	94.6	0.214	0.214	0.224	0.212
		0.9	1	90.3	94.1	95.8	94.5	0.195	0.195	0.201	0.194
			2	90.2	94.2	95.9	94.8	0.195	0.195	0.201	0.194
			3	90.0	94.2	95.9	94.8	0.196	0.196	0.202	0.194
	200	0.7	1	87.6	94.3	96.8	94.6	0.181	0.181	0.193	0.180
			2	87.3	94.2	96.9	94.8	0.182	0.182	0.193	0.180
			3	87.6	94.2	97.1	94.9	0.182	0.182	0.193	0.180
		0.8	1	88.4	94.7	96.6	94.9	0.166	0.166	0.174	0.165
			2	88.5	94.7	96.7	95.0	0.166	0.166	0.175	0.165
			3	88.5	94.6	96.8	95.0	0.166	0.166	0.175	0.165
		0.9	1	91.8	95.6	96.8	95.9	0.152	0.152	0.156	0.151
			2	91.9	95.4	96.9	95.9	0.152	0.152	0.156	0.151
			3	91.8	95.3	96.9	95.7	0.152	0.152	0.157	0.151
5	120	0.7	1	93.1	94.2	95.8	95.4	0.217	0.217	0.218	0.214
			2	93.3	94.3	95.6	95.0	0.217	0.217	0.218	0.214
			3	93.5	94.4	95.6	95.1	0.217	0.217	0.218	0.214
		0.8	1	92.2	93.6	94.6	94.3	0.202	0.202	0.203	0.200
			2	92.2	93.7	94.6	94.4	0.202	0.202	0.203	0.200
			3	92.3	93.7	94.7	94.5	0.202	0.202	0.203	0.200
		0.9	1	93.1	93.8	94.2	94.2	0.189	0.189	0.189	0.188
			2	93.1	93.8	94.5	94.4	0.189	0.189	0.189	0.188
			3	93.3	93.8	94.6	94.5	0.189	0.189	0.190	0.188
	200	0.7	1	92.5	94.6	95.6	95.3	0.169	0.169	0.170	0.167
			2	92.9	94.8	95.7	95.3	0.169	0.169	0.170	0.167
			3	92.6	94.7	95.6	95.2	0.169	0.169	0.170	0.167
		0.8	1	92.9	94.2	95.4	94.9	0.157	0.157	0.158	0.156
			2	92.5	94.4	95.7	95.1	0.157	0.157	0.158	0.156
			3	92.5	94.6	95.5	94.9	0.157	0.157	0.158	0.156
		0.9	1	94.5	95.0	96.0	95.7	0.147	0.147	0.147	0.146
			2	94.4	95.1	96.0	95.7	0.147	0.147	0.147	0.146
			3	94.4	95.0	95.9	95.7	0.147	0.147	0.148	0.146

Table 2.4: Bootstrap confidence interval coverage probability and average interval length for the mean $F3 = 0.75$ under random (J=1) and fractional (J=5) imputations with sample size n, response probability p, $B = 1000k$ bootstrap repetitions.

J	n	p	k	Coverage (%)				Average Length			
				OrdBP	AdjBP	OrdEL	AdjEL	OrdBP	AdjBP	OrdEL	AdjEL
1	120	0.7	1	86.5	93.4	96.8	95.5	0.202	0.202	0.214	0.201
			2	87.0	93.7	96.7	95.5	0.202	0.202	0.214	0.201
			3	86.6	93.4	96.7	95.6	0.202	0.202	0.215	0.201
		0.8	1	87.3	93.1	96.2	94.2	0.185	0.185	0.194	0.183
			2	87.4	93.1	96.3	94.5	0.185	0.185	0.194	0.184
			3	87.3	93.2	96.4	94.5	0.185	0.185	0.194	0.184
		0.9	1	89.8	93.4	95.6	94.7	0.169	0.169	0.175	0.169
			2	90.0	93.5	95.9	94.8	0.169	0.169	0.175	0.169
			3	89.9	93.6	95.8	94.8	0.169	0.169	0.175	0.169
	200	0.7	1	88.0	94.0	96.5	94.9	0.157	0.157	0.168	0.156
			2	87.9	93.9	96.5	94.9	0.157	0.157	0.168	0.156
			3	88.0	94.1	96.5	95.1	0.157	0.157	0.168	0.156
		0.8	1	88.5	94.0	96.0	94.8	0.144	0.144	0.152	0.143
			2	88.8	94.2	96.0	95.0	0.144	0.144	0.152	0.143
			3	88.8	94.4	96.0	94.9	0.144	0.144	0.152	0.143
		0.9	1	90.4	94.3	95.9	94.9	0.131	0.131	0.135	0.131
			2	90.6	94.2	96.1	94.7	0.131	0.131	0.136	0.131
			3	90.6	94.3	96.0	94.8	0.131	0.131	0.136	0.131
5	120	0.7	1	92.3	94.0	95.1	94.4	0.188	0.188	0.189	0.186
			2	92.1	94.0	95.3	94.5	0.188	0.188	0.190	0.186
			3	92.3	94.0	95.3	94.6	0.188	0.188	0.190	0.186
		0.8	1	92.6	93.5	94.7	94.3	0.175	0.175	0.176	0.173
			2	92.5	93.5	94.7	94.4	0.175	0.175	0.176	0.174
			3	92.3	93.4	94.7	94.4	0.175	0.175	0.176	0.174
		0.9	1	93.0	93.5	94.9	94.7	0.164	0.164	0.164	0.163
			2	93.2	93.7	95.0	94.8	0.164	0.164	0.164	0.163
			3	93.2	93.7	95.1	94.9	0.164	0.164	0.165	0.163
	200	0.7	1	91.7	93.1	95.3	94.7	0.146	0.146	0.148	0.145
			2	91.9	93.3	95.1	94.9	0.146	0.146	0.148	0.145
			3	91.8	93.2	95.2	94.7	0.146	0.146	0.148	0.145
		0.8	1	92.5	94.0	94.9	94.6	0.136	0.136	0.137	0.135
			2	92.5	94.1	95.1	94.6	0.136	0.136	0.137	0.135
			3	92.7	94.0	95.0	94.7	0.136	0.136	0.137	0.135
		0.9	1	94.2	94.5	95.5	95.4	0.127	0.127	0.128	0.126
			2	94.4	94.7	95.6	95.4	0.127	0.127	0.128	0.127
			3	94.5	94.8	95.6	95.3	0.127	0.127	0.128	0.127

2.5 Conclusion

We proposed asymptotically correct adjusted bootstrap percentile and bootstrap empirical likelihood confidence intervals on the mean $\mu = E(Y)$ and the distribution function $F(y)$, $y \in R$, under fractional imputation. We constructed the adjusted confidence intervals based on bootstrap data obtained by imitating the process of imputing the original data set in bootstrap resampling. Our simulation study demonstrated that the proposed method led to better coverage and shorter confidence intervals for the mean and distribution function under fractional imputation. The EL method performed better than the BP method in terms of coverage. Therefore, we recommend that the proposed adjusted-EL confidence intervals for μ and θ should be used with fractionally-imputed data, particularly for small $J > 1$ or for random hot-deck imputation.

2.6 Appendix

2.6.1 Additional Theorems and Results

The following theorems and results will be used in the proofs of the theory presented in Chapters 2-4.

Theorem 2.5 (Chen and Rao) *(Chen and Rao, 2007)*

Let U_n and V_n be two sequences of random variables and B_n be a σ-algebra. Assume that: (1) There exists σ_{1n}>0 such that $\sigma_{1n}^{-1} V_n \xrightarrow{d} N(0,1)$ as $n \to \infty$ and V_n is B_n -measurable (2) $E[U_n|B_n] = 0$ and $Var(U_n|B_n) = \sigma_{2n}^2$ such that $\sup_t |P(\sigma_{2n}^{-1} U_n \leq$

$t|B_n) - \Phi(t)| = o_p(1)$ *(3)* $\gamma_n^2 = \frac{\sigma_{1n}^2}{\sigma_{2n}^2} = \gamma^2 + o_p(1)$. *Then as* $n \to \infty$,

$$\frac{U_n + V_n}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \xrightarrow{d} N(0,1).$$

Theorem 2.6 (Berry-Esseen) *Suppose* $X_{1,}X_2,...$ *are independent random variables with* $E[X_i] = 0$ *and* $Var(X_i) = \sigma_i^2 < \infty$. *Let* $B_n^2 = \sum_i \sigma_i^2 > 0$. *Then there exists a constant* $c > 0$ *such that*

$$\sup_t \left| P\left\{ \frac{1}{B_n} \sum X_i \leq t \right\} - \Phi(t) \right| \leq \frac{c}{B_n^3} \sum E|X_i|^3.$$

Theorem 2.7 (Polya) *(Gupta, 2008)*

If X *has a continuous distribution function then* $X_n \xrightarrow{d} X$ *if and only if*

$$\sup_{x \in R} |P\{X_n \leq x\} - P\{X \leq x\}| \to 0.$$

Theorem 2.8 (Slutsky) *(Serfling, 2002)*

Suppose that $X_n \xrightarrow{d} X$, *and* $Y_n \xrightarrow{P} c$. *Then,*

(i) $X_n + Y_n \xrightarrow{d} X + c$,

(ii) $X_n/Y_n \xrightarrow{d} X/c$ *provided that* $c \neq 0$,

(iii) $X_n Y_n \xrightarrow{d} cX$.

Special cases:

If $X_n \xrightarrow{d} X$ *and* $Y_n = o_p(1)$ *then* $X_n + Y_n \xrightarrow{d} X$.

If $X_n \xrightarrow{d} X$ *and* $Y_n = c + o_p(1)$, $c \neq 0$, *then* $X_n/Y_n \xrightarrow{d} X/c$.

Theorem 2.9 (Lebesgue's dominated convergence) *Suppose* $f_n : \mathcal{R} \to (-\infty, \infty)$

are measurable functions such that the pointwise limit $f(x) = \lim_{n\to\infty} f(n)$ *exists. Assume there is an integrable* $g : \mathcal{R} \to [0, \infty)$ *with* $|f_n(x)| \le g(x)$ *for each* $x \in \mathcal{R}$*. Then* f *is integrable as is* f_n *for each* n *and*

$$\lim_{n\to\infty} \int_{\mathcal{R}} f_n d\mu = \int_{\mathcal{R}} \lim_{n\to\infty} f_n d\mu = \int_{\mathcal{R}} f d\mu. \tag{2.32}$$

Proposition 2.10 (C_r−Inequality)

$$E|X+Y|^r \le c_r \left(E|X|^r + E|Y|^r\right), \tag{2.33}$$

$$\textit{where } c_r = \begin{cases} 1 \textit{ if } 0 < r \le 1, \\ 2^{r-1} \textit{ if } r > 1. \end{cases}$$

2.6.2 Proof of Theorem 2.1

The proof of Theorem 2.1 follows closely the proof of Theorem 2.1 in Qin et al. (2008) and is based on Theorem 2.5.

We start by forming the following decomposition

$$\hat{\mu} - \mu = \left(\hat{\mu} - E^*[\hat{\mu}]\right) + \left(E^*[\hat{\mu}] - \mu\right).$$

Define

$$U_n = \sqrt{n}\left(\hat{\mu} - E^*[\hat{\mu}]\right),$$

and

$$V_n = \sqrt{n}\left(E^*[\hat{\mu}] - \mu\right),$$

so that

$$\sqrt{n}(\hat{\mu} - \mu) = U_n + V_{n.} \tag{2.34}$$

Let us now show that the conditions of Chen and Rao Theorem are met. As shown in Section 2.2, $E^*[Y_{ij}] = \frac{1}{r}\sum_{i\in s_r} Y_i = \frac{1}{r}\sum_{i=1}^{n} \delta_i Y_i = \bar{Y}_r$ and $E^*[\hat{\mu}] = \bar{Y}_r$ where E^* denotes expectation with respect to randomness in the imputation process. Let $B_n = ((\delta_i, Y_i), i = 1, 2, ..., n)$ so that V_n is B_n-measurable. We have

$$\begin{aligned} V_n &= \sqrt{n}\left(E^*[\hat{\mu}] - \mu\right) = \sqrt{n}\left(\bar{Y}_r - \mu\right) \\ &= \sqrt{n}\left(\frac{1}{r}\sum_{i=1}^{n} \delta_i Y_i - \frac{1}{r}\sum_{i=1}^{n} \delta_i \mu\right) = \frac{1}{\sqrt{n}}\frac{n}{r}\left\{\sum_{i=1}^{n} \delta_i (Y_i - \mu)\right\} \\ &= \frac{1}{\sqrt{n}}\left(\frac{1}{p} + o_p(1)\right)\left\{\sum_{i=1}^{n} \delta_i (Y_i - \mu)\right\}, \end{aligned}$$

after noting that $\dfrac{r}{n} = p + o_p(1)$ where $p = P(\delta = 1)$.

We note that the random variables $\{\delta_i(Y_i - \mu), i = 1, ..., n\}$ are i.i.d. and, by the MCAR assumption,

$$E\left[\delta_i(Y_i - \mu)\right] = E[\delta_i Y_i] - E[\delta_i]\mu = p\mu - p\mu = 0$$

and

$$
\begin{aligned}
Var\left[\delta_i(Y_i-\mu)\right] &= E\left[\delta_i(Y_i-\mu)\right]^2 - E^2\left[\delta_i(Y_i-\mu)\right] \\
&= E\left[\delta_i Y_i\right]^2 - 2E[\delta_i Y_i \mu] + E\left[\delta_i \mu\right]^2 \\
&= pE\left[Y_i\right]^2 - 2p\mu^2 + p\mu^2 = p(\sigma^2+\mu^2) - p\mu^2 = p\sigma^2.
\end{aligned}
$$

That is, $E\left[V_n\right] = 0$ and $Var(V_n) = \sigma^2/p$. By the Central Limit Theorem, we have $\sqrt{n}\left(E^*[\hat{\mu}] - \mu\right) \xrightarrow{d} N(0, \frac{\sigma^2}{p})$ or

$$
\frac{\sqrt{n}V_n}{\sigma_{1n}} \xrightarrow{d} N(0,1), \tag{2.35}
$$

where $\sigma_{1n}^2 = Var(V_n)$. This verifies the first condition of Chen and Rao Theorem. Next, we will verify the second condition. Note that

$$
U_n = \sqrt{n}\left(\hat{\mu} - E^*[\hat{\mu}]\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ (1-\delta_i)\frac{1}{J}\sum_{j=1}^{J}\left(Y_{ij} - \bar{Y}_r\right) \right\}.
$$

It can be easily seen that $E^*[U_n|B_n] = 0$ and

$$
\begin{aligned}
Var^*[U_n|B_n] &= \frac{1}{n}\sum_{i=1}^{n}\left\{(1-\delta_i)Var^*(\frac{1}{J}\sum_{j=1}^{J}Y_{ij} - \bar{Y}_r)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{(1-\delta_i)\frac{1}{J}Var^*(Y_{ij} - \bar{Y}_r)\right\} \\
&= \frac{n-r}{nJ}\left(E^*(Y_{ij} - \bar{Y}_r)^2 - \left(E^*(Y_{ij} - \bar{Y}_r)\right)^2\right) \\
&= \frac{n-r}{nJ}\left(\frac{1}{r}\sum_{i\in s_r}(Y_i - E^*(Y_{ij}))^2\right) = \frac{(1-p)}{J}\sigma^2 + o_p(1).
\end{aligned}
$$

Let us denote $\sigma_{2n}^2 = Var^*[U_n|B_n]$. By the Berry-Esseen Theorem,

$$\sup_x \left| P^* \left(\frac{U_n}{\sigma_{2n}} \le x \right) - \Phi(x) \right| \le \frac{c\rho_n}{(\sigma_{2n}^2)^{3/2}}, \tag{2.36}$$

where

$$\rho_n = n^{-3/2} \sum_{i=1}^{n} (1-\delta_i) E^* \left| \left(\frac{1}{J} \sum_{j=1}^{J} Y_{ij} - E^* \left[\frac{1}{J} \sum_{j=1}^{J} Y_{ij} \right] \right) \right|^3 .$$

We note that by the C_r-inequality (2.33),

$$\begin{aligned}
\rho_n &= n^{-3/2} \sum_{i=1}^{n} (1-\delta_i) E^* \left| \frac{1}{J} \sum_{j=1}^{J} Y_{ij} - \bar{Y}_r \right|^3 \\
&\le Cn^{-3/2} \sum_{i=1}^{n} (1-\delta_i) \left(E^* \left| \frac{1}{J} \sum_{j=1}^{J} Y_{ij} \right|^3 + \left| \bar{Y}_r \right|^3 \right) \\
&\le Cn^{-3/2} \sum_{i=1}^{n} (1-\delta_i) \left(J^{-2} E^* \left| Y_{ij} \right|^3 + \left| E^* \left(Y_{ij} \right) \right|^3 \right) \\
&\le Cn^{-3/2}(n-r) \le Cn^{-1/2}.
\end{aligned}$$

Therefore, $c\rho_n(\sigma_{2n})^{-3} \le cn^{-1/2}$ and so

$$\sup_x \left| P^* \left(\frac{U_n}{\sigma_{2n}} \le x \right) - \Phi(x) \right| = o_p(1). \tag{2.37}$$

Hence, the conditions of Theorem 2.5 are satisfied and as $n \to \infty$,

$$\frac{U_n + V_n}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \xrightarrow{d} N(0,1). \tag{2.38}$$

Recall that $\sqrt{n}(\hat{\mu} - \mu) = U_n + V_n$ or

$$\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \xrightarrow{d} N(0, 1). \tag{2.39}$$

Applying Slutsky Theorem 2.8 to the denominator, we obtain as $n \to \infty$,

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N\left(0, \sigma_\mu^2\right),$$

with $\sigma_\mu^2 = \left\{p^{-1} + J^{-1}(1 - p)\right\} \sigma^2$.

The asymptotic normality result for θ can be proved in a similar way, by replacing Y_i with $I(Y_i \leq y)$. ■

2.6.3 Proof of Theorem 2.2

To prove that (2.15) of Theorem 2.2 holds, first, we decompose $\hat{\mu}_b - \overline{Y}_r$ into a sum $U_n + V_n + R_{n1}$. We verify the conditions of Chen and Rao Theorem about the limiting distributions of U_n and V_n using Berry-Esseen Theorem 2.6. Let $r_b = \sum_{i=1}^n \delta_{b,i}$ represent the number of respondents in the bootstrap sample $\mathcal{D}^* = \{(Y_{b,i}, \delta_{b,i}), b = 1, .., B, i = 1, ..., n\}$ and $\bar{\delta} = n^{-1} \sum_{i=1}^n \delta_i$ be the average number of respondents in the original sample $\mathcal{D} = \{(Y_i, \delta_i), i = 1, ..., n\}$.

Note that $\bar{\delta} = r/n = E_b\left[\delta_{b,i}\right]$, and $n^{-1} \sum_{i=1}^n \delta_i Y_i = E_b\left[\delta_{b,i} Y_{b,i}\right]$ where E_b denotes expectation taken with respect to the resampling distribution and conditional on $\mathcal{D}$. Also,

$$\frac{E_b\left[\delta_{b,i} Y_{b,i}\right]}{\bar{\delta}} = \frac{\sum_{i=1}^n \delta_i Y_i}{\sum_{i=1}^n \delta_i} = \overline{Y}_r.$$

We begin by forming the following decomposition

$$\hat{\mu}_b - \overline{Y}_r = \left(\hat{\mu}_b - E_b^*\left[\hat{\mu}_b\right]\right) + \left(E_b^*\left[\hat{\mu}_b\right] - \overline{Y}_r\right),$$

where E_b^* denotes expectation taken with respect to the imputation process for the bootstrap sample.

$$\begin{aligned} E_b^*\left[\hat{\mu}_b\right] &= E_b^*\left[n^{-1}\sum_{i=1}^{n}\left\{\delta_{b,i}Y_{b,i} + (1-\delta_{b,i})\frac{1}{J}\sum_{j=1}^{J}Y_{b,ij}\right\}\right] \\ &= n^{-1}\sum_{i=1}^{n}\delta_{b,i}Y_{b,i} + n^{-1}\sum_{i=1}^{n}(1-\delta_{b,i})\frac{1}{J}\sum_{j=1}^{J}E_b^*\left[Y_{b,ij}\right] \\ &= rn^{-1}\bar{Y}_{r,b} + (n-r)\,n^{-1}\bar{Y}_{r,b} = \bar{Y}_{r,b}. \end{aligned}$$

Using Taylor expansion, we obtain

$$\begin{aligned} E_b^*\left[\hat{\mu}_b\right] &= \frac{E_b\left[\delta_{b,i}Y_{b,i}\right]}{E_b\left[\delta_{b,i}\right]} - \frac{E_b\left[\delta_{b,i}Y_{b,i}\right]}{E_b^2\left[\delta_{b,i}\right]}n^{-1}\sum\left(\delta_{b,i} - E_b\left[\delta_{b,i}\right]\right) + \\ &\quad + \left(E_b\left[\delta_{b,i}\right]\right)^{-1}n^{-1}\sum\left(\delta_{b,i}Y_{b,i} - E_b\left[\delta_{b,i}Y_{b,i}\right]\right) + R_{n1} \\ &= \overline{Y}_r - \overline{Y}_r\bar{\delta}^{-1}n^{-1}\sum\left(\delta_{b,i} - E_b\left[\delta_{b,i}\right]\right) + \\ &\quad + \bar{\delta}^{-1}n^{-1}\sum\left(\delta_{b,i}Y_{b,i} - E_b\left[\delta_{b,i}Y_{b,i}\right]\right) + R_{n1} \\ &= \bar{\delta}^{-1}n^{-1}\sum\left(\delta_{b,i}\left(Y_{b,i} - \overline{Y}_r\right) - E_b\left[\delta_{b,i}\left(Y_{b,i} - \overline{Y}_r\right)\right]\right) + \bar{Y}_r + R_{n1}, \end{aligned}$$

where $P_b\left[\sqrt{n}|R_{n1}| > \varepsilon\right] \to 0$ a.s. in $[P_b]$ for any $\varepsilon > 0$.

Let

$$V_n = \bar{\delta}^{-1} n^{-1} \sum \left(\delta_{b,i} \left(Y_{b,i} - \overline{Y}_r \right) - E_b \left[\delta_{b,i} \left(Y_{b,i} - \overline{Y}_r \right) \right] \right),$$

and

$$U_n = (\hat{\mu}_b - E_b^* \left[\hat{\mu}_b \right]).$$

Note that $V_n + R_{n1} = E_b^* \left[\hat{\mu}_b \right] - \bar{Y}_r$. Hence, we have

$$\hat{\mu}_b - \overline{Y}_r = (\hat{\mu}_b - E_b^* \left[\hat{\mu}_b \right]) + \left(E_b^* \left[\hat{\mu}_b \right] - \overline{Y}_r \right) = U_n + V_n + R_{n1}. \tag{2.40}$$

Next, we will investigate the limiting distribution of $\sqrt{n} V_n$. Based on the bootstrap approximation procedure, $\{\delta_{b,i} \left(Y_{b,i} - \overline{Y}_r \right) - E_b \left[\delta_{b,i} \left(Y_{b,i} - \overline{Y}_r \right) \right], 1 \leq i \leq n\}$ is an i.i.d. random variable sequence given $\mathcal{D}$. By the MCAR assumption and the law of large numbers

$$\begin{aligned}
Var_b(\sqrt{n} V_n) &= nVar_b(\bar{\delta}^{-1} n^{-1} \sum \delta_{b,i} \left(Y_{b,i} - \overline{Y}_r \right)) \\
&= nr^{-2} nVar_b \left(\delta_{b,i} \left(Y_{b,i} - \overline{Y}_r \right) \right) \\
&= (n/r)^2 \left(E_b \left[\delta_{b,i} \left(Y_{b,i} - \overline{Y}_r \right) \right]^2 - E_b^2 \left[\delta_{b,i} \left(Y_{b,i} - \overline{Y}_r \right) \right] \right) \\
&= \bar{\delta}^{-2} \left(E_b \left[\delta_{b,i} Y_{b,i} \right]^2 - 2E_b[\delta_{b,i} Y_{b,i} \overline{Y}_r] + E_b \left[\delta_{b,i} \overline{Y}_r \right]^2 \right) \\
&= \bar{\delta}^{-2} \left(\bar{\delta} E_b \left[Y_{b,i} \right]^2 - 2\bar{\delta} \overline{Y}_r^2 + \bar{\delta} \overline{Y}_r^2 \right) \\
&= \bar{\delta}^{-1} Var_b \left[Y_{b,i} \right].
\end{aligned}$$

Noting that $\bar{\delta}^{-1} = p^{-1} + o_p(1)$, we have

$$\sigma_{1n}^2 := Var_b(\sqrt{n} V_n) = \sigma^2 p^{-1} + o_p(1).$$

By the Berry-Esseen Theorem, we have

$$\sup_{x\in R}\left|P_b\left(\frac{\sqrt{n}V_n}{\sigma_{1n}}\leq x\right)-\Phi(x)\right|\leq\frac{c\rho_n}{(\sigma_{1n}^2)^{3/2}n^{1/2}},$$

where

$$\sqrt{n}V_n=\sqrt{n}\bar{\delta}^{-1}n^{-1}\sum_{i=1}^{n}\left(\delta_{b,i}\left(Y_{b,i}-\overline{Y}_r\right)-E_b\left[\delta_{b,i}\left(Y_{b,i}-\overline{Y}_r\right)\right]\right),$$

and, by the C_r- inequality,

$$\begin{aligned}\rho_n n^{-1/2} &= n^{-1/2}\sum_{i=1}^{n}E_b\left|(r/n)^{-1}n^{-1}\left(\delta_{b,i}\left(Y_{b,i}-\overline{Y}_r\right)-E_b\left[\delta_{b,i}\left(Y_{b,i}-\overline{Y}_r\right)\right]\right)\right|^3\\ &\leq Cn^{-1/2}(r/n)^{-3}n^{-2}\left(E_b\left|\delta_{b,i}\left(Y_{b,i}-\overline{Y}_r\right)\right|^3+\left|E_b\left[\delta_{b,i}\left(Y_{b,i}-\overline{Y}_r\right)\right]\right|^3\right)\\ &= Cn^{-1/2}(r/n)^{-3}n^{-2}(r/n)\left(E_b\left|Y_{b,i}-\overline{Y}_r\right|^3+\left|E_b\left[\delta_{b,i}\left(Y_{b,i}-\overline{Y}_r\right)\right]\right|^3\right)\\ &\leq Cn^{-1/2}r^{-2}\leq Cn^{-1/2}.\end{aligned}$$

Therefore,

$$\sup_x\left|P_b\left\{\sqrt{n}V_n\sigma_{1n}^{-1}\leq x\right\}-\Phi(x)\right|\xrightarrow{P}0. \tag{2.41}$$

By Polya Theorem,

$$\sqrt{n}V_n\sigma_{1n}^{-1}\xrightarrow{d}N(0,1).$$

Next, we want to show that

$$\sup_x|P_b^*\left(\sqrt{n}U_n\sigma_{2n}^{-1}\leq x\right)-\Phi(x)|\to 0,\ a.s.[P], \tag{2.42}$$

where

$$\begin{aligned}\sqrt{n}U_n &= \sqrt{n}\left(\hat{\mu}_b - E_b^*[\hat{\mu}_b]\right)\\ &= n^{-1/2}\sum_{i=1}^{n}\left\{(1-\delta_{b,i})\left(J^{-1}\sum_{j=1}^{J}Y_{b,ij} - E_b^*\left[J^{-1}\sum_{j=1}^{J}Y_{b,ij}\right]\right)\right\}\end{aligned}$$

and

$$\begin{aligned}\sigma_{2n}^2 &= Var_b^*[\sqrt{n}U_n] = Var_b^*\left[n^{-1/2}\sum_{i=1}^{n}\left\{(1-\delta_{b,i})J^{-1}\sum_{j=1}^{J}Y_{b,ij} - \overline{Y}_{r,b}\right\}\right]\\ &= n^{-1}J^{-2}\sum_{i=1}^{n}\left\{(1-\delta_{b,i})\sum_{j=1}^{J}Var_b^*(Y_{b,ij})\right\}\\ &= (n-r_b)\,n^{-1}J^{-1}\left(r_b^{-1}\sum_{i=1}^{n}\delta_{b,i}Y_{b,i} - \left(r_b^{-1}\sum_{i=1}^{n}\delta_{b,i}Y_{b,i}\right)^2\right),\end{aligned}$$

and so $\sigma_{2n}^2 = (1-p)J^{-1}\sigma^2 + o_p(1)$ a.s. [P]. By the Berry-Esseen Theorem,

$$\sup_x\left|P_b^*\left(\frac{\sqrt{n}U_n}{\sigma_{2n}} \le x\right) - \Phi(x)\right| \le \frac{c\rho_n^*}{(\sigma_{2n})^3}, \tag{2.43}$$

where $\rho_n^* = n^{-3/2}\sum_{i=1}^{n}(1-\delta_{b,i})E_b^*\left|\left(J^{-1}\sum_{j=1}^{J}Y_{b,ij} - E_b^*\left[J^{-1}\sum_{j=1}^{J}Y_{b,ij}\right]\right)\right|^3$. We note that, by the C_r- inequality again,

$$\begin{aligned}
\rho_n^* &= \sum_{i=1}^{n} E_b^* \left| n^{-1/2}(1-\delta_{b,i}) \left(J^{-1}\sum_{j=1}^{J} Y_{b,ij} - E_b^* \left[J^{-1}\sum_{j=1}^{J} Y_{b,ij} \right] \right) \right|^3 \\
&= n^{-3/2}\sum_{i=1}^{n}(1-\delta_{b,i})E_b^* \left| J^{-1}\sum_{j=1}^{J} Y_{b,ij} - E_b^* \left[J^{-1}\sum_{j=1}^{J} Y_{b,ij} \right] \right|^3 \\
&\leq Cn^{-3/2}\sum_{i=1}^{n}(1-\delta_{b,i}) \left(E_b^* \left| J^{-1}\sum_{j=1}^{J} Y_{b,ij} \right|^3 + \left| E_b^* \left[J^{-1}\sum_{j=1}^{J} Y_{b,ij} \right] \right|^3 \right) \\
&\leq Cn^{-3/2}(n-r_b) \leq Cn^{-1/2}.
\end{aligned}$$

Hence, result (2.41). So, by Chen and Rao Theorem, as $n \to \infty$,

$$\frac{\sqrt{n}\,(U_n+V_n)}{\sqrt{\sigma_{1n}^2+\sigma_{2n}^2}} \xrightarrow{d} N(0,1). \tag{2.44}$$

Recall that $\sqrt{n}(\hat{\mu}_b - \overline{Y}_r) = \sqrt{n}\,(U_n+V_n) + \sqrt{n}R_{n1}$. Since it was assumed that $P_b\left[\sqrt{n}|R_{n1}| > \varepsilon\right] \to 0$ a.s in [P] for any $\varepsilon > 0$, Cràmer Convergence theorem, together with (2.44), give

$$\frac{\sqrt{n}(\hat{\mu}_b - \overline{Y}_r)}{\sqrt{\sigma_{1n}^2+\sigma_{2n}^2}} \xrightarrow{d} N(0,1).$$

Polya's theorem gives

$$\sup_x \left| P_b\left(\frac{\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right)}{\sqrt{\sigma_{1n}^2+\sigma_{2n}^2}} \leq x \right) - \Phi(x) \right| \to 0.$$

Applying Slutsky Theorem 2.8 to the denominator, we obtain

$$\sup_x \left| P_b\left(\frac{\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right)}{\sqrt{\sigma^2/p+(1-p)\sigma^2/J}} \leq x \right) - \Phi(x) \right| \to 0.$$

From Theorem 2.1 we have, as $n \to \infty$, $\sqrt{n}(\hat{\mu}-\mu) \xrightarrow{d} \Phi_{\sigma_\mu^2}(x)$ where $\Phi_{\sigma_\mu^2}(x)$ denotes $N\left(0,\sigma_\mu^2\right)$ with $\sigma_\mu^2 = \left(\frac{1}{p}+\frac{1-p}{J}\right)\sigma^2$, and so by Polya Theorem 2.7, we have

$$\sup_x \left| P_b\left(\frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{\sigma^2/p+(1-p)\sigma^2/J}} \le x\right) - \Phi(x)\right| \to 0.$$

Now, since

$$\begin{aligned}
& \sup_x \left| P_b\left(\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right) \le x\right) - P\left(\sqrt{n}\left(\hat{\mu}-\mu\right) \le x\right)\right| \\
= & \sup_x \left| P_b\left(\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right) \le x\right) - \Phi_{\sigma_\mu^2}(x) + \Phi_{\sigma_\mu^2}(x) - P\left(\sqrt{n}\left(\hat{\mu}-\mu\right) \le x\right)\right| \\
\le & \sup_x \left| P_b\left(\frac{\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right)}{\sigma_\mu} \le x\right) - \Phi(x)\right| + \sup_x \left| P\left(\frac{\sqrt{n}\left(\hat{\mu}-\mu\right)}{\sigma_\mu} \le x\right) - \Phi(x)\right|,
\end{aligned}$$

we obtain

$$\sup_x \left| P_b\left(\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right) \le x\right) - P\left(\sqrt{n}\left(\hat{\mu}-\mu\right) \le x\right)\right| \xrightarrow{P} 0. \blacksquare$$

2.6.4 Proof of Theorem 2.3

The proof of Theorem 2.3 follows closely the proof of Theorem 3.1 in Qin et al. (2008) and is based on Owen (1990). Denote $\overline{Z}^2_{i,m}(\mu) = \frac{1}{n}\sum_{i=1}^n \left(Z_{i,m}(\mu)\right)^2$, that is

$$\begin{aligned}
\bar{Z}^2_{i,m}(\mu) &= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i Y_i + (1-\delta_i)J^{-1}\sum_{j=1}^{J} Y_{ij} - \mu\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i \left(Y_i - \mu\right)^2 + (1-\delta_i)J^{-2}\sum_{j=1}^{J} \left(Y_{ij} - \mu\right)^2\right) \\
&= \frac{r}{n}\frac{1}{r}\sum_{i\in s_r}\left(Y_i - \mu\right)^2 + \frac{n-r}{n}J^{-2}\sum_{j=1}^{J}\frac{1}{n-r}\sum_{i\in s_m}\left(Y_{ij} - \mu\right)^2 \\
&= p\sigma^2 + o_p(1) + J^{-1}(1-p)\sigma^2 + o_p(1) = \sigma^2\left(p + J^{-1}(1-p)\right) + o_p(1),
\end{aligned}$$

and so

$$\bar{Z}^2_{i,m}(\mu) = \sigma_1^2 + o_p(1). \tag{2.45}$$

From Theorem 2.1, we have $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N\left(0, \sigma^2_\mu\right)$ where $\sigma^2_\mu = \frac{1}{p} + \frac{1}{J}(1-p)\sigma^2$. That is, denoting $\overline{Z}_{i,m}(\mu) = \frac{1}{n}\sum_{i=1}^{n} Z_{i,m}(\mu)$, we have

$$n^{1/2}\sigma_\mu^{-1}\bar{Z}_{i,m}(\mu) \xrightarrow{d} N\left(0,1\right). \tag{2.46}$$

Assume $E\left[Y_i - \mu\right]^2 < \infty$. Similarly as in Owen (1990), we note that this condition implies $\sum P\left(\left(Y_i - \mu\right)^2 > n\right) < \infty$ and thus $\sum P\left(|Y_i - \mu| > n^{1/2}\right) < \infty$. Then, by the Borel-Cantelli Theorem, $|Y_i - \mu| > n^{1/2}$ finitely often with probability 1. This implies $|\underset{1\le i\le n}{max}(Y_i - \mu)| > n^{1/2}$ finitely often or $|\underset{1\le i\le n}{max} Z_{i,m}(\mu)| > n^{1/2}$ finitely often. Similarly, $|Z_{i,m}(\mu)| > cn^{1/2}$ finitely often for any $c > 0$. Therefore, $\limsup |Z_{i,m}(\mu)|/\sqrt{n} \le c$ with probability 1. This holds for any countable set of values c, hence, with probability 1,

$$\underset{1\le i\le n}{max}|Z_{i,m}(\mu)| = o_p(n^{1/2}). \tag{2.47}$$

Following the steps of the proof of Theorem 1 in Owen (1990), we will now show that $\left(\sigma_\mu^2/\sigma_1^2\right)^{-1} l_{n,m}(\mu) \xrightarrow{d} \chi_1^2$. Define $\gamma_i := \lambda_{n,m} Z_{i,m}(\mu)$. We have

$$\begin{aligned} l_{n,m}(\mu) &= -2 \max_{\sum p_i Z_{im}(\mu)=0, \sum p_i=1} \sum_{i=1}^n \log(np_i) \\ &= 2\sum_{i=1}^n \log\left\{1+\lambda_{n,m} Z_{i,m}(\mu)\right\} = 2\sum_{i=1}^n \log\left\{1+\gamma_i\right\} \\ &= 2\sum_{i=1}^n \gamma_i - \sum_{i=1}^n \gamma_i^2 + 2\sum_{i=1}^n \eta_i, \end{aligned}$$

where, for some $0 < c < \infty$, $P\left[|\eta_i| \le c|\gamma_i|^3, 1 \le i \le n\right] \to 1$ as $n \to \infty$. The last expression of the equation is based on the Maclaurin series expansion as we have

$$max_{1\le i\le n}|\gamma_i| = max_{1\le i\le n}|\lambda_{n,m} Z_{i,m}(\mu)| = O_p(n^{-1/2}) o_p(n^{1/2}) = o_p(1),$$

which results from $||\lambda_{n,m}|| = O_p(n^{-1/2})$ and (2.47).

Owen (1990) expands $g(\lambda_{n,m})$ as follows

$$\begin{aligned} g(\lambda_{n,m}) &= \frac{1}{n}\sum_{i=1}^n \frac{Z_{i,m}(\mu)}{1+\lambda_{n,m} Z_{i,m}(\mu)} \\ &= \frac{1}{n}\sum_{i=1}^n \frac{Z_{i,m}(\mu)}{1+\gamma_i} \\ &= \frac{1}{n}\sum_{i=1}^n Z_{i,m}(\mu)\left(1-\gamma_i+\frac{\gamma_i^2}{1-\gamma_i}\right) \\ &= \overline{Z}_{i,m}(\mu) - \overline{Z}^2_{i,m}(\mu)\lambda_{n,m} + \frac{1}{n}\sum_{i=1}^n Z_{i,m}(\mu)\frac{\gamma_i^2}{1-\gamma_i} \\ &= \overline{Z}_{i,m}(\mu) - \overline{Z}^2_{i,m}(\mu)\lambda_{n,m} + \beta, \end{aligned}$$

where $||\beta|| = o_p(n^{-1/2})$ since, as shown in Owen (1990),

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n} Z_{i,m}(\mu)\frac{\gamma_i^2}{1-\gamma_i}\right\| &\leq \frac{1}{n}\sum_{i=1}^{n} ||Z_{i,m}(\mu)||^3||\lambda||^2 \left\|\frac{1}{1-\gamma_i}\right\| \\ &= o(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2}).\end{aligned}$$

Therefore, since $g(\lambda_{n,m}) := 0$

$$\lambda_{n,m} = \left(\overline{Z}^2_{i,m}(\mu)\right)^{-1} \bar{Z}_{i,m}(\mu) + \beta.$$

Going back to the equation for $l_{n,m}(\mu)$, back substituting for γ_i and expanding $\lambda_{n,m}$, we obtain

$$\begin{aligned} l_{n,m}(\mu) &= 2n\lambda_{n,m}\overline{Z}_{i,m}(\mu) - n\lambda^2_{n,m}\overline{Z}^2_{i,m}(\mu) + 2\sum_{i=1}^{n}\eta_i \\ &= 2n\left(\overline{Z}_{i,m}(\mu)\right)^2\left(\overline{Z}^2_{i,m}(\mu)\right)^{-1} + 2n\beta\overline{Z}_{i,m}(\mu) - n\left(\overline{Z}_{i,m}(\mu)\right)^2\left(\overline{Z}^2_{i,m}(\mu)\right)^{-1} + \\ &\quad -2n\beta\overline{Z}_{i,m}(\mu) - n\beta^2\left(\overline{Z}^2_{i,m}(\mu)\right)^{-1} + 2\sum_{i=1}^{n}\eta_i \\ &= n\left(\overline{Z}_{i,m}(\mu)\right)^2\left(\overline{Z}^2_{i,m}(\mu)\right)^{-1} - n\beta^2\overline{Z}^2_{i,m}(\mu)^{-1} + 2\sum_{i=1}^{n}\eta_i \\ &= n\left(\overline{Z}_{i,m}(\mu)\right)^2\left(\overline{Z}^2_{i,m}(\mu)\right)^{-1} + R_{n1}. \end{aligned}$$

Next, we note that $n\beta^2\left(\overline{Z}^2_{i,m}(\mu)\right)^{-1} = o_p(1)$ and by definition of η_i,

$$\left|2\sum_{i=1}^{n}\eta_i\right| \leq 2c||\lambda||^3\sum_{i=1}^{n}||Z_{i,m}(\mu)||^3 = 2cO_p(n^{-3/2})o_p(n^{3/2}) = o_p(1).$$

Therefore, $P\left[|R_{n1}| > \epsilon\right] \to 0$ a.s. [P]. Also,

$$\frac{n\left(\overline{Z}_{i,m}(\mu)\right)^2}{\overline{Z}^2_{i,m}(\mu)} = \frac{n\left(\frac{1}{n}\sum_{i=1}^{n} Z_{i,m}(\mu)\right)^2}{n^{-1}\sum_{i=1}^{n}\left(Z_{i,m}(\mu)\right)^2},$$

and using results (2.45) and (2.46), we have $\left(\sigma^2_\mu/\sigma^2_1\right)^{-1} l_{n,m}(\mu) \xrightarrow{d} \chi^2_1$ as $n \to \infty$. That is, $l_{n,m}(\mu) \xrightarrow{d} c_m\chi^2_1$.

The results corresponding to θ can be proved similarly to those for μ with Y replaced by $I\left(Y \leq y\right)$.

2.6.5 Proof of Theorem 2.4

Similarly to the Proof of Theorem 2.3, we first show

$$\begin{aligned}
\overline{Z}^2_{b,i,m}(\widehat{\mu}) \quad : \quad &= \frac{1}{n}\sum_{i=1}^{n}\left(Z_{b,i,m}(\widehat{\mu})\right)^2 = \frac{1}{n}\sum_{i=1}^{n} E^*_b\left[Z_{b,i,m}(\widehat{\mu})\right]^2 + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(Var^*_b\left[Z_{b,i,m}(\widehat{\mu})\right] + \left(E^*_b\left[Z_{b,i,m}(\widehat{\mu})\right]\right)^2\right) + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(Var^*_b\left[\delta_{b,i}\left(Y_{b,i} - \overline{Y}_r\right) + (1-\delta_{b,i})\frac{1}{J}\sum_{j=1}^{J}\left(Y_{b,ij} - \overline{Y}_r\right)\right]\right. \\
&\quad \left. + \left(E^*_b\left[\delta_{b,i}\left(Y_{b,i} - \overline{Y}_r\right) + (1-\delta_{b,i})\frac{1}{J}\sum_{j=1}^{J}\left(Y_{b,ij} - \overline{Y}_r\right)\right]\right)^2\right) + o_p(1) \\
&= \frac{1}{J}\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{b,i})Var^*_b\left[Y_{b,ij}\right] + \frac{1}{n}\sum_{i=1}^{n}\delta_{b,i}\left(Y_{b,i} - \overline{Y}_r\right)^2 + o_p(1) \\
&= \frac{1}{J}(1-p)\sigma^2 + o_p(1) + p\sigma^2 + o_p(1) \\
&= \sigma^2\left(p + \frac{1}{J}(1-p)\right) + o_p(1).
\end{aligned}$$

That is,

$$\overline{Z}^2_{b,i,m}(\widehat{\mu}) = \sigma_1^2 + o_p(1). \tag{2.48}$$

From the Proof of Theorem 2.2, we have as $n \to \infty$,

$$\frac{\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right)}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \xrightarrow{d} N(0,1),$$

where $\sigma_{1n}^2+\sigma_{2n}^2 = \frac{1}{p}+\frac{1}{J}(1-p)\sigma^2 := \sigma_\mu^2$. That is, denoting $\bar{Z}_{b,i,m}(\widehat{\mu}) := \frac{1}{n}\sum_{i=1}^n Z_{b,i,m}(\widehat{\mu})$, we have

$$\frac{\sqrt{n}\bar{Z}_{b,i,m}(\widehat{\mu})}{\sqrt{\sigma_\mu^2}} \xrightarrow{d} N(0,1). \tag{2.49}$$

Assume $E\left[Y_i - \overline{Y}_r\right]^2 < \infty$. Similarly to as it was shown in Owen (1990) and in the Proof of Theorem 2.3, we obtain

$$\max_{1\le i\le n} |Z_{b,i,m}(\widehat{\mu})| = o_p(n^{1/2}). \tag{2.50}$$

Continuing to follow the steps of the Proof of Theorem 1 in Owen (1990), and similar to the Proof of Theorem 2.3, we can show that

$$\begin{aligned} l_{b,n,m}(\widehat{\mu}) &= 2\sum_{i=1}^n \log\left\{1 + \lambda_{b,n,m} Z_{b,i,m}(\widehat{\mu})\right\} \\ &= \frac{\left(\sqrt{n}\overline{Z}_{b,i,m}(\widehat{\mu})\right)^2}{\overline{Z}^2_{b,i,m}(\widehat{\mu})} + R_{n2}, \end{aligned}$$

with $P\left[|R_{n2}| > \epsilon\right] \to 0$ a.s. [P]. Also

$$\frac{\left(\sqrt{n}\overline{Z}_{b,i,m}(\widehat{\mu})\right)^2}{\overline{Z}^2_{b,i,m}(\widehat{\mu})} = \frac{\sigma^2_\mu n\left(\sqrt{n}\bar{Z}_{b,i,m}(\widehat{\mu})\left(\sigma^2_\mu\right)^{-1/2}\right)^2}{\overline{Z}^2_{b,i,m}(\widehat{\mu})},$$

and using results (2.48) and (2.49), we have

$$\sup_{x\in R}\left|P_b\left\{\frac{\overline{Z}^2_{b,i,m}(\widehat{\mu})}{\sigma^2_\mu}l_{b,n,m}(\hat{\mu}) \le x\right\} - P\left\{\chi^2_1 \le x\right\}\right| \xrightarrow{P} 0,$$

where $\left(\overline{Z}^2_{b,i,m}(\widehat{\mu})/\sigma^2_\mu\right)^{-1} = \sigma^2_\mu/\sigma^2_1 + o_p(1) = c_m + o_p(1)$ as defined in Theorem 2.3. By Theorem 2.3, $l_{n,m}(\mu) \xrightarrow{d} c_m\chi^2_1$, therefore

$$\sup_{x\in R}\left|P_b\left\{l_{b,n,m}(\hat{\mu}) \le x\right\} - P\left\{l_{n,m}(\mu) \le x\right\}\right| \xrightarrow{P} 0.$$

We follow the same steps to prove the result for distribution function.■

Chapter 3

Confidence Intervals for Population Mean and Distribution Function with Imputation Classes

3.1 Introduction

To improve the accuracy of imputation in practice, units are often divided into homogenous groups, called imputation classes, such that the missing values can be imputed independently, using separate imputation procedures within each class. Haziza and Beaumont (2007) compared different methods that can be used to construct imputation classes. In this chapter, we extend the theory introduced in Chapter 2 and form bootstrap percentile and bootstrap-calibrated empirical likelihood confidence intervals on the mean $\mu = E(Y)$ and the distribution function $\theta = F(y) = P(Y \leq y)$, $y \in R$, based on data with imputation classes.

The chapter is organized as follows. In Section 3.2, we establish the asymptotic normality of fractionally imputed estimators $\hat{\mu}$ and $\hat{\theta}$, and construct asymptotically valid bootstrap percentile confidence intervals on μ and θ. In Section 3.3, we obtain the empirical likelihood ratio statistics and their limiting distributions to construct asymptotically valid bootstrap-calibrated empirical likelihood confidence intervals on μ and θ. We report the results of a small simulation study on the finite sample performance of the proposed confidence intervals in Section 3.4. The proofs of theoretical results are deferred to the Appendix.

3.1.1 Framework

In this chapter we suppose that the population Y can be divided into subpopulations, called imputation classes, $\mathcal{P}_s$, $s = 1, ..., S$, according to the values of an auxiliary variable X with known distribution, therefore the classes are fixed, and that all $\{Y_{si}, i = 1, ..., n_s,\ s = 1, ..., S\ \}$ have the same distribution as the population Y. We assume a missing completely at random (MCAR) mechanism within each class. That is, the class response probability is given by

$$p_s = P(\delta_{si} = 1|Y_{si}) = P(\delta_{si} = 1), 0 < p_s \leq 1, \tag{3.1}$$

where $\delta_{si} = 1$ if Y_{si} is observed and $\delta_{si} = 0$ if Y_{si} is missing, $s = 1, ..., S$. Thus, we have an i.i.d. sample of incomplete data in $\mathcal{P}_s$: $\{(Y_{si}, \delta_{si}), i = 1, ..., n_s\}$ and all $\{(Y_{si}, \delta_{si}), i = 1, ..., n_s\}, s = 1, ..., S$, are independent with random frequencies $\{n_s\}$

such that $\sum_{s=1}^{S} n_s = n$. We also assume that

$$W_s = P(Y \in \mathcal{P}_s) > 0. \tag{3.2}$$

Further, we define

$$s_{rs} = \{i : \delta_{si} = 1\} \text{ and } s_{ms} = \{i : \delta_{si} = 0\} \tag{3.3}$$

which respectively are the sets of respondents and non-respondents in class s, and let r_s denote the number of respondents in class s, that is

$$r_s = \sum_{i=1}^{n_s} \delta_{si}, \;\; s = 1, ..., S. \tag{3.4}$$

3.1.2 Fractional Imputation with Imputation Classes

After the data are partitioned into S imputation classes, within each class, missing values are replaced by the values selected randomly from the set of respondents within that class. Similarly as in Chapter 2, we use the fractional imputation method (Kim and Fuller, 2004); however here, the procedure is done separately in each class and independently across classes. In particular, for class s, we generate $J \geq 1$ imputed values $Y_{sij} = \bar{Y}_{rs} + \epsilon^*_{sij}$, $j = 1, ..., J$, for each missing Y_{si} where $\{\epsilon^*_{sij}, \; j = 1, ..., J\}$ are drawn by simple random sampling with replacement from the set of donor residuals $\{\hat{\epsilon}_{si} = Y_{si} - \bar{Y}_{rs}, i \in s_{rs}\}$ formed within class s with

$$\bar{Y}_{rs} = \frac{\sum_{i=1}^{n_s} \delta_{si} Y_{si}}{r_s}. \tag{3.5}$$

The fractionally-imputed data consists of $\{(\tilde{Y}_{si}, \delta_{si}); i = 1, 2, ..., n_s, s = 1, ..., S\}$, where $\tilde{Y}_{si} = Y_{si}$ if $\delta_{si} = 1$ or $\tilde{Y}_{si} = (Y_{si1}...Y_{siJ})$ if $\delta_{si} = 0$.

3.2 Normal Approximation Confidence Intervals

The estimators of the mean μ and distribution function θ under fractional imputation with imputation classes are respectively given by

$$\hat{\mu} = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left\{\delta_{si}Y_{si} + (1-\delta_{si})\frac{1}{J}\sum_{j=1}^{J}Y_{sij}\right\}, \tag{3.6}$$

and

$$\hat{\theta} = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left\{\delta_{si}I(Y_{si} \leq y) + (1-\delta_{si})\frac{1}{J}\sum_{j=1}^{J}I(Y_{sij} \leq y)\right\}. \tag{3.7}$$

Let E^* denote the expectation with respect to randomness in the imputation procedure and define $\bar{Y}_r = \frac{1}{n}\sum_{s=1}^{S} n_s\bar{Y}_{rs}$. Since, $E^*[Y_{sij}] = r_s^{-1}\sum_{i=1}^{n_s}\delta_{si}Y_{si} = \bar{Y}_{rs}$, we have

$$\begin{aligned} E^*[\hat{\mu}] &= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\delta_{si}Y_{si} + \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{si})\frac{1}{J}\sum_{j=1}^{J}\bar{Y}_{rs} && (3.8)\\ &= \frac{1}{n}\sum_{s=1}^{S}r_s\bar{Y}_{rs} + \frac{1}{n}\sum_{s=1}^{S}(n_s - r_s)\bar{Y}_{rs} = \frac{1}{n}\sum_{s=1}^{S}n_s\bar{Y}_{rs} = \bar{Y}_r. && (3.9)\end{aligned}$$

Similarly for the distribution function, $E^*[I(Y_{sij} \leq y)] = r_s^{-1}\sum_{i=1}^{n}\delta_{si}I(Y_{si} \leq y) :=$

$\bar{\theta}_{rs}$, so that $E^*[\hat{\theta}] = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left(\delta_{si}I(Y_{si} \leq y) + (1-\delta_{si})\bar{\theta}_{rs}\right) = \bar{\theta}_r$. Here

$$\bar{Y}_r = n^{-1}\sum_{s=1}^{S} n_s r_s^{-1}\sum_{i=1}^{n_s}\delta_{si}Y_{si}, \tag{3.10}$$

and

$$\bar{\theta}_r = n^{-1}\sum_{s=1}^{S} n_s r_s^{-1}\sum_{i=1}^{n_s}\delta_{si}I(Y_i \leq y). \tag{3.11}$$

3.2.1 Ordinary Confidence Intervals

The results on the asymptotic normality of $\hat{\mu}$ and $\hat{\theta}$ are summarized in Theorem 3.1.

Theorem 3.1 *Assume that* $0 < p_s = P(\delta_{si} = 1) \leq 1$, $W_s = P(Y \in \mathcal{P}_s) > 0$, $0 < \sigma^2 = Var(Y) < \infty$, *and that there exists an* $\alpha_0 > 0$ *such that* $E|Y|^{2+\alpha_0} < \infty$. *Then, as* $n \to \infty$,

$$\sqrt{n}\sigma_{nm}^{-1}(\hat{\mu} - \mu) \xrightarrow{d} N(0,1), \tag{3.12}$$

and

$$\sqrt{n}\sigma_{nd}^{-1}(\hat{\theta} - \theta) \xrightarrow{d} N(0,1), \tag{3.13}$$

where

$$\sigma_{nm}^2 = \sigma^2\sum_{s=1}^{S} W_s p_s^{-1} + \sigma_{2nm}^2 \tag{3.14}$$

$$\text{with } \sigma_{2nm}^2 = \sum_{s=1}^{S} W_s(1-p_s)J^{-1}r_s^{-1}\sum_{i\in s_{rs}}\left(Y_{si} - r_s^{-1}\sum_{i\in s_{rs}}Y_{si}\right)^2,$$

and

$$\sigma_{nd}^2 = \theta(1-\theta)\sum_{s=1}^{S} W_s p_s^{-1} + \sigma_{2nd}^2 \tag{3.15}$$

with $\sigma^2_{2nd} = \sum_{s=1}^{S} W_s(1-p_s)J^{-1}r_s^{-1}\sum_{i\in s_{rs}}(I(Y_{si}\le y) - r_s^{-1}\sum_{i\in s_{rs}} I(Y_{si}\le y))^2$.

Let $\widehat{W}_s = n_s/n$ and $\hat{p}_s = r_s/n_s$, $s = 1, ..., S$. We assume that the observed response rates $\hat{p}_s = r_s/n_s$, and the frequencies n_s, for imputation classes $s = 1, ..., S$ are reported in the data file. Based on Theorem 3.1, the ordinary normal approximation confidence intervals for μ and θ, with asymptotically correct coverage probability $(1-\alpha)$, are respectively given by

$$\mu \in \left(\hat{\mu} - z_{\alpha/2}\hat{\sigma}_{1m}n^{-1/2}, \hat{\mu} + z_{\alpha/2}\hat{\sigma}_{1m}n^{-1/2}\right) \tag{3.16}$$

and

$$\theta \in \left(\hat{\theta} - z_{\alpha/2}\hat{\sigma}_{1d}n^{-1/2}, \hat{\theta} + z_{\alpha/2}\hat{\sigma}_{1d}n^{-1/2}\right) \tag{3.17}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile from the standard normal distribution,

$$\hat{\sigma}^2_{1m} = \sum_{s=1}^{S} \widehat{W}_s(\hat{p}_s^{-1} + J^{-1}(1-\hat{p}_s))\,(r_s\text{-}1)^{-1}\sum_{i\in s_{rs}}(Y_{si} - \bar{Y}_{rs})^2, \tag{3.18}$$

and

$$\hat{\sigma}^2_{1d} = \sum_{s=1}^{S} \widehat{W}_s(\hat{p}_s^{-1} + J^{-1}(1-\hat{p}_s))\,(r_s\text{-}1)^{-1}\sum_{i\in s_{rs}}(I(Y_{si}\le y) - \bar{\theta}_{rs})^2, \tag{3.19}$$

where $\bar{\theta}_{rs} = r_s^{-1}\sum_{i\in s_{rs}} I(Y_{si}\le y))^2$. Note that for $J > 1$, the individual response identification flags, δ_{si}, are needed in the construction of confidence intervals.

3.2.2 Bootstrap Confidence Intervals

We now use the bootstrap method (Shao and Sitter, 1996) to approximate the asymptotic distributions of $\sqrt{n}(\hat{\mu}-\mu)$ and $\sqrt{n}(\hat{\theta}-\theta)$ under fractional imputation with imputation classes. In the procedure, the bootstrap data sets are imputed in the same way as the original data set as follows:

1. Set $b = 1$.

2. Independently within each imputation class $s = 1, ..., S$, draw simple random samples $D_s^* = \{(Y_{b,si}, \delta_{b,si}), i = 1, ..., n_s\}$ with replacement from the imputed data set $D_s = \{(\tilde{Y}_{si}, \delta_{si}), i = 1, ..., n_s\}$. Denote $D = \{D_s, s = 1, ..., S\}$ and $D^* = \{D_s^*, s = 1, ..., S\}$.

3. Within each imputation class $s = 1, ..., S$: when $\delta_{b,si} = 0$, generate $J \geq 1$ imputed values $Y_{b,sij} = \bar{Y}_{bsr} + \epsilon_{b,sij},\ j = 1, ..., J$, where $\{\epsilon_{b,sij},\ j = 1, ..., J\}$ are drawn by simple random sampling with replacement from donor residuals $\{\hat{\epsilon}_{b,si} = Y_{bsj} - \bar{Y}_{brs}, i \in s_{brs}\}$ with $\bar{Y}_{brs} = \sum_{i=1}^{n_s} \delta_{b,si} Y_{b,si} / \sum_{i=1}^{n_s} \delta_{b,si}$ and $s_{brs} = \{i : \delta_{b,si} = 1\}$.

4. Compute the imputed bootstrap estimators of μ and θ from the fractionally-imputed bootstrap data:

$$\hat{\mu}_b = \frac{1}{n} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \left\{ \delta_{b,si} Y_{b,si} + (1 - \delta_{b,si}) \frac{1}{J} \sum_{j=1}^{J} Y_{b,ij} \right\},$$

and

$$\hat{\theta}_b = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left\{\delta_{b,si}I(Y_{b,si} \leq y) + (1-\delta_{b,si})\frac{1}{J}\sum_{j=1}^{J} I(Y_{b,sij} \leq y)\right\}.$$

5. Repeat steps 2-4 for $b = 2, ..., B$ with large B.

The usual bootstrap analogues of $\hat{\mu} - \mu$ and $\hat{\theta} - \theta$ are respectively given by $\hat{\mu}_b - \hat{\mu}$ and $\hat{\theta}_b - \hat{\theta}$. We now show that in the presence of missing data, under fractional imputation, the distributions of $\sqrt{n}(\hat{\mu} - \mu)$ and $\sqrt{n}(\hat{\theta} - \theta)$ can be respectively approximated by the modified-bootstrap versions $\sqrt{n}(\hat{\mu}_b - \overline{Y}_r)$ and $\sqrt{n}(\hat{\theta} - \bar{\theta}_r)$ where $\hat{\mu}$, $\hat{\theta}$, $\overline{Y}_r$ and $\bar{\theta}_r$ are respectively given by (3.6), (3.7), (3.10) and (3.11).

Theorem 3.2 *Suppose that the conditions in Theorem 3.1 are satisfied, then as* $n \to \infty$,

$$\sup_{x \in R}\left|P_b\left\{\sqrt{n}(\hat{\mu}_b - \overline{Y}_r) \leq x\right\} - P\left\{\sqrt{n}(\hat{\mu} - \mu) \leq x\right\}\right| \xrightarrow{P} 0, \tag{3.20}$$

and

$$\sup_{x \in R}\left|P_b\left\{\sqrt{n}(\hat{\theta}_b - \hat{\theta}) \leq x\right\} - P\left\{\sqrt{n}(\hat{\theta} - \theta) \leq x\right\}\right| \xrightarrow{P} 0. \tag{3.21}$$

That is, the proposed adjustments to Shao and Sitters's (1996) statistics are given by

$$\mu_{nb} = \overline{Y}_r - \hat{\mu} \text{ and } F_{nb} = \bar{\theta}_r - \hat{\theta}. \tag{3.22}$$

Further, we have

$$\sqrt{n}\mu_{nb} \xrightarrow{d} N\left(0, \left\{J^{-1}\sum_{s=1}^{S} W_s(1-p_s)\right\}\sigma^2\right),$$

and

$$\sqrt{n}F_{nb} \xrightarrow{d} N\left(0, \left\{J^{-1}\sum_{s=1}^{S} W_s(1-p_s)\right\}\theta\left\{1-\theta\right\}\right).$$

Theorem 3.2 states that we need the adjusted bootstrap pivotals to approximate $\sqrt{n}(\hat{\mu}-\mu)$ and $\sqrt{n}(\hat{\theta}-\theta)$. We note that when J is large, $\sqrt{n}\mu_{nb} = o_p(1)$ and $\sqrt{n}F_{nb} = o_p(1)$ as $n \to \infty$ and so the usual bootstrap statistic could be used. Also, in the case of deterministic imputation, $\mu_{nb} = 0$ which means that we could use $\hat{\mu}_b - \mu$ in place of $\hat{\mu}_b - \overline{Y}_r$. However, deterministic imputation leads to an inconsistent estimator of the distribution function of Y.

We form adjusted bootstrap percentile confidence intervals on μ as follows. We repeat the bootstrap process independently B times to obtain $\hat{\mu}_1, ..., \hat{\mu}_B$ and select the $100(1-\alpha/2)$ and $100(\alpha/2)$ sample quantiles of $\{\hat{\mu}_b, 1 \le b \le B\}$. The $(1-\alpha)$-level bootstrap percentile confidence interval on μ is given by

$$(\hat{\mu} - (\hat{\mu}_{b,1-\alpha/2} - \bar{Y}_r), \hat{\mu} - (\hat{\mu}_{b,\alpha/2} - \bar{Y}_r)).$$

The $(1-\alpha)$-level adjusted bootstrap interval on θ can be constructed similarly.

3.3 Empirical Likelihood Confidence Intervals

3.3.1 Ordinary Confidence Intervals

We extend the method presented in Chapter 2 to the case with imputation classes and form empirical likelihood ratios for μ and θ as follows. Let

$$Z_{si,m}(\mu) = \delta_{si} Y_{si} + (1-\delta_{si}) J^{-1} \sum_{j=1}^{J} Y_{sij} - \mu,$$

and

$$Z_{si,d}(\theta) = \delta_{si} I(Y_{si} \leq y) + (1-\delta_{si}) J^{-1} \sum_{j=1}^{J} I(Y_{sij} \leq y) - \theta,$$

where $1 \leq s \leq S$, $1 \leq i \leq n_s$. The empirical log-likelihood ratio for μ is given by

$$l_{n,m}(\mu) = -2 \sum_{s=1}^{S} \sum_{i=1}^{n_s} \log(n p_{si,m}), \tag{3.23}$$

where $\{p_{si,m}, 1 \leq s \leq S, 1 \leq i \leq n_s\}$ maximize the log-EL function

$$l_m(p) = \sum_{s=1}^{S} \sum_{i=1}^{n_s} \log(n p_{si,m}),$$

subject to the following constraints: $p_{si,m} > 0$, $\sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si,m} Z_{si,m}(\mu) = 0$, and $\sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si,m} = 1$.

Similarly for θ,

$$l_{n,d}(\theta) = -2 \sum_{s=1}^{S} \sum_{i=1}^{n_s} \log(n p_{si,d}), \tag{3.24}$$

where $\{p_{si,d}, 1 \leq s \leq S, 1 \leq i \leq n_s\}$ maximize the log-EL function

$$l_d(p) = \sum_{s=1}^{S} \sum_{i=1}^{n_s} \log(np_{si,d}),$$

subject to the following constraints: $p_{si,d} > 0$, $\sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si,d} Z_{si,d}(\theta) = 0$, and $\sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si,d} = 1$.

It can be shown, using the Lagrange multiplier method, that

$$l_{n,m}(\mu) = 2 \sum_{s=1}^{S} \sum_{i=1}^{n_s} \log\left\{1 + \lambda_{n,m} Z_{si,m}(\mu)\right\},$$

and

$$l_{n,d}(\theta) = 2 \sum_{s=1}^{S} \sum_{i=1}^{n_s} \log\left\{1 + \lambda_{n,d} Z_{si,d}(\theta)\right\},$$

where $\lambda_{n,m}$ and $\lambda_{n,d}$ are solutions to equations:

$$\frac{1}{n} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \frac{Z_{si,m}(\mu)}{1 + \lambda_{n,m} Z_{si,m}(\mu)} = 0 \text{ and } \frac{1}{n} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \frac{Z_{si,d}(\theta)}{1 + \lambda_{n,d} Z_{si,d}(\theta)} = 0$$

The results on the asymptotic distributions of $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$ are stated in Theorem 3.3.

Theorem 3.3 *Suppose that the conditions in Theorem 3.1 are satisfied. Then as* $n \to \infty$,

$$l_{n,m}(\mu) \xrightarrow{d} c_m \chi_1^2, \tag{3.25}$$

and

$$l_{n,d}(\theta) \xrightarrow{d} c_d \chi_1^2, \tag{3.26}$$

where $c_m = \sigma^2_{nm}/\sigma^2_{2m}$, $c_d = \sigma^2_{nd}/\sigma^2_{2d}$,

$$\sigma^2_{2m} = \sigma^2 \sum_s W_s \left((1 - p_s) J^{-1} + p_s \right), \tag{3.27}$$

$$\sigma^2_{2d} = \theta(1-\theta) \sum_s W_s \left((1 - p_s) J^{-1} + p_s \right), \tag{3.28}$$

and σ^2_{nm} *and* σ^2_{nd} *are respectively defined by equations (3.14) and (3.15) of Theorem 3.1.*

Using Theorem 3.3, a (1-α)- level confidence interval on μ ,with asymptotically correct coverage probability, can be constructed as

$$\left\{ \mu | \left(\hat{\sigma}^2_2/\hat{\sigma}^2_{1m} \right) l_{n,m}(\mu) \leq \chi^2_\alpha(1) \right\}, \tag{3.29}$$

where $\chi^2_\alpha(1)$ is the upper α quantile of the χ^2 distribution with one degree of freedom, $\hat{\sigma}^2_{1m}$ is given by (3.18) and

$$\hat{\sigma}^2_2 = \sum_s \widehat{W}_s \left((1 - \hat{p}_s) J^{-1} + \hat{p}_s \right) (r_s - 1)^{-1} \sum_{i \in s_{rs}} (Y_{si} - \bar{Y}_{rs})^2, \tag{3.30}$$

with $\widehat{W}_s = n_s/n$, $\hat{p}_s = r_s/n_s$ and $\bar{Y}_{rs} = {r_s}^{-1} \sum_{i \in s_{rs}} Y_{si}$. The (1-$\alpha$)- level confidence interval on θ with asymptotically correct coverage probability can be constructed similarly.

3.3.2 Bootstrap Calibrated Confidence Intervals

We now approximate the asymptotic distributions of $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$ using the bootstrap sample data. Let

$$\begin{aligned} Z_{b,si,m}(\hat{\mu}) &= \delta_{b,si}Y_{b,si} + (1-\delta_{b,si})J^{-1}\sum_{j=1}^{J} Y_{b,sij} - \overline{Y}_r, &(3.31)\\ \text{and } Z_{b,si,d}(\hat{\theta}) &= \delta_{b,si}I(Y_{b,si} \leq y) + (1-\delta_{b,si})J^{-1}\sum_{j=1}^{J} I(Y_{b,sij} \leq y) - \bar{\theta}_r, &(3.32) \end{aligned}$$

for $s = 1, ..., S$, $i = 1, ..., n_s$. Then the proposed adjusted bootstrap analog of $l_{n,m}(\mu)$ is given by

$$l_{b,n,m}(\hat{\mu}) = -2\sum_{s=1}^{S}\sum_{i=1}^{n_s}\log(np_{si,m}),$$

where $\{p_{si,m}, 1 \leq s \leq S, 1 \leq i \leq n_s\}$ maximize $\sum_{s=1}^{S}\sum_{i=1}^{n_s}\log(np_{si,m})$ subject to the following constraints: $p_{si,m} > 0$, $\sum_{s=1}^{S}\sum_{i=1}^{n_s} p_{si,m}Z_{b,si,m}(\mu) = 0$ and $\sum_{s=1}^{S}\sum_{i=1}^{n_s} p_{si,m} = 1$.

Similarly, the proposed bootstrap analog of $l_{n,d}(\theta)$ is

$$l_{b,n,d}(\hat{\theta}) = -2\sum_{s=1}^{S}\sum_{i=1}^{n_s}\log(np_{si,d}),$$

where $\{p_{si,d}, 1 \leq s \leq S, 1 \leq i \leq n_s\}$ maximize $\sum_{s=1}^{S}\sum_{i=1}^{n_s}\log(np_{si,d})$ subject to the constraints: $p_{si,d} > 0$, $\sum_{s=1}^{S}\sum_{i=1}^{n_s} p_{si,d}Z_{b,si,d}(\theta) = 0$, and $\sum_{s=1}^{S}\sum_{i=1}^{n_s} p_{si,d} = 1$.

It can be shown, using the Lagrange multiplier method, that

$$l_{b,n,m}(\hat{\mu}) = 2\sum_{s=1}^{S}\sum_{i=1}^{n_s}\log\left\{1+\lambda_{b,n,m}Z_{b,si,m}(\hat{\mu})\right\},$$

and

$$l_{b,n,d}(\hat{\theta}) = 2\sum_{s=1}^{S}\sum_{i=1}^{n_s}\log\left\{1+\lambda_{b,n,d}Z_{b,si,d}(\hat{\theta})\right\},$$

where $\lambda_{b,n,m}$ and $\lambda_{b,n,d}$ are solutions to equations:

$$\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\frac{Z_{b,si,m}(\hat{\mu})}{1+\lambda_{b,n,m}Z_{b,si,m}(\hat{\mu})} = 0 \text{ and } \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\frac{Z_{bsi,d}(\hat{\theta})}{1+\lambda_{b,n,d}Z_{b,si,d}(\hat{\theta})} = 0.$$

Theorem 3.2 below states that $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$ can be approximated by their adjusted bootstrap analogs $l_{b,n,m}(\hat{\mu})$ and $l_{b,n,d}(\hat{\theta})$.

Theorem 3.4 *Suppose that the conditions in Theorem 3.1 are satisfied. Then as* $n \to \infty$,

$$\sup_{x\in R}\left|P_b\left\{l_{b,n,m}(\hat{\mu}) \le x\right\} - P\left\{l_{n,m}(\mu) \le x\right\}\right| \xrightarrow{P} 0, \tag{3.33}$$

and

$$\sup_{x\in R}\left|P_b\left\{l_{b,n,d}(\hat{\theta}) \le x\right\} - P\left\{l_{n,d}(\theta) \le x\right\}\right| \xrightarrow{P} 0. \tag{3.34}$$

This shows that the ordinary bootstrap EL statistics cannot be used to approximate the distributions of $l_{n,m}(\mu)$ and $l_{n,d}(\theta)$ unless $J \to \infty$.

The adjusted bootstrap calibrated empirical likelihood confidence interval on μ and θ can be constructed as follows. We repeat the bootstrap process independently B times and obtain $l^1_{b,n,m}(\hat{\mu}), ..., l^B_{b,n,m}(\hat{\mu})$ and $l^1_{b,n,d}(\hat{\theta}), ..., l^B_{b,n,d}(\hat{\theta})$. Let $l_{1-\alpha,m}$

and $l_{1-\alpha,d}$ be respectively the $100(1-\alpha)\%$ sample quantiles of $\left\{l^k_{b,n,m}(\hat{\mu}), 1 \leq k \leq B\right\}$ and $\left\{l^k_{b,n,d}(\hat{\theta}), 1 \leq k \leq B\right\}$. The $(1-\alpha)$-level adjusted bootstrap EL intervals on μ and θ are respectively given by

$$\left\{\mu \middle| \, l_{n,m}(\mu) \leq l_{1-\alpha,m}\right\}, \text{ and } \left\{\theta \middle| \, l_{n,d}(\theta) \leq l_{1-\alpha,d}\right\}. \tag{3.35}$$

3.4 Simulation Study

A simulation was conducted to study the performance of bootstrap confidence intervals on the population mean μ and the distribution function θ, for fixed y, based on fractional-imputed data with imputation classes. In particular, we compared the performance of the proposed adjusted bootstrap 95% confidence intervals versus their ordinary (unadjusted) counterparts based on two methods: the bootstrap percentile (BP) and the empirical likelihood (EL). The confidence intervals were examined in terms of their coverage probabilities and their average lengths. In our simulations, the precision of comparisons among the same test procedures at different settings was achieved by re-using the values of input random numbers, in the sense that the results were correlated by having common observations for each of the simulation runs.

The results were based on 2000 simulations on data imputed using random imputation (fractional imputation with $J = 1$) utilizing $B = 3000$ bootstrap repetitions. The standard errors for simulated coverage of the 95% confidence intervals were approximately 0.010 with 2000 simulation runs. We considered three imputation classes, $S = 3$, and total sample size $n = n_1 + n_2 + n_3 = 300$, as well as five different cases of class response probabilities p_s, $s = 1, 2, 3$ as shown in Table 3.1. In the first scenario we considered full response in each class, we can describe scenario 2 as high response, scenarios 3 and 4 as medium response since both share same set of probabilities assigned to different classes, and scenario 5 as low response with 50% chance of response in each class. Note that under full response, the proposed adjustments cancel out.

Table 3.1: Class response probabilities considered in simulation scenarios.

scenario	p_1	p_2	p_3
1	1	1	1
2	0.8	0.7	0.6
3	0.6	0.7	0.5
4	0.5	0.6	0.7
5	0.5	0.5	0.5

3.4.1 Data Frame

The data was generated based on the simulation setup presented in Section 5 of Fang et al. (2009). We considered a total sample size of 300. That is, for the three classes, the sum of random class sample sizes: $n = n_1 + n_2 + n_3 = 300$. For each simulation, we considered three imputation classes ($S = 3$), and generated n values of Y from gamma distribution with shape parameter 43 and scale parameter 0.20. The sample data was divided into imputation classes according to the value of an auxiliary variable $X \in \{1, 2, 3\}$ which was generated by the proportional odds model

$$\log \frac{P(X \leq j|Y = y)}{P(X > j|Y = y)} = j + \beta y, \tag{3.36}$$

with $j = 1, 2$ and $\beta = -0.4$. In particular, for $U_k \sim$uniform(0,1) and $P_{kj} = P(X \leq j|Y = y_k) = \exp(j + \beta y_k)/\left(1 + \exp(j + \beta y_k)\right)$, $k = 1, ..., n$ and $j = 1, 2$, we assigned classes to observations $k = 1, ..., n$ in the data file according to the following pseudo code : `If` $U_k \leq P_{k1}$ `then` $class = 1$; `Else if` $U_k \leq P_{k2}$ `then` $class = 2$; `Else` $class = 3$.

Note that the class sample sizes n_1, n_2 and n_3 were different for each simulation run. We assumed that Y_{si} is MCAR within each class, that is $P(\delta_{si} = 1|Y_{si}) = P(\delta_{si} = 1) = p_s$, $0 < p_s \leq 1$ where $\delta_{si} = 1$ if Y_{si} is observed and $\delta_{si} = 0$ if Y_{si} is missing, $s = 1, 2, 3$, $i = 1, ..., n_s$. Response flags δ_{si} were generated, independently within each class $s = 1, 2, 3$, from three Bernoulli distributions with corresponding success probabilities p_s. We considered five different combinations of class response probabilities $p = (p_1, p_2, p_3)$ in the simulation scenarios as shown in Table 3.1.

3.4.2 Simulations

The ordinary versions of confidence intervals were obtained by ignoring the proposed adjustments μ_{nb} and F_{nb} defined by (3.22). In particular, the 95% BP confidence interval for μ for the adjusted method was $\left(\hat{\mu} - P_{0.975},\ \hat{\mu} - P_{0.025}\right)$, where P_α was the 100α percentile of the sampling distribution of $\hat{\mu}_b - \bar{Y}_r$; while for the ordinary method, P_α was based on the sampling distribution of $\hat{\mu}_b - \hat{\mu}$. Similarly, for θ. Note that the lengths of the BP confidence intervals were the same under the ordinary and adjusted methods as the proposed adjustment, present in the upper and lower bounds, cancelled each other out in the calculation of the interval length.

The bisection method proposed by Wu (2005) was used to obtain $\lambda_{n,m,b}$ ($\lambda_{n,d,b}$)

and find lower and upper bounds of the (1-α)% empirical likelihood confidence intervals using bootstrap sample percentiles of $\lambda_{n,m,b}$ ($\lambda_{n,d,b}$) as cut-off values of the χ_1^2 distribution. Note that the ordinary analogs of $Z_{b,i,m}(\hat{\mu})$ and $Z_{b,i,d}(\hat{\theta})$ (3.31) after setting the adjustments (3.22) to zero are

$$\tilde{Z}_{b,si,m}(\hat{\mu}) = \delta_{b,si}Y_{b,si} + (1-\delta_{b,si})J^{-1}\sum_{j=1}^{J} Y_{b,sij} - \hat{\mu}, \quad (3.37)$$

$$\text{and } \tilde{Z}_{b,si,d}(\hat{\theta}) = \delta_{b,si}I(Y_{b,si} \leq y) + (1-\delta_{b,si})J^{-1}\sum_{j=1}^{J} I(Y_{b,sij} \leq y) - \hat{\theta}. \quad (3.38)$$

3.4.3 Results

Table 3.2 displays the coverage probabilities and average lengths of the 95% confidence intervals for the population mean, $\mu = E(Y) = 8.6$ and Tables 3.3- 3.5 display the coverage probabilities and average lengths of the 95% confidence intervals for the distribution functions: $F1 = F(7.68) = 0.25$, $F2 = F(8.53) = 0.5$ and $F3 = F(9.45) = 0.75$ respectively.

Under full response, the coverages and lengths of the BP and bootstrap EL intervals for the population mean were similarly very good. Generally, in the presence of missing data, the adjusted BP confidence intervals for μ led to very good coverage close to nominal 95%. Compared to the ordinary methods, the adjusted methods resulted in smaller departures from the nominal level for all simulation cases. For all the missing data cases, the ordinary EL led to overcoverage while the ordinary BP produced significant undercoverage. For example, scenario 5 with the lowest class response probabilities, generated coverage of 88% for the ordinary BP versus 97% for the ordinary EL.

In terms of the average lengths of the confidence intervals for μ, the adjusted EL intervals were slightly longer relative to the corresponding BP intervals with percent change of less than 1%. The adjusted EL generated shorter confidence intervals, for example, under scenario 4, the ordinary and adjusted EL average interval lengths were respectively 0.476 and 0.441 (compared to 0.437 for BP). As we dealt with increasing non-response under scenarios 1 to 5, the resulting confidence intervals were longer for all methods.

For the distribution functions, we observed that both adjusted methods resulted in coverage probabilities that were very close to, or above, the nominal value of 95%. The ordinary EL method led to coverage probabilities that were greater than 95%, while the ordinary BP method had low coverage. In terms of the average lengths of confidence intervals for the distribution functions, the adjusted EL method performed similarly compared to the BP, and better than the ordinary EL.

Table 3.2: Bootstrap confidence interval coverage probability and average interval length for the mean μ under random imputation with imputation classes for different class response probability scenarios.

	Coverage (%)				Average Length			
sc.	OrdBP	AdjBP	OrdEL	AdjEL	OrdBP	AdjBP	OrdEL	AdjEL
1	95.2	95.2	95.4	95.4	0.294	0.294	0.296	0.296
2	89.2	94.8	96.2	94.9	0.379	0.379	0.408	0.382
3	88.1	94.7	96.6	94.9	0.410	0.410	0.442	0.413
4	89.5	95.3	96.9	95.4	0.437	0.437	0.476	0.441
5	87.9	95.0	97.1	95.3	0.464	0.464	0.507	0.469

Table 3.3: Bootstrap confidence interval coverage probability and average interval length for distribution function $F1$ under random imputation with imputation classes for different class response probability scenarios.

	Coverage (%)				Average Length			
sc.	OrdBP	AdjBP	OrdEL	AdjEL	OrdBP	AdjBP	OrdEL	AdjEL
1	93.8	93.8	95.5	95.5	0.097	0.097	0.097	0.097
2	87.8	94.9	96.5	95.4	0.125	0.125	0.132	0.126
3	88.5	94.6	96.3	95.9	0.137	0.137	0.146	0.143
4	86.2	93.9	96.1	94.8	0.145	0.145	0.156	0.147
5	86.4	94.5	96.8	95.1	0.154	0.154	0.165	0.154

Table 3.4: Bootstrap confidence interval coverage probability and average interval length for distribution function $F2$ under random imputation with imputation classes fordifferent class response probability scenarios.

	Coverage (%)				Average Length			
sc.	OrdBP	AdjBP	OrdEL	AdjEL	OrdBP	AdjBP	OrdEL	AdjEL
1	95.8	94.8	94.1	94.1	0.112	0.112	0.112	0.112
2	87.1	94.4	96.5	94.6	0.145	0.145	0.154	0.146
3	87.0	93.9	96.5	95.1	0.157	0.157	0.167	0.164
4	87.3	94.1	96.8	94.7	0.167	0.167	0.179	0.169
5	87.4	95.2	97.5	95.6	0.178	0.178	0.191	0.177

Table 3.5: Bootstrap confidence interval coverage probability and average interval length for distribution function $F3$ under random imputation with imputation classes for different class response probability scenarios.

	Coverage (%)				Average Length			
sc.	OrdBP	AdjBP	OrdEL	AdjEL	OrdBP	AdjBP	OrdEL	AdjEL
1	95.8	94.8	94.8	94.8	0.097	0.097	0.097	0.097
2	89.3	94.9	97.1	95.9	0.126	0.126	0.134	0.127
3	88.6	96.1	97.6	97.1	0.135	0.135	0.144	0.140
4	88.2	95.9	97.6	96.4	0.144	0.144	0.154	0.145
5	87.7	93.9	96.8	95.3	0.154	0.154	0.165	0.154

3.5 Conclusions

We proposed asymptotically correct normal approximation and empirical likelihood confidence intervals on the mean $\mu = E(Y)$ and the distribution function $F(y) = P(Y \leq y)$, $y \in R$, under fractional imputation with imputation classes. We constructed adjusted confidence intervals based on the bootstrap data obtained by imitating the process of imputing the original data set in bootstrap resampling. Our simulation study demonstrated that, in terms of coverage probabilities for the mean and the distribution function, the ordinary EL method performed better than the ordinary BP method, and that the proposed adjustments to the confidence intervals have brought the coverage closer to the nominal level.

3.6 Appendix: Proofs

The following proofs are based on the theorems and results stated in the appendix to Chapter 2 (Section 2.6.1).

3.6.1 Proof of Theorem 3.1

Let P^* denote the probability with respect to the randomness in the imputation process, and similarly for E^* and Var^*.

The proof of Theorem 3.1 follows closely the proof of Theorem 2.1 in Chapter 2. We start by forming the following decomposition

$$\hat{\mu} - \mu = (\hat{\mu} - E^*[\hat{\mu}]) + (E^*[\hat{\mu}] - \mu)\,.$$

Define $V_n = \sqrt{n}\,(E^*[\hat{\mu}] - \mu)$ and $U_n = \sqrt{n}\,(\hat{\mu} - E^*[\hat{\mu}])$ so that $\sqrt{n}(\hat{\mu} - \mu) = U_n + V_n$.

Let us now show that the conditions of Chen and Rao Theorem, stated in the appendix of Chapter 2, are met. We have $E^*[Y_{sij}] = \frac{1}{r_s}\sum_{i=1}^{n_s} \delta_{si} Y_{si} = \bar{Y}_{rs}$ and define $\bar{Y}_r = \frac{1}{n}\sum_{s=1}^{S} n_s \bar{Y}_{rs}$. Then, as shown in Section 3.2, $E^*[\hat{\mu}] = \bar{Y}_r$. Let $B_n = \{(\delta_{si}, Y_{si}), i = 1, 2, ..., n_s, s = 1, ..., S\}$ so that V_n is B_n-measurable. We have

$$
\begin{aligned}
V_n &= \sqrt{n}\left(E^*[\hat{\mu}] - \mu\right) = \sqrt{n}\left(\bar{Y}_r - \mu\right) \\
&= \frac{1}{\sqrt{n}}\left(\sum_{s=1}^{S}\frac{n_s}{r_s}\sum_{i=1}^{n_s}\delta_{si}Y_{si} - \sum_{s=1}^{S}\frac{n_s}{r_s}\sum_{i=1}^{n_s}\delta_{si}\mu\right) \\
&= \sum_{s=1}^{S}\sqrt{\frac{n_s}{n}}\frac{n_s}{r_s}\frac{1}{\sqrt{n_s}}\sum_{i=1}^{n_s}\delta_{si}(Y_{si}-\mu) \\
&= \sum_{s=1}^{S}\left(W_s^{1/2} + o_p(1)\right)\left(p_s^{-1} + o_p(1)\right)\frac{1}{\sqrt{n_s}}\sum_{i=1}^{n_s}\delta_{si}(Y_{si}-\mu),
\end{aligned}
$$

after noting that $\frac{n_s}{n} = W_s + o_p(1)$ and $\frac{n_s}{r_s} = p_s^{-1} + o_p(1)$, where $W_s = P(Y \in P_s)$, $p_{si} = P(\delta_{si} = 1)$.

The random variables $\{\delta_{si}(Y_{si} - \mu),\ i = 1, ..., n_s,\ s = 1, ..., S\}$ are i.i.d. within classes and, by the MCAR assumption,

$$
E\left[\delta_{si}(Y_{si} - \mu)\right] = 0,
$$

and

$$
\begin{aligned}
Var\left[\delta_{si}(Y_{si} - \mu)\right] &= E\left[\delta_{si}(Y_{si} - \mu)\right]^2 - E^2\left[\delta_{si}(Y_{si} - \mu)\right] \\
&= E\left[\delta_{si}Y_{si}\right]^2 - 2E[\delta_{si}Y_{si}\mu] + E\left[\delta_{si}\mu\right]^2 \\
&= p_s E\left[Y_{si}\right]^2 - 2p_s\mu^2 + p_s\mu^2 \\
&= p_s(\sigma^2 + \mu^2) - p_s\mu^2 \\
&= p_s\sigma^2.
\end{aligned}
$$

Therefore, $E[V_n] = 0$ and $Var(V_n) \approx \sigma^2 \sum_{s=1}^{S} W_s p_s^{-1} = \sigma_{1nm}^2$. By the Central Limit Theorem, conditional on $\{n_s, s = 1, .., S\}$, we have $\sqrt{n}\sigma_{1nm}^{-1}\left(\bar{Y}_r - \mu\right) \xrightarrow{d} N(0,1)$ which verifies the first condition of Chen and Rao Theorem.

Note that $U_n = \sqrt{n}\left(\hat{\mu} - E^*[\hat{\mu}]\right) = \dfrac{1}{\sqrt{n}} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \left\{(1-\delta_{si})\frac{1}{J}\sum_{j=1}^{J}\left(Y_{sij} - \bar{Y}_{rs}\right)\right\}$. It can be easily seen that $E^*[U_n|B_n] = 0$ and

$$\begin{aligned}
Var^*[U_n|B_n] &= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left\{(1-\delta_{si})Var^*(\frac{1}{J}\sum_{j=1}^{J}Y_{sij} - \bar{Y}_{rs})\right\} \\
&= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left\{(1-\delta_i)\frac{1}{J}Var^*(Y_{sij})\right\} \\
&= \frac{1}{n}\frac{1}{J}\sum_{s=1}^{S}(n_s - r_s)\left(E^*(Y_{sij} - \bar{Y}_{rs})^2\right) \\
&= \sum_{s=1}^{S}\frac{n_s - r_s}{nJ}\frac{1}{r_s}\sum_{i\in r_s}(Y_{si} - \bar{Y}_{rs})^2.
\end{aligned}$$

Let us denote $\sigma_{2nm}^2 = \sum_{s=1}^{S} W_s(1-p_s)\frac{1}{J}\dfrac{1}{r_s}\sum_{i\in r_s}(Y_{si} - \bar{Y}_{rs})^2$. By the Berry-Esseen Theorem,

$$\sup_x \left| P\left(\frac{U_n}{\sigma_{2nm}} \le x\right) - \Phi(x)\right| \le \frac{c\rho_n}{(\sigma_{2nm}^2)^{3/2}}, \tag{3.39}$$

where

$$\rho_n = n^{-3/2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{si})E^*\left|\frac{1}{J}\sum_{j=1}^{J}Y_{ijs} - \bar{Y}_{rs}\right|^3.$$

We note that, by the C_r-inequality (2.33),

$$
\begin{aligned}
\rho_n &= n^{-3/2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{si})E^*\left|\frac{1}{J}\sum_{j=1}^{J}Y_{sij}-\bar{Y}_{rs}\right|^3 \\
&\leq Cn^{-3/2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{si})\left(E^*\left|\frac{1}{J}\sum_{j=1}^{J}Y_{sij}\right|^3+\left|\bar{Y}_{rs}\right|^3\right) \\
&\leq Cn^{-3/2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{si})\left(\frac{1}{J^2}E^*\left|Y_{sij}\right|^3+\left|E^*\left(Y_{sij}\right)\right|^3\right) \\
&\leq Cn^{-3/2}(n-r)\leq Cn^{-1/2}.
\end{aligned}
$$

Therefore, $c\rho_n(\sigma_{2nm})^{-3}\leq cn^{-1/2}$ and so

$$
\sup_x\left|P^*\left(\frac{U_n}{\sigma_{2n}}\leq x|B_n\right)-\Phi(x)\right|=o_p(1), \tag{3.40}
$$

where $B_n=\{(\delta_{si},Y_{si}),i=1,2,...,n_s,s=1,...,S\}$. This result is also true unconditionally, by applying the Lebesque's dominated convergence theorem (2.9).

Hence, as $n\to\infty$,

$$
\frac{U_n+V_n}{\sqrt{\sigma_{1nm}^2+\sigma_{2nm}^2}}\xrightarrow{d}N(0,1). \tag{3.41}
$$

Recall that $\sqrt{n}(\hat{\mu}-\mu)=U_n+V_n$ or

$$
\frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{\sigma_{1nm}^2+\sigma_{2nm}^2}}\xrightarrow{d}N(0,1). \tag{3.42}
$$

Applying Slutsky Theorem to the denominator, we obtain,

$$
\sqrt{n}\sigma_{nm}^{-1}(\hat{\mu}-\mu)\xrightarrow{d}N\left(0,1\right),
$$

as $n \to \infty$, with $\sigma^2_{nm} = \sigma^2_{1nm} + \sigma^2_{2nm}$.

The asymptotic normality of $\hat{\theta}$ can be proved similarly. ■

3.6.2 Proof of Theorem 3.2

Let E_b denote expectation taken with respect to the resampling distribution and conditional on $\mathcal{D}$. Similarly, we denote probability P_b.

To prove that equation (3.20) of Theorem 3.2 holds, first, we will decompose $\hat{\mu}_b - \overline{Y}_r$ into a sum $U_n + V_n + R_{n1}$. We will verify the conditions of Chen and Rao Theorem about the limiting distributions of U_n and V_n using Berry-Esseen Theorem. We will then use Chen and Rao Theorem and Theorem 3.1 to show (3.20).

Let

$$r_{b,s} = \sum_{i=1}^{n_s} \delta_{b,si}$$

represent the number of respondents in the bootstrap sample $\mathcal{D}^* = \{(Y_{b,si}, \delta_{b,si}), b = 1, .., B, i = 1, ..., n_s, s = 1, ..., S\}$, and let

$$\mu_{s1n} = \frac{1}{n_s} \sum_{i=1}^{n_s} \delta_{si} Y_{si} \text{ and } \mu_{s2n} = \frac{1}{n_s} \sum_{i=1}^{n_s} \delta_{si}$$

be respectively the average response, and average number of respondents, in imputation class $s = 1, ..., S$ of the original sample $\mathcal{D} = \{(Y_{si}, \delta_{si}), i = 1, ..., n_s\}$. Note that $\mu_{s1n} = E_b [\delta_{b,si} Y_{b,si}]$ and $\mu_{s2n} = E_b [\delta_{b,si}]$ and $\mu_{s1n}/\mu_{s2n} = \bar{Y}_{rs}$.

We begin by forming the following decomposition

$$\hat{\mu}_b - \overline{Y}_r = (\hat{\mu}_b - E_b^* [\hat{\mu}_b]) + \left(E_b^* [\hat{\mu}_b] - \overline{Y}_r\right),$$

where E_b^* denotes expectation taken with respect to the imputation process for the bootstrap sample.

Let $\bar{Y}_{b,sr} = \sum \delta_{b,si} Y_{b,si} / \sum \delta_{b,si}$ denote the average response in a bootstrap sample $b = 1, ..., B$. Using Taylor expansion, we obtain

$$\begin{aligned}
E_b^*\left[\hat{\mu}_b\right] &= E_b^*\left[\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left\{\delta_{b,si}Y_{b,si} + (1-\delta_{b,si})\frac{1}{J}\sum_{j=1}^{J}Y_{b,sij}\right\}\right] \\
&= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\delta_{b,si}Y_{b,si} + \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{b,si})\frac{1}{J}\sum_{j=1}^{J}E_b^*\left[Y_{b,sij}\right] \\
&= \frac{1}{n}\sum_{s=1}^{S} r_{b,s}\bar{Y}_{b,sr} + \frac{1}{n}\sum_{s=1}^{S}\left(n_s - r_{b,s}\right)\bar{Y}_{b,sr} = \frac{1}{n}\sum_{s=1}^{S} n_s \frac{\sum \delta_{b,si}Y_{b,si}}{\sum \delta_{b,si}} \\
&= \frac{1}{n}\sum_{s=1}^{S} n_s \left(\frac{E_b\left[\delta_{b,si}Y_{b,si}\right]}{E_b\left[\delta_{b,si}\right]} - \frac{E_b\left[\delta_{b,si}Y_{b,si}\right]}{E_b^2\left[\delta_{b,si}\right]}\frac{1}{n_s}\sum_{i=1}^{n_s}\left(\delta_{b,si} - E_b\left[\delta_{b,si}\right]\right) + \right. \\
&\quad \left. + \frac{1}{E_b\left[\delta_{b,i}\right]}\frac{1}{n_s}\sum_{i=1}^{n_s}\left(\delta_{b,i}Y_{b,i} - E_b\left[\delta_{b,i}Y_{b,i}\right]\right) + R_{n1}\right) \\
&= \frac{1}{n}\sum_{s=1}^{S} n_s \left(\overline{Y}_{rs} \text{-} \frac{1}{r_s}\sum_{i=1}^{n_s}\left(\delta_{b,si}\left(Y_{b,si} - \frac{\mu_{s1n}}{\mu_{s2n}}\right) - E_b\left(\delta_{b,si}\left(Y_{b,si} - \frac{\mu_{s1n}}{\mu_{s2n}}\right)\right)\right) + R_{n1}\right) \\
&= \frac{1}{n}\sum_{s=1}^{S}\frac{1}{\mu_{s2n}}\sum_{i=1}^{n_s}\left(\delta_{b,si}\left(Y_{b,si} - \overline{Y}_{rs}\right) - E_b\left(\delta_{b,si}\left(Y_{b,si} - \overline{Y}_{rs}\right)\right)\right) + \bar{Y}_r + R_{n1},
\end{aligned}$$

where $P_b\left[\sqrt{n}|R_{n1}| > \varepsilon\right] \to 0$ a.s. in [P] for any $\varepsilon > 0$.

Let

$$V_n = \frac{1}{n}\sum_{s=1}^{S}\frac{1}{\mu_{s2n}}\sum_{i=1}^{n_s}\left(\delta_{b,si}\left(Y_{b,si} - \overline{Y}_{rs}\right) - E_b\left(\delta_{b,si}\left(Y_{b,si} - \overline{Y}_{rs}\right)\right)\right),$$

and $U_n = (\hat{\mu}_b - E_b^*[\hat{\mu}_b])$. Note that $V_n + R_{n1} = E_b^*[\hat{\mu}_b] - \bar{Y}_r$. Hence, we have

$$\hat{\mu}_b - \overline{Y}_r = (\hat{\mu}_b - E_b^*[\hat{\mu}_b]) + \left(E_b^*[\hat{\mu}_b] - \overline{Y}_r\right) = U_n + V_n + R_{n1}. \tag{3.43}$$

Next, we will investigate the limiting distribution of $\sqrt{n}V_n$. By the MCAR assumption and the law of large numbers

$$\begin{aligned}
Var_b(\sqrt{n}V_n) &= nVar_b(\frac{1}{n}\sum_{s=1}^{S}\frac{1}{\mu_{s2n}}\sum_{i=1}^{n_s}\left(\delta_{b,si}\left(Y_{b,si}-\overline{Y}_{rs}\right)\right)) \\
&= \frac{1}{n}\sum_{s=1}^{S}\frac{1}{\mu_{s2n}^2}\sum_{i=1}^{n_s}Var_b\left(\delta_{b,si}\left(Y_{b,si}-\overline{Y}_{rs}\right)\right) \\
&= \frac{1}{n}\sum_{s=1}^{S}\frac{1}{\mu_{s2n}^2}n_s\left(E_b\left[\delta_{b,si}\left(Y_{b,si}-\overline{Y}_{rs}\right)\right]^2 - E_b^2\left[\delta_{b,si}\left(Y_{b,si}-\overline{Y}_{rs}\right)\right]\right) \\
&= \frac{1}{n}\sum_{s=1}^{S}\frac{1}{\mu_{s2n}^2}n_s\left(E_b\left[\delta_{b,si}Y_{b,si}\right]^2 - 2E_b[\delta_{b,si}Y_{b,si}\overline{Y}_{rs}] + E_b\left[\delta_{b,si}\overline{Y}_{rs}\right]^2\right) \\
&= \frac{1}{n}\sum_{s=1}^{S}\frac{1}{\mu_{s2n}^2}n_s\left(\mu_{s2n}E_b\left[Y_{b,si}\right]^2 - \mu_{s2n}\overline{Y}_{rs}^2\right) = \frac{1}{n}\sum_{s=1}^{S}\frac{n_s}{\mu_{s2n}}\left(Var_b\left[Y_{b,si}\right]\right).
\end{aligned}$$

Noting that $\dfrac{1}{\mu_{s2n}} = p_s^{-1} + o_p(1)$ and $\dfrac{n_s}{n} = W_s + o_p(1)$, we have

$$\sigma_{a1n}^2 := Var_b(\sqrt{n}V_n) = \sigma^2\sum_{s=1}^{S}W_s p_s^{-1} + o_p(1).$$

By the Berry-Esseen Theorem, we have

$$\sup_{x\in R}\left|P_b\left(\frac{\sqrt{n}V_n}{\sigma_{a1n}} \le x\right) - \Phi(x)\right| \le \frac{c\rho_n}{(\sigma_{a1n}^2)^{2/3}n^{1/2}},$$

where, by the C_r- inequality again,

$$
\begin{aligned}
\rho_n n^{-1/2} &= n^{-1/2} \sum_{s=1}^{S} \sum_{i=1}^{n_s} E_b |\mu_{s2n}^{-1} n^{-1} \left(\delta_{b,si} \left(Y_{b,si} - \overline{Y}_{rs} \right) - E_b \left[\delta_{b,si} \left(Y_{b,si} - \overline{Y}_{rs} \right) \right] \right) |^3 \\
&\leq Cn^{-1/2} n^{-3} \sum_{s=1}^{S} \mu_{s2n}^{-2} n_s \left(E_b |Y_{b,si} - \overline{Y}_{rs}|^3 + |E_b \left[\delta_{b,si} \left(Y_{b,si} - \overline{Y}_{rs} \right) \right] |^3 \right) \\
&\leq Cn^{-1/2} r^{-2} \leq Cn^{-1/2}.
\end{aligned}
$$

Note that the above inequality is obtained by applying (2.33). We have,

$$
\sup_x \left| P_b \left\{ \frac{\sqrt{n} V_n}{\sigma_{a1n}} \leq x \right\} - \Phi(x) \right| \xrightarrow{P} 0. \tag{3.44}
$$

By Polya Theorem,

$$
\frac{\sqrt{n} V_n}{\sigma_{a1n}} \xrightarrow{d} N(0, 1).
$$

Next, we want to show that

$$
\sup_x \left| P_b^* \left(\frac{\sqrt{n} U_n}{\sigma_{a2n}} \leq x \right) - \Phi(x) \right| \xrightarrow{P} 0, \tag{3.45}
$$

where

$$
\begin{aligned}
U_n &= (\hat{\mu}_b - E_b^*[\hat{\mu}_b]) \\
&= \frac{1}{n} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \left\{ (1 - \delta_{b,si}) \left(\frac{1}{J} \sum_{j=1}^{J} Y_{b,sij} - E_b^* \left[\frac{1}{J} \sum_{j=1}^{J} Y_{b,sij} \right] \right) \right\},
\end{aligned}
$$

and

$$\begin{aligned}\sigma_{a2n}^2 &= Var_b^*[\sqrt{n}U_n] = Var_b^*\left[\frac{1}{\sqrt{n}}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left\{(1-\delta_{b,si})\frac{1}{J}\sum_{j=1}^{J}Y_{b,sij} - \overline{Y}_{b,sr}\right\}\right] \\ &= \frac{1}{nJ^2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left\{(1-\delta_{b,si})\sum_{j=1}^{J}Var_b^*(Y_{b,sij})\right\} \approx \sum_{s=1}^{S}\frac{1}{J}\frac{n_s}{n}\frac{n_s - r_{b,s}}{n}\sigma^2,\end{aligned}$$

where $\bar{Y}_{b,sr} = \sum \delta_{b,si}Y_{b,si}/r_{b,s}$, and $r_{b,s} = \sum \delta_{b,si}$, so $\sigma_{a2n}^2 = \sum_{s=1}^{S} J^{-1}W_s(1-p_s)\sigma^2 + o_p(1)$ a.s [P].

By Berry-Esseen Theorem,

$$\sup_x \left| P_b^*\left(\frac{\sqrt{n}U_n}{\sigma_{a2n}} \leq x\right) - \Phi(x)\right| \leq \frac{c\rho_n^*}{(\sigma_{a2n})^3}, \tag{3.46}$$

where $\rho_n^* = n^{-3/2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{b,si})E_b^*\left|\left(\frac{1}{J}\sum_{j=1}^{J}Y_{b,sij} - E_b^*\left[\frac{1}{J}\sum_{j=1}^{J}Y_{b,sij}\right]\right)\right|^3$.

Applying the C_r-inequality (2.33), we note that

$\rho_n^* = \sum_{s=1}^{S}\sum_{i=1}^{n_s}E_b^*\left|n^{-1/2}(1-\delta_{b,si})\left(\frac{1}{J}\sum_{j=1}^{J}Y_{b,sij} - E_b^*\left[\frac{1}{J}\sum_{j=1}^{J}Y_{b,sij}\right]\right)\right|^3$

$= n^{-3/2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{b,si})E_b^*\left|J^{-1}\sum_{j=1}^{J}Y_{b,ij} - E_b^*\left[J^{-1}\sum_{j=1}^{J}Y_{b,ij}\right]\right|^3$

$\leq Cn^{-3/2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{b,si})\left(J^{-2}E_b^*\left|Y_{b,ij}\right|^3 - \left|E_b^*\left[Y_{b,ij}\right]\right|^3\right)$

$\leq Cn^{-3/2}(n-r_b) \leq Cn^{-1/2}$,

hence result (3.45). So, by Chen and Rao Theorem, as $n \to \infty$,

$$\frac{\sqrt{n}\,(U_n + V_n)}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \xrightarrow{d} N(0,1). \tag{3.47}$$

Recall that $\sqrt{n}(\hat{\mu}_b - \overline{Y}_r) = \sqrt{n}\,(U_n + V_n) + \sqrt{n}R_{n1}$. Since it was assumed that $P_b\left[\sqrt{n}|R_{n1}| > \varepsilon\right] \to 0$ a.s in [P] for any $\varepsilon > 0$, Cràmer Convergence Theorem, together with (3.47), give

$$\frac{\sqrt{n}(\hat{\mu}_b - \overline{Y}_r)}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \xrightarrow{d} N(0,1).$$

Polya Theorem gives

$$\sup_x \left| P_b\left(\frac{\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right)}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \le x \right) - \Phi(x) \right| \xrightarrow{P} 0.$$

Applying Slutsky Theorem to the denominator, we obtain

$$\sup_x \left| P_b\left(\frac{\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right)}{\sqrt{\sigma^2/p + (1-p)\sigma^2/J}} \le x \right) - \Phi(x) \right| \xrightarrow{P} 0.$$

From Theorem 3.1 we have, as $n \to \infty$, $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \Phi_{\sigma_\mu^2}(x)$ where $\Phi_{\sigma_\mu^2}(x)$ denotes $N\left(0, \sigma_\mu^2\right)$ with $\sigma_\mu^2 = \left(\frac{1}{p} + \frac{1-p}{J}\right)\sigma^2$, and so by Polya Theorem, we have

$$\sup_x \left| P_b\left(\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\sigma^2/p + (1-p)\sigma^2/J}} \le x \right) - \Phi(x) \right| \xrightarrow{P} 0.$$

Now, since

$$\begin{aligned} & \sup_x \left| P_b\left(\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right) \le x\right) - P\left(\sqrt{n}\left(\hat{\mu} - \mu\right) \le x\right) \right| \\ = & \sup_x \left| P_b\left(\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right) \le x\right) - \Phi_{\sigma_\mu^2}(x) + \Phi_{\sigma_\mu^2}(x) - P\left(\sqrt{n}\left(\hat{\mu} - \mu\right) \le x\right) \right| \qquad (3.48) \\ \le & \sup_x \left| P_b\left(\frac{\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right)}{\sigma_\mu} \le x \right) - \Phi(x) \right| + \sup_x \left| P\left(\frac{\sqrt{n}\left(\hat{\mu} - \mu\right)}{\sigma_\mu} \le x \right) - \Phi(x) \right|, \end{aligned}$$

we obtain

$$\sup_x \left| P_b\left(\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right) \le x\right) - P\left(\sqrt{n}\left(\hat{\mu} - \mu\right) \le x\right) \right| \xrightarrow{P} 0.$$

The rest of Theorem 3.2 can be proved similarly. ■

3.6.3 Proof of Theorem 3.3

The proof of Theorem 3.3 follows closely the proof of Theorem 3.1 in Qin et al. (2008) and is based on theory presented in Owen (1990). Denote $\overline{Z}^2_{si,m}(\mu) = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s} Z^2_{si,m}(\mu)$, that is

$$\begin{aligned}
\overline{Z}^2_{si,m}(\mu) &= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s} E^*\left(Z^2_{si,m}(\mu)\right) + o_p(1) \\
&= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left[Var^*\left(Z_{si,m}(\mu)\right) + \left(E^*\left(Z_{si,m}(\mu)\right)\right)^2\right] + o_p(1) \\
&= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left[(1-\delta_{si})J^{-1}Var^*\left(Y_{sij}\right) + \left(E^*\left(Z_{si,m}(\mu)\right)\right)^2\right] + o_p(1) \\
&= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left[(1-\delta_{si})J^{-1}\sigma^2 + \delta_{si}\left(Y_{si}-\mu\right)^2\right] + o_p(1) \\
&= \sigma^2\sum_{s=1}^{S} W_s\left[(1-p_s)J^{-1} + p_s\right] + o_p(1).
\end{aligned} \tag{3.49}$$

From Theorem 3.1, we have $\sqrt{n}(\hat{\mu}-\mu) \xrightarrow{d} N\left(0, \sigma^2_{nm}\right)$ where σ^2_{nm} is defined by (3.14). That is, denoting $\overline{Z}_{si,m}(\mu) = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s} Z_{si,m}(\mu)$, we have

$$n^{1/2}\sigma^{-1}_{nm}\bar{Z}_{si,m}(\mu) \xrightarrow{d} N\left(0,1\right). \tag{3.50}$$

Assume $E\left[Y_i-\mu\right]^2 < \infty$. Similarly as in Owen (1990), we note that this condition implies $\sum_{s=1}^{S}\sum_{i=1}^{n_s} P\left(\left(Y_{si}-\mu\right)^2 > n\right) < \infty$ and thus $\sum_{s=1}^{S}\sum_{i=1}^{n_s} P\left(\left|Y_{si}-\mu\right| > n^{1/2}\right) <$

∞. Then, by the Borel-Cantelli Theorem, $|Y_{si} - \mu| > n^{1/2}$ finitely often with probability 1. This implies $|max(Y_{si} - \mu)| > n^{1/2}$ finitely often or $|max Z_{si,m}(\mu)| > n^{1/2}$ finitely often. Similarly, $|Z_{si,m}(\mu)| > cn^{1/2}$ finitely often for any $c > 0$.

Therefore, $\limsup |Z_{si,m}(\mu)|/\sqrt{n} \leq c$ with probability 1. This holds for any countable set of values c, hence, with probability 1,

$$max|Z_{si,m}(\mu)| = o_p(n^{1/2}). \tag{3.51}$$

Following the steps of the proof of Theorem 1 in Owen (1990), we will now show that $\left(\sigma_{nm}^2/\sigma_{2m}^2\right)^{-1} l_{n,m}(\mu) \xrightarrow{d} \chi_1^2$. Let us introduce $\gamma_{si} := \lambda_{n,m} Z_{si,m}(\mu)$ with $\lambda_{n,m}$ defined in Section 3.3, then

$$\begin{aligned} l_{n,m}(\mu) &= 2\sum_{s=1}^{S}\sum_{i=1}^{n_s} \log\left\{1 + \lambda_{n,m} Z_{si,m}(\mu)\right\} \\ &= 2\sum_{s=1}^{S}\sum_{i=1}^{n_s} \log\left\{1 + \gamma_{si}\right\} \\ &= 2\sum_{s=1}^{S}\sum_{i=1}^{n_s} \gamma_{si} - \sum_{s=1}^{S}\sum_{i=1}^{n_s} \gamma_{si}^2 + 2\sum_{s=1}^{S}\sum_{i=1}^{n_s} \eta_{si}, \end{aligned}$$

where, for some $0 < c < \infty$, $P\left[|\eta_{si}| \leq c|\gamma_{si}|^3\right] \to 1$ as $n \to \infty$. Following Owen

(1990) let

$$\begin{aligned} g(\lambda_{n,m}) &= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\frac{Z_{si,m}(\mu)}{1+\lambda_{n,m}Z_{si,m}(\mu)} \\ &= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\frac{Z_{si,m}(\mu)}{1+\gamma_{si}} \\ &= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}Z_{si,m}(\mu)\left(1-\gamma_{si}+\frac{\gamma_{si}^2}{1-\gamma_{si}}\right) \\ &= \bar{Z}_{si,m}(\mu)-\bar{Z}^2_{si,m}(\mu)\lambda_{n,m}+\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}Z_{si,m}(\mu)\frac{\gamma_{si}^2}{1-\gamma_{si}} \\ &= \bar{Z}_{si,m}(\mu)-\bar{Z}^2_{si,m}(\mu)\lambda_{n,m}+\beta, \end{aligned}$$

where $||\beta|| = o_p(n^{-1/2})$ since, based on as shown in Owen (1990),

$$\begin{aligned} \left\|\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}Z_{si,m}(\mu)\frac{\gamma_{si}^2}{1-\gamma_{si}}\right\| &\leq \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}||Z_{si,m}(\mu)||^3||\lambda||^2\left\|\frac{1}{1-\gamma_{si}}\right\| \\ &= o(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2}). \end{aligned}$$

Therefore, as $g(\lambda_{n,m}) := 0$

$$\lambda_{n,m} = \left(\bar{Z}^2_{si,m}(\mu)\right)^{-1}\bar{Z}_{si,m}(\mu)+\beta.$$

Going back to the equation for $l_{n,m}(\mu)$, back substituting for γ_{si} and expanding $\lambda_{n,m}$, we obtain

$$
\begin{aligned}
l_{n,m}(\mu) &= 2n\lambda_{n,m}\overline{\overline{Z}}_{si,m}(\mu) - n\lambda_{n,m}^2\overline{\overline{Z}}^2_{si,m}(\mu) + 2\sum_{s=1}^{S}\sum_{i=1}^{n_s}\eta_{si} \\
&= 2n\left(\bar{Z}_{si,m}(\mu)\right)^2\left(\bar{Z}^2_{si,m}(\mu)\right)^{-1} + 2n\beta\overline{\overline{Z}}_{si,m}(\mu) - n\left(\overline{\overline{Z}}_{si,m}(\mu)\right)^2\left(\bar{Z}^2_{si,m}(\mu)\right)^{-1} + \\
&\quad -2n\beta\overline{\overline{Z}}_{si,m}(\mu) - n\beta^2\left(\overline{\overline{Z}}^2_{si,m}(\mu)\right)^{-1} + 2\sum_{s=1}^{S}\sum_{i=1}^{n_s}\eta_{si} \\
&= n\left(\overline{\overline{Z}}_{si,m}(\mu)\right)^2\left(\overline{\overline{Z}}^2_{si,m}(\mu)\right)^{-1} - n\beta^2\overline{\overline{Z}}^2_{si,m}(\mu)^{-1} + 2\sum_{s=1}^{S}\sum_{i=1}^{n_s}\eta_{si} \\
&= n\left(\overline{\overline{Z}}_{si,m}(\mu)\right)^2\left(\overline{\overline{Z}}^2_{si,m}(\mu)\right)^{-1} + R_{n1}.
\end{aligned}
$$

Now we note that $n\beta^2\left(\overline{Z}^2_{si,m}(\mu)\right)^{-1} = o_p(1)$ and by definition of η_i,

$$
|2\sum_{s=1}^{S}\sum_{i=1}^{n_s}\eta_{si}| \leq 2c||\lambda||^3\sum_{s=1}^{S}\sum_{i=1}^{n_s}||Z_{si,m}(\mu)||^3 = 2cO_p(n^{-3/2})o_p(n^{3/2}) = o_p(1).
$$

Therefore, $P\left[|R_{n1}| > \epsilon\right] \to 0$ a.s. [P].

Also,

$$
\frac{n\left(\overline{Z}_{si,m}(\mu)\right)^2}{\overline{Z}^2_{si,m}(\mu)} = \frac{n\left(\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}Z_{si,m}(\mu)\right)^2}{\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left(Z_{si,m}(\mu)\right)^2}
$$

and using results 3.49 and 3.50, we have $\left(\frac{\sigma^2_{nm}}{\sigma^2_{2m}}\right)^{-1} l_{n,m}(\mu) \xrightarrow{d} \chi^2_1$ as $n \to \infty$. That is, $l_{n,m}(\mu) \xrightarrow{d} c_m\chi^2_1$. The results corresponding to θ can be proved similarly. ■

3.6.4 Proof of Theorem 3.4

Similarly to the Proof of Theorem 3.3, it can be shown that

$$\begin{aligned}
\overline{Z}^2_{b,si,m}(\widehat{\mu}) \quad : \quad &= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left(Z_{b,si,m}(\widehat{\mu})\right)^2 = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}E_b^*\left[Z_{b,si,m}(\widehat{\mu})\right]^2 + o_p(1) \\
&= \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left(Var_b^*\left[Z_{b,si,m}(\widehat{\mu})\right] + \left(E_b^*\left[Z_{b,si,m}(\widehat{\mu})\right]\right)^2\right) + o_p(1) \\
&= \frac{1}{J}\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(1-\delta_{b,si})Var_b^*\left[Y_{b,sij}\right] + \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\delta_{b,si}\left(Y_{b,si} - \overline{Y}_{rs}\right)^2 + o_p(1) \\
&= \sigma^2_{a2} + o_p(1),
\end{aligned}$$

with $\sigma^2_{a2} \xrightarrow{P} \sigma^2_{2m}$. That is,

$$\overline{Z}^2_{b,si,m}(\widehat{\mu}) = \sigma^2_{a2} + o_p(1). \tag{3.52}$$

From the Proof of Theorem 3.1, we have as $n \to \infty$, $\frac{\sqrt{n}\left(\hat{\mu}_b - \overline{Y}_r\right)}{\sqrt{\sigma^2_{1nm} + \sigma^2_{2nm}}} \xrightarrow{d} N(0,1)$ where $\sigma^2_{nm} = \sigma^2_{1nm} + \sigma^2_{2nm} = \sigma^2 \sum_{s=1}^{S} W_s p_s^{-1} + \sum_{s=1}^{S} W_s(1-p_s)\dfrac{1}{J}\dfrac{1}{r_s}\sum_{i\in r_s}(Y_{ij} - \bar{Y}_{rs})^2$. That is, denoting $\bar{Z}_{b,si,m}(\widehat{\mu}) = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s} Z_{b,si,m}(\widehat{\mu})$, we have

$$\frac{\sqrt{n}\bar{Z}_{b,si,m}(\widehat{\mu})}{\sqrt{\sigma^2_{nm}}} \xrightarrow{d} N(0,1). \tag{3.53}$$

Assume $E\left[Y_{si} - \overline{Y}_r\right]^2 < \infty$. Similarly, as in Owen (1990) and in the Proof of Theorem 3.3, we obtain

$$max|Z_{b,si,m}(\widehat{\mu})| = o_p(n^{1/2}). \tag{3.54}$$

and following the steps of the Proof of Theorem 1 in Owen (1990), and similar to the Proof of Theorem 3.3, we can show that

$$l_{b,n,m}(\widehat{\mu}) = \frac{\left(\sqrt{n}\overline{Z}_{b,si,m}(\widehat{\mu})\right)^2}{\overline{Z}^2_{b,si,m}(\widehat{\mu})} + R_{n2},$$

with $P\left[|R_{n2}| > \epsilon\right] \xrightarrow{P} 0$. Noting that $\left(\sqrt{n}\overline{Z}_{b,si,m}(\widehat{\mu})\right)^2 = \sigma^2_{nm}\left(\sqrt{n}\bar{Z}_{b,si,m}(\widehat{\mu})/\sqrt{\sigma^2_{nm}}\right)^2$, and using results (3.52) and (3.53), we obtain

$$\sup_{x \in R}\left| P_b\left\{\frac{\overline{Z}^2_{b,si,m}(\widehat{\mu})}{\sigma^2_{nm}} l_{b,n,m}(\hat{\mu}) \le x\right\} - P\left\{\chi^2_1 \le x\right\}\right| \xrightarrow{P} 0,$$

where $\left(\overline{Z}^2_{b,si,m}(\widehat{\mu})/\sigma^2_{nm}\right)^{-1} = \sigma^2_{nm}/\sigma^2_{2m} + o_p(1) = c_m + o_p(1)$ as defined in Theorem 3.3.

By Theorem 3.3, $l_{n,m}(\mu) \xrightarrow{d} c_m\chi^2_1$, therefore

$$\sup_{x \in R}\left|P_b\left\{l_{b,n,m}(\hat{\mu}) \le x\right\} - P\left\{l_{n,m}(\mu) \le x\right\}\right| \xrightarrow{P} 0. \tag{3.55}$$

We follow the same steps to prove the corresponding result for the distribution function. ■

Chapter 4

Confidence Intervals for Correlation Coefficient Under Joint Regression Imputation

4.1 Introduction

We have presented the theory to obtain bootstrap percentile and empirical likelihood confidence intervals on the univariate mean and distribution function of y in the presence of missing data in Chapters 2 and 3. In this chapter, we consider the case of bivariate (y, z), with possible nonresponse in y and z, and construct confidence intervals on the correlation coefficient

$$\rho = \frac{E(yz) - E(y)E(z)}{\sqrt{\left(E(y^2) - E^2(y)\right)\left(E(z^2) - E^2(z)\right)}}. \tag{4.1}$$

To compensate for missing data, we use joint regression imputation (Shao and Wang, 2002) which preserves unbiasedness for each component of the correlation coefficient equation (4.1). An imputed estimator that is unbiased for the first and second marginal moments and the cross-product moment for y and z, is approximately unbiased for ρ. Shao and Wang (2002) investigated estimation of sample correlation coefficient based on survey data under joint regression imputation and showed that the usual estimators for sample correlation coefficients are consistent. They showed that their imputation method is model unbiased for marginal totals, second moments and cross-product moments. However, they did not study asymptotic confidence intervals for these parameters. In this chapter, we investigate asymptotic properties of these estimators and construct bootstrap percentile and bootstrap empirical-likelihood confidence intervals on the correlation coefficient ρ after applying joint regression imputation to the data.

The chapter is organized as follows. In Section 4.2, we study asymptotic normality of the estimators and construct bootstrap percentile confidence interval on ρ. In Section 4.3, we form the empirical likelihood ratio statistic, obtain its limiting distribution and construct bootstrap-calibrated empirical likelihood confidence intervals on ρ. We show that the confidence intervals have asymptotically correct coverage accuracy. Results of a simulation study to assess the performance of the proposed confidence intervals are presented in Section 4.4. All proofs are differed to the Appendix (Section 4.6).

4.1.1 Framework

In this chapter, we focus on the correlation coefficient (4.1) between two study variables, y and z, both with possible nonresponse. Let a be response indicator variable for y, that is $a = 0$ if y is missing and $a = 1$ otherwise, and similarly let b be response indicator for z with $b = 0$ if z is missing and $b = 1$ otherwise. Further, we suppose that x is a vector of fully observed covariates. We assume that (y, z) are missing at random (MAR) given x that is, (a, b) and (y, z) are conditionally independent given x, or

$$P\left((a,b) = (k,l) \mid y, z, x\right) = P\left((a,b) = (k,l) \mid x\right) \text{ for any } k, l = 0, 1. \quad (4.2)$$

4.1.2 Joint Regression Imputation

The joint regression imputation is an extension to the popular marginal random regression imputation, the latter being unbiased only for the marginal first and second moments, the first also for the cross-product moment. The method's advantages are that it is unbiased for the correlation coefficient and, unlike most known imputation methods, it does not impose any distributional assumptions or parametric modeling; consequently, it is especially applicable when dealing with data from complex surveys (after incorporating survey weights) (Shao and Wang, 2002). However, while this method preserves well the relationships between two variables, it may lead to inefficient estimators as it suffers from additional variability due to the random selection of residuals (Chauvet and Haziza, 2012).

We form two linear regression models for y and z to apply joint regression imputation. In particular

$$y = \beta' x + v^{1/2}\epsilon \text{ and } z = \gamma' x + u^{1/2}\eta, \tag{4.3}$$

where β and γ are $d \times 1$ vectors of regression parameters, $v = v(x)$ and $u = u(x)$ are known strictly positive functions of x, the random terms ϵ and η have zero means and finite, independent of x, variances. Under models (4.2) and (4.3), we generate an i.i.d. sample of incomplete observations, $\{(x_i, y_i, z_i, a_i, b_i), i = 1, ..., n\}$, from (x, y, z, a, b) where all the x_i's are observed. The respective estimators of β and γ in (4.3) can be obtained using the weighted least squares (WLS) on completely observed pairs of data, that is

$$\hat{\beta}_r = \left(\sum_{i=1}^{n} \frac{a_i x_i x_i'}{v_i}\right)^{-1} \sum_{i=1}^{n} \frac{a_i x_i y_i}{v_i}, \text{ and } \hat{\gamma}_r = \left(\sum_{i=1}^{n} \frac{b_i x_i x_i'}{u_i}\right)^{-1} \sum_{i=1}^{n} \frac{b_i x_i y_i}{u_i}, \tag{4.4}$$

where $v_i = v(x_i)$ and $u_i = u(x_i)$. We can then use these statistics to estimate the covariance matrix of random terms ϵ and η by letting

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_\epsilon^2 & \hat{\sigma}_{\epsilon,\eta} \\ \hat{\sigma}_{\epsilon,\eta} & \hat{\sigma}_\eta^2 \end{pmatrix} \tag{4.5}$$

$$= \sum_{i=1}^{n} a_i b_i \begin{pmatrix} r_{yi}^2 & r_{yi} r_{zi} \\ r_{yi} r_{zi} & r_{zi}^2 \end{pmatrix} \left(\sum_{i=1}^{n} a_i b_i\right)^{-1}, \tag{4.6}$$

where $r_{yi} = v_i^{-1/2}(y_i - \hat{\beta}_r' x_i)$,and $r_{zi} = u_i^{-1/2}(z_i - \hat{\gamma}_r' x_i)$.

The joint regression imputation procedure, for i.i.d. data, consists of the following cases of non-response treatment for each pair of observations (y_i, z_i), $i = 1, ..., n$:

1. If only y_i is missing, that is $(a_i, b_i) = (0, 1)$, the missing y_i is imputed by

$$y_i^* = \hat{\beta}_r' x_i + \frac{v_i^{1/2} \hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2} \hat{\sigma}_\eta^2} (z_i - \hat{\gamma}_r' x_i) + v_i^{1/2} \tilde{\epsilon}_i^*,$$

   where, given the observed data, the $\tilde{\epsilon}_i^*$s are independently generated from a population with mean 0 and variance $\left(\hat{\sigma}_\epsilon^2 - \frac{\hat{\sigma}_{\epsilon,\eta}^2}{\hat{\sigma}_\eta^2} \right)$.

2. If only z_i is missing, that is $(a_i, b_i) = (1, 0)$, the missing value z_i is imputed by

$$z_i^* = \hat{\gamma}_r' x_i + \frac{u_i^{1/2} \hat{\sigma}_{\epsilon,\eta}}{v_i^{1/2} \hat{\sigma}_\epsilon^2} (y_i - \hat{\beta}_r' x_i) + u_i^{1/2} \tilde{\eta}_i^*,$$

   where, given the observed data, the $\tilde{\eta}_i^*$s are independently generated from a population with mean 0 and variance $\left(\hat{\sigma}_\eta^2 - \frac{\hat{\sigma}_{\epsilon,\eta}^2}{\hat{\sigma}_\epsilon^2} \right)$.

3. If both y_i and z_i are missing, that is $(a_i, b_i) = (0, 0)$, the missing values (y_i, z_i) are imputed by

$$(y_i^*, z_i^*) = \left(\hat{\beta}_r' x_i,\ \hat{\gamma}_r' x_i \right) + \left(v_i^{1/2} \epsilon_i^*,\ u_i^{1/2} \eta_i^* \right),$$

   where, given the observed data, the (ϵ_i^*; η_i^*)'s are independently generated from a population with mean $(0, 0)$ and covariance $\hat{\Sigma}$.

After applying the joint regression imputation, the imputed data are denoted by

$$(\tilde{y}_i, \tilde{z}_i), \ i = 1, ..., n, \tag{4.7}$$

where $(\tilde{y}_i, \tilde{z}_i) = a_i b_i (y_i, z_i) + (1-a_i) b_i (y_i^*, z_i) + a_i (1-b_i)(y_i, z_i^*) + (1-a_i)(1-b_i)(y_i^*, z_i^*)$.

4.2 Normal Approximation

Let us begin by defining the vector of components of ρ, that is the first and the second marginal moments and the cross-product moment for y and z, namely

$$\theta = (\theta_1, ..., \theta_5)' = \left(E(y), E(z), E(y^2), E(z^2), E(yz)\right)'. \tag{4.8}$$

After applying the joint regression imputation method to the original data, the imputed estimator of θ is given by

$$\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5\right)', \tag{4.9}$$

where

$$
\begin{aligned}
\hat{\theta}_1 &= \frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}a_i y_i + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i y_i^* + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)y_i^* \\
&= \frac{1}{n}\sum_{i=1}^{n}a_i y_i + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i) + v_i^{1/2}\tilde{\epsilon}_i^*\right) + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\hat{\beta}_r' x_i + v_i^{1/2}\epsilon_i^*\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{\theta}_3 &= \frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}a_i y_i^2 + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i y_i^{*2} + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)y_i^{*2} \\
&= \frac{1}{n}\sum_{i=1}^{n}a_i y_i^2 + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i) + v_i^{1/2}\tilde{\epsilon}_i^*\right)^2 + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\hat{\beta}_r' x_i + v_i^{1/2}\epsilon_i^*\right)^2.
\end{aligned}
$$

Note that $\hat{\theta}_2$ and $\hat{\theta}_4$ can be obtained in the similar way as $\hat{\theta}_1$ and $\hat{\theta}_3$ above. Finally,

$$
\begin{aligned}
\hat{\theta}_5 &= \frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i\tilde{z}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}a_ib_iy_iz_i + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_iy_i^*z_i + \frac{1}{n}\sum_{i=1}^{n}a_i(1-b_i)y_iz_i^* + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)y_i^*z_i^* \\
&= \frac{1}{n}\sum_{i=1}^{n}a_ib_iy_iz_i + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i\left(\hat{\beta}_r'x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i-\hat{\gamma}_r'x_i) + v_i^{1/2}\tilde{\epsilon}_i^*\right)z_i + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}a_i(1-b_i)y_i\left(\hat{\gamma}_r'x_i + \frac{u_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{v_i^{1/2}\hat{\sigma}_\epsilon^2}(y_i-\hat{\beta}_r'x_i) + u_i^{1/2}\tilde{\eta}_i^*\right) + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\hat{\beta}_r'x_i + v_i^{1/2}\epsilon_i^*\right)\left(\hat{\gamma}_r'x_i + u_i^{1/2}\eta_i^*\right).
\end{aligned}
$$

We will now show that joint random regression imputation is model unbiased for the marginal first and second moments as well as for the cross-product moment. Let E_m denote expectation under models (4.2) and (4.3) and E^* be the expectation with respect to randomness in the imputation procedure. For the first marginal moment, we have

$$
\begin{aligned}
E^*\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i\right) &= \frac{1}{n}\sum_{i=1}^{n}E^*(\tilde{y}_i) \qquad (4.10)\\
&= \frac{1}{n}\sum_{i=1}^{n}a_iy_i + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_iE^*\left(y_i^*\right) + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left(y_i^*\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}a_iy_i + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_iE^*\left(\hat{\beta}_r'x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i-\hat{\gamma}_r'x_i) + v_i^{1/2}\tilde{\epsilon}_i^*\right) +\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left(\hat{\beta}_r'x_i + v_i^{1/2}\epsilon_i^*\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(a_iy_i + (1-a_i)b_i\left(\hat{\beta}_r'x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}^2}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i-\hat{\gamma}_r'x_i)\right) + \right.\\
&\quad \left. +(1-a_i)(1-b_i)\left(\hat{\beta}_r'x_i\right)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(a_iy_i + (1-a_i)\hat{\beta}_r'x_i + (1-a_i)b_i\frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i-\hat{\gamma}_r'x_i)\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E_mE^*\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i\right) &\simeq E_m\left(\frac{1}{n}\sum_{i=1}^{n}\left(a_i\left(\beta'x_i + v_i^{1/2}\epsilon_i\right) + \right.\right.\\
&\quad \left.\left. + (1-a_i)\beta'x_i + (1-a_i)b_i\frac{v_i^{1/2}\sigma_{\epsilon,\eta}}{\sigma_\eta^2}\eta_i\right)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}E_m(\beta'x_i) = E_m\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right).
\end{aligned}
$$

The result is obtained by expanding $y_i = \beta'x_i + v_i^{1/2}\epsilon_i$, substituting for $(z_i - \gamma'x_i)\,u_i^{-1/2} = \eta_i$, $E_m(\epsilon_i) = 0$, $E_m(\eta_i) = 0$ and after noting that $E_m\left(y_i\right) = E_m\left(\beta'x_i + v_i^{1/2}\epsilon_i\right) = n^{-1}\sum_{i=1}^{n}E_m(\beta'x_i) = E_m(\beta'x_i)$. Similarly, we can show that $E_mE^*\left(n^{-1}\sum_{i=1}^{n}\tilde{z}_i\right) = E_m(n^{-1}\sum_{i=1}^{n}z_i)$. For the second moment, noting that $E^*(\tilde{\epsilon}_i^{*2}) = Var^*(\tilde{\epsilon}_i^*) + E^*(\tilde{\epsilon}_i^*)^2 =$

$\hat{\sigma}_\epsilon^2 - \hat{\sigma}_{\epsilon,\eta}^2/\hat{\sigma}_\eta^2$ since $E^*(\tilde{\epsilon}_i^*) = 0$, we obtain

$$\begin{aligned}
E^*\left(\frac{1}{n}\sum_{i=1}^n \tilde{y}_i^2\right) &= \frac{1}{n}\sum_{i=1}^n a_i y_i^2 + \frac{1}{n}\sum_{i=1}^n (1-a_i) b_i E^*\left(y_i^{*2}\right) + \\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-a_i)(1-b_i) E^*\left(y_i^{*2}\right) \\
&= \frac{1}{n}\sum_{i=1}^n a_i y_i^2 + \frac{1}{n}\sum_{i=1}^n (1-a_i) b_i E^*\left\{\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)^2 + \right. \\
&\quad \left. + 2\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)\left(v_i^{1/2}\tilde{\epsilon}_i^*\right) + v_i \tilde{\epsilon}_i^{*2}\right\} + \\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-a_i)(1-b_i) E^*\left\{\left(\hat{\beta}_r' x_i\right)^2 + 2\left(\hat{\beta}_r' x_i\right)\left(v_i^{1/2}\epsilon_i^*\right) + v_i \epsilon_i^{*2}\right\} \\
&= \frac{1}{n}\sum_{i=1}^n a_i y_i^2 + \frac{1}{n}\sum_{i=1}^n (1-a_i) b_i \left(\left(\hat{\beta}_r' x_i\right)^2 + 2\hat{\beta}_r' x_i \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i) + \right. \\
&\quad \left. + \left(\frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)^2 + v_i Var^*(\tilde{\epsilon}_i^*)\right) + \\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-a_i)(1-b_i)\left\{\left(\hat{\beta}_r' x_i\right)^2 + v_i Var^*(\epsilon_i^*)\right\} \\
&= \frac{1}{n}\sum_{i=1}^n a_i y_i^2 + \frac{1}{n}\sum_{i=1}^n (1-a_i) b_i \left(\left(\hat{\beta}_r' x_i\right)^2 + 2\hat{\beta}_r' x_i \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\eta^2}\eta_i + \right. \\
&\quad \left. + \frac{v_i \hat{\sigma}_{\epsilon,\eta}^2}{\hat{\sigma}_\eta^4}\eta_i^2 + v_i\left(\hat{\sigma}_\epsilon^2 - \frac{\hat{\sigma}_{\epsilon,\eta}^2}{\hat{\sigma}_\eta^2}\right)\right) + \\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-a_i)(1-b_i)\left\{\left(\hat{\beta}_r' x_i\right)^2 + v_i \hat{\sigma}_\epsilon^2\right\}.
\end{aligned} \tag{4.11}$$

and so

$$
\begin{aligned}
E_m E^*\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i^2\right) &\simeq E_m\left(\frac{1}{n}\sum_{i=1}^{n}a_i y_i^2\right) + \\
&+\frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i\left(\left(\beta' x_i\right)^2 + \frac{v_i\sigma_{\epsilon,\eta}^2}{\sigma_\eta^4}E_m\left(\eta_i^2\right) + v_i\left(\sigma_\epsilon^2 - \frac{\sigma_{\epsilon,\eta}^2}{\sigma_\eta^2}\right)\right) + \\
&+\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\left(\beta' x_i\right)^2 + v_i\sigma_\epsilon^2\right) \\
&= E_m\left(\frac{1}{n}\sum_{i=1}^{n}a_i y_i^2\right) + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)\left(E_m\left(\beta' x_i\right)^2 + v_i\sigma_\epsilon^2\right) \\
&= E_m\left(\frac{1}{n}\sum_{i=1}^{n}y_i^2\right).
\end{aligned}
$$

since $E_m\left(\eta_i^2\right) = \sigma_\eta^2$ and after noting that $E_m\left(y_i^2\right) = E_m\left(\beta' x_i + v_i^{1/2}\epsilon_i\right)^2 = E_m\left(\beta' x_i\right)^2 + v_i Var_m(\epsilon_i) = E_m\left(\beta' x_i\right)^2 + v_i\sigma_\epsilon^2$. Similarly, we can show that $E_m E^*\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{z}_i^2\right) = E_m(\frac{1}{n}\sum_{i=1}^{n}z_i^2)$. Finally, for the cross-product moment, we have

$$
\begin{aligned}
E^*\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i\tilde{z}_i\right) &= \frac{1}{n}\sum_{i=1}^{n}a_i b_i y_i z_i + \\
&+\frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i E^*\left(y_i^*\right)z_i + \frac{1}{n}\sum_{i=1}^{n}a_i(1-b_i)y_i E^*\left(z_i^*\right) + \\
&+\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left(y_i^* z_i^*\right).
\end{aligned}
\tag{4.12}
$$

The second component of the above summation can be written as

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_iE^*\left(y_i^*\right)z_i &= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_iz_iE^*\left\{\hat{\beta}_r'x_i+\frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}\left(z_i-\hat{\gamma}_rx_i\right)+v_i^{1/2}\tilde{\epsilon}_i^*\right\}\\ &= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_iz_i\left(\hat{\beta}_r'x_i+\frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\eta^2}\eta_i\right).\end{aligned}$$

Similarly, the third component is

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}a_i(1-b_i)y_iE^*\left(z_i^*\right) &= \frac{1}{n}\sum_{i=1}^{n}a_i(1-b_i)y_iE^*\left\{\hat{\gamma}_r'x_i+\frac{u_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{v_i^{1/2}\hat{\sigma}_\epsilon^2}\left(y_i-\hat{\beta}_rx_i\right)+u_i^{1/2}\tilde{\eta}_i^*\right\}\\ &= \frac{1}{n}\sum_{i=1}^{n}a_i(1-b_i)y_i\left(\hat{\gamma}_r'x_i+\frac{u_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\epsilon^2}\varepsilon_i\right).\end{aligned}$$

Finally, for the last component, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left(y_i^*z_i^*\right) &= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E_*\left\{\left(\hat{\beta}_r'x_i\right)\left(\hat{\gamma}_r'x_i\right)+\hat{\beta}_r'x_iu_i^{1/2}\eta_i^*+\right.\\ &\quad\left.+\hat{\gamma}_r'x_iv_i^{1/2}\epsilon_i^*+u_i^{1/2}v_i^{1/2}\eta_i^*\epsilon_i^*\right\}\\ &= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\left(\hat{\beta}_r'x_i\right)\left(\hat{\gamma}_r'x_i\right)+u_i^{1/2}v_i^{1/2}Cov^*(\eta_i^*,\epsilon_i^*)\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\left(\hat{\beta}_r'x_i\right)\left(\hat{\gamma}_r'x_i\right)+u_i^{1/2}v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}\right).\end{aligned}$$

So that

$$
\begin{aligned}
E_m E^* \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{y}_i \tilde{z}_i \right) \simeq\ & E_m \left(\frac{1}{n} \sum_{i=1}^{n} a_i b_i y_i z_i + \frac{1}{n} \sum_{i=1}^{n} (1-a_i) b_i z_i \left(\beta' x_i + \frac{v_i^{1/2} \sigma_{\epsilon,\eta}}{\sigma_\eta^2} \eta_i \right) + \right. \\
& + \frac{1}{n} \sum_{i=1}^{n} a_i (1-b_i) y_i \left(\gamma' x_i + \frac{u_i^{1/2} \sigma_{\epsilon,\eta}}{\sigma_\epsilon^2} \varepsilon_i \right) + \\
& \left. + \frac{1}{n} \sum_{i=1}^{n} (1-a_i)(1-b_i) \left((\beta' x_i)(\gamma' x_i) + u_i^{1/2} v_i^{1/2} \sigma_{\epsilon,\eta} \right) \right) \\
=\ & E_m \left(\frac{1}{n} \sum_{i=1}^{n} a_i b_i \left((\beta' x_i)(\gamma' x_i) + u_i^{1/2} v_i^{1/2} \sigma_{\epsilon,\eta} \right) \right) + \\
& + E_m \left(\frac{1}{n} \sum_{i=1}^{n} (1-a_i) b_i \left(\gamma' x_i + u_i^{1/2} \eta_i \right) \left(\beta' x_i + \frac{v_i^{1/2} \sigma_{\epsilon,\eta}}{\sigma_\eta^2} \eta_i \right) \right) + \\
& + E_m \left(\frac{1}{n} \sum_{i=1}^{n} a_i (1-b_i) \left(\beta' x_i + v_i^{1/2} \varepsilon_i \right) \left(\gamma' x_i + \frac{u_i^{1/2} \sigma_{\epsilon,\eta}}{\sigma_\epsilon^2} \varepsilon_i \right) \right) + \\
& + E_m \left(\frac{1}{n} \sum_{i=1}^{n} (1-a_i)(1-b_i) \left((\beta' x_i)(\gamma' x_i) + u_i^{1/2} v_i^{1/2} \sigma_{\epsilon,\eta} \right) \right) \\
=\ & E_m \left(\frac{1}{n} \sum_{i=1}^{n} (\beta' x_i)(\gamma' x_i) + \frac{1}{n} \sum_{i=1}^{n} u_i^{1/2} v_i^{1/2} \sigma_{\epsilon,\eta} \right) \\
=\ & E_m \left(\frac{1}{n} \sum_{i=1}^{n} y_i z_i \right).
\end{aligned}
$$

4.2.1 Ordinary Confidence Intervals

The result on the asymptotic normality of $\hat{\theta}$ is summarized in Theorem 4.1.

Theorem 4.1 *Assume that* $E(ab) > 0$, $0 < Ey^4 < \infty$, $0 < Ez^4 < \infty$. *Then there exists a* 5×5 *matrix* $\Sigma > 0$ *such that as* $n \to \infty$,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma). \tag{4.13}$$

Let us define a function

$$h(g) = h(u_1, ..., u_5) = \frac{u_5 - u_1 u_2}{\sqrt{(u_3 - u_1^2)\,(u_4 - u_2^2)}}. \tag{4.14}$$

The correlation coefficient ρ is a smooth function of θ, and the estimator of ρ is $h(\hat{\theta})$ with h defined by (4.14), that is

$$\hat{\rho} = h(\hat{\theta}) = \frac{\hat{\theta}_5 - \hat{\theta}_1\hat{\theta}_2}{\sqrt{(\hat{\theta}_3 - \hat{\theta}_1^2)(\hat{\theta}_4 - \hat{\theta}_2^2)}}, \tag{4.15}$$

where $\hat{\theta}_i$ is the i^{th} component of $\hat{\theta}$, $i = 1, ..., 5$. Using this approach together with Theorem 4.1 we can obtain the result on asymptotic normality for $\hat{\rho}$.

Theorem 4.2 *Suppose conditions of Theorem 4.1. are satisfied, then as* $n \to \infty$

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, \sigma^2) \tag{4.16}$$

where $\sigma^2 = c_0' \Sigma c_0$, *and*

$$c_0 = \left(\frac{\partial h(\theta)}{\partial \theta_1}, ..., \frac{\partial h(\theta)}{\partial \theta_5}\right)'. \tag{4.17}$$

The standard $(1-\alpha)$-level normal approximation confidence interval on $\hat{\rho}$ is given by

$$\left(\hat{\rho} - z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}},\ \hat{\rho} + z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}\right), \tag{4.18}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile from the standard normal distribution. However, since the asymptotic variance of $\sqrt{n}(\hat{\rho} - \rho)$ is too complicated to use in practice, we need to use bootstrap method to approximate the asymptotic distribution of $\sqrt{n}(\hat{\rho} - \rho)$.

4.2.2 Bootstrap Confidence Intervals

As discussed in Shao and Sitter (1996), applying the usual bootstrap method leads to invalid results with missing data. Therefore, to overcome this issue, we impute each bootstrap sample in the same way as the original data set. The procedure is as follows. Let $b = 1, ..., B$ where B represents a large number of bootstrap samples.

1. We draw a simple random sample $D^* = \{(x_{i,b}, y_{i,b}, z_{i,b}, a_{i,b}, b_{i,b}),\ i = 1, ..., n\}$ with replacement from the data set $D = \{(x_i, \tilde{y}_i, \tilde{z}_i, a_i, b_i)\}$ with $\tilde{y}_i$ and $\tilde{z}_i$ defined in (4.7). Let $v_{i,b}^{1/2} = v^{1/2}(x_{i,b})$ and $u_{i,b}^{1/2} = u^{1/2}(x_{i,b})$.

2. When $(a_{i,b}, b_{i,b}) = (0, 1)$, the missing values $y_{i,b}$'s are imputed by

$$y_{i,b}^* = \hat{\beta}_{r,b}' x_{i,b} + \frac{v_{i,b}^{1/2} \hat{\sigma}_{\epsilon,\eta,b}}{u_{i,b}^{1/2} \hat{\sigma}_{\eta,b}^2} (z_{i,b} - \hat{\gamma}_{r,b}' x_{i,b}) + v_{i,b}^{1/2} \tilde{\epsilon}_{i,b}^*,$$

   where, given D^*, the $\tilde{\epsilon}_{i,b}^*$s are independently generated from a population with mean 0 and variance $\left(\hat{\sigma}_{\epsilon,b}^2 - \hat{\sigma}_{\epsilon,\eta,b}^2 / \hat{\sigma}_{\eta,b}^2\right)$,with $\hat{\beta}_{r,b}'$ and $\hat{\gamma}_{r,b}'$ defined similarly as in (4.4) with data replaced by the counterpart data in D^* , and $\hat{\sigma}_{\epsilon,b}^2$, $\hat{\sigma}_{\gamma,b}^2$ and $\hat{\sigma}_{\epsilon,\eta,b}$ are defined similarly to (4.5) with data replaced by the counterpart data in D^*. In cases when $(a_{i,b}, b_{i,b}) = (1, 0)$, or when $(a_{i,b}, b_{i,b}) = (0, 0)$, the missing values in $z_{i,b}'$s, or $(y_{i,b}, z_{i,b})'$s respectively, are imputed similarly as per the procedure described before with data D replaced by D^*.

After the imputation is applied, the bootstrap imputed data are denoted by

$$(\tilde{y}_{i,b}, \tilde{z}_{i,b}),\ i = 1, 2, ..., n,\ b = 1, ..., B, \tag{4.19}$$

and the estimator of θ based on the bootstrap imputed data are

$$\hat{\theta}_b = \begin{bmatrix} \hat{\theta}_{1b} \\ \hat{\theta}_{2b} \\ \hat{\theta}_{3b} \\ \hat{\theta}_{4b} \\ \hat{\theta}_{5b} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^{n} \tilde{y}_{i,b} \\ \sum_{i=1}^{n} \tilde{z}_{i,b} \\ \sum_{i=1}^{n} \tilde{y}_{i,b}^2 \\ \sum_{i=1}^{n} \tilde{z}_{i,b}^2 \\ \sum_{i=1}^{n} \tilde{y}_{i,b}\tilde{z}_{i,b} \end{bmatrix}. \tag{4.20}$$

The usual bootstrap analogues of $(\hat{\theta} - \theta)$ are given by $(\hat{\theta}_b - \hat{\theta})$; however, in the presence of missing data, under joint regression imputation, we show that the distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is approximated by its modified bootstrap version $\sqrt{n}(\hat{\theta}_b - E^*(\hat{\theta}))$. This result is stated in Theorem 4.3. below.

Theorem 4.3 *Suppose that the conditions of Theorem 4.1 are satisfied. Then, conditioning on D, as $n \to \infty$,*

$$\Sigma_b^{-1/2}\sqrt{n}\left(\hat{\theta}_b - E^*\left(\hat{\theta}\right)\right) \xrightarrow{d} N(0, I_5), \tag{4.21}$$

where $\Sigma_b \xrightarrow{P} \Sigma$ for Σ defined in Theorem 4.1, and $\sqrt{n}\left(E^*\left(\hat{\theta}\right) - \hat{\theta}\right) \xrightarrow{d} N(0, \Sigma_0)$ for some $\Sigma_0 > 0$.

Note that $E^*\left(\hat{\theta}\right) = \hat{\theta} + \mu_{n,b}$, where $\mu_{n,b}$ represents the proposed adjustment to the ordinary method that needs to be applied to the imputed estimator $\hat{\theta}$ in order to obtain the asymptotic normality result (4.21). In particular,

$$\begin{aligned} \mu_{n,b} \quad &: \quad = E^*(\hat{\theta}) - \hat{\theta} \\ &= \quad (\mu_{n,b,1}, \mu_{n,b,2}, \mu_{n,b,3}, \mu_{n,b,4}, \mu_{n,b,5})'. \end{aligned} \tag{4.22}$$

Using previous derivations for $E^*\left(\hat{\theta}\right)$ (see equations (4.10),(4.11) and (4.12)), we obtain

$$\begin{aligned} \mu_{n,b,1} \quad &: \quad = E^*\left(\hat{\theta}_1\right) - \hat{\theta}_1 \\ = \quad & \frac{1}{n}\sum_{i=1}^{n} a_i y_i + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i E^*\left\{\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i) + v_i^{1/2}\tilde{\epsilon}_i^*\right\} + \\ & + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left\{\left(\hat{\beta}_r' x_i\right) + v_i^{1/2}\epsilon_i^*\right\} + \\ & - \frac{1}{n}\sum_{i=1}^{n} a_i y_i + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i\left\{\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i) + v_i^{1/2}\tilde{\epsilon}_i^*\right\} + \\ & - \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left\{\left(\hat{\beta}_r' x_i\right) + v_i^{1/2}\epsilon_i^*\right\} \\ = \quad & -\frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i v_i^{1/2}\tilde{\epsilon}_i^* - \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)v_i^{1/2}\epsilon_i^*. \end{aligned}$$

Similarly,

$$\begin{aligned} \mu_{n,b,2} \quad &: \quad = E^*\left(\hat{\theta}_2\right) - \hat{\theta}_2 \\ &= \quad -\frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i)\sqrt{u_i}\tilde{\eta}_i^* - \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\sqrt{u_i}\eta_i^*, \end{aligned}$$

and

$$
\begin{aligned}
\mu_{n,b,3} \quad : \quad & = E^*\left(\hat{\theta}_3\right) - \hat{\theta}_3 \\
= \quad & \frac{1}{n}\sum_{i=1}^{n} a_i y_i^2 + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i E^*\left[\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)^2 + \right. \\
& \left. +2\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)\left(v_i^{1/2}\tilde{\epsilon}_i^*\right) + v_i\tilde{\epsilon}_i^{*2}\right] + \\
& +\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left[\left(\hat{\beta}_r' x_i\right)^2 + 2\left(\hat{\beta}_r' x_i\right)\left(v_i^{1/2}\epsilon_i^*\right) + v_i\epsilon_i^{*2}\right] + \\
& -\frac{1}{n}\sum_{i=1}^{n} a_i y_i^2 + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i\left[\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)^2 + \right. \\
& \left. -2\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)\left(v_i^{1/2}\tilde{\epsilon}_i^*\right) + v_i\tilde{\epsilon}_i^{*2}\right] + \\
& -\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left[\left(\hat{\beta}_r' x_i\right)^2 + 2\left(\hat{\beta}_r' x_i\right)\left(v_i^{1/2}\epsilon_i^*\right) + v_i\epsilon_i^{*2}\right] \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i\left(-2\left(\hat{\beta}_r' x_i + \frac{\sqrt{v_i}\hat{\sigma}_{\epsilon,\eta}}{\sqrt{u_i}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)\sqrt{v_i}\tilde{\epsilon}_i^* + \right. \\
& \left. -v_i\left(\tilde{\epsilon}_i^{*2} - \hat{\sigma}_\epsilon^2 + \frac{\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\eta^2}\right)\right) + \\
& +\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left[-2\left(\hat{\beta}_r' x_i\right)\sqrt{v_i}\epsilon_i^* - v_i\epsilon_i^{*2} + v_i\hat{\sigma}_\epsilon^2\right].
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mu_{n,b,4} \quad : \quad & = E^{*}\left(\hat{\theta}_4\right) - \hat{\theta}_4 \\
& = \frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i)\left[-2\left(\hat{\gamma}_r' x_i + \frac{\sqrt{u_i}\hat{\sigma}_{\epsilon,\eta}}{\sqrt{v_i}\hat{\sigma}_{\epsilon}^2}(y_i - \hat{\beta}_r' x_i)\right)\sqrt{u_i}\tilde{\eta}_i^{*} + \right. \\
& \quad \left. -u_i\left(\tilde{\eta}_i^{*2} - \hat{\sigma}_{\eta}^2 + \frac{\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_{\epsilon}^2}\right)\right] + \\
& \quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left[-2\left(\hat{\gamma}_r' x_i\right)\sqrt{u_i}\eta_i^{*} - u_i\eta_i^{*2} + u_i\hat{\sigma}_{\eta}^2\right].
\end{aligned}
$$

Finally, we can show that

$$
\begin{aligned}
\mu_{n,b,5} \quad : \quad & = E^{*}\left(\hat{\theta}_5\right) - \hat{\theta}_5 \\
& = \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i(E^{*}\left(y_i^{*}\right)z_i - y_i^{*}z_i) + \frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i)(y_i E^{*}\left(z_i^{*}\right) - y_i z_i^{*}) + \\
& \quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)(E^{*}\left(y_i^{*}z_i^{*}\right) - y_i^{*}z_i^{*}) \\
& = -\frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i z_i\sqrt{v_i}\tilde{\epsilon}_i^{*} - \frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i)y_i\sqrt{u_i}\tilde{\eta}_i^{*} + \\
& \quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i) \times \\
& \quad \times\left(\left(-\hat{\beta}_r' x_i\right)\sqrt{u_i}\eta_i^{*} - \left(\hat{\gamma}_r' x_i\right)\sqrt{v_i}\epsilon_i^{*} - \sqrt{u_i}\sqrt{v_i}\epsilon_i^{*}\eta_i^{*} + \sqrt{u_i}\sqrt{v_i}\hat{\sigma}_{\epsilon,\eta}\right).
\end{aligned}
$$

Let $\hat{\rho}_b = h(\hat{\theta}_b)$ with h defined by (4.14). The following theorem states that the distribution of $\sqrt{n}(\hat{\rho}-\rho)$ can be approximated by its modified bootstrap counterpart $\sqrt{n}(\hat{\rho}_b - E^{*}\left(\hat{\rho}\right))$.

Theorem 4.4 *Suppose that the conditions of Theorem 4.1 are satisfied. Then, con-*

ditioning on D, as $n \to \infty$,

$$\sigma_b^{-1}\sqrt{n}(\hat{\rho}_b - E^*(\hat{\rho})) \xrightarrow{d} N(0,1), \tag{4.23}$$

where $\sigma_b^2 \xrightarrow{P} \sigma^2$ with σ^2 defined in Theorem 4.2. Also $\sqrt{n}\left(E^*(\hat{\rho}) - \hat{\rho}\right) \xrightarrow{d} N(0, c_0'\Sigma_0 c_0)$ where c_0 and Σ_0 are respectively defined by (4.17) and in Theorem 4.3.

Using this result, a $(1-\alpha)$-level bootstrap percentile confidence interval on ρ is given by

$$\left(\hat{\rho} - P_{1-\alpha/2},\ \hat{\rho} - P_{\alpha/2}\right), \tag{4.24}$$

where P_α is the 100α percentile of the sampling distribution of $(\hat{\rho}_b - E^*(\hat{\rho}))$, that is $P_\alpha \simeq \hat{\rho}_{[\alpha B]} - E^*(\hat{\rho})$, where $\hat{\rho}_{[1]} \leq ... \leq \hat{\rho}_{[B]}$ is the sequence of ordered $\hat{\rho}_b's$ obtained by drawing bootstrap samples, $b = 1, ..., B$.

4.3 Empirical Likelihood

4.3.1 Ordinary Confidence Intervals

Owen (2001) has shown that empirical likelihood ratio statistics for parameters based on i.i.d. data, including multidimensional estimators, have limiting chi-square distribution under mild conditions. In the previous chapters, we used this result to obtain bootstrap confidence intervals for the univariate population mean and distribution function in the presence of missing data. Qin and Lawless (1994) formed empirical likelihood ratio test statistic for obtaining confidence limits for fully observed i.i.d. data through linking estimating equations and empirical likelihood. In this section, we will further extend this theory to form empirical likelihood confidence intervals for the correlation coefficient ρ under joint regression imputation.

We begin by introducing the framework. Let us denote

$$\theta^{(1)} = (\theta_1, \theta_2, \theta_3, \theta_4)' \text{ and } \phi = (\theta^{(1)}, \rho)' \ , \tag{4.25}$$

with components θ_i defined by (4.8) and ρ as per (4.1). Since we are interested in constructing the confidence intervals for ρ, the components of $\theta^{(1)}$ represent nuisance parameters (Owen, 2001).

Let

$$w_i = (\tilde{y}_i, \tilde{z}_i, \tilde{y}_i^2, \tilde{z}_i^2, \tilde{y}_i \tilde{z}_i)', \tag{4.26}$$

be a vector based on imputed data after applying the joint regression imputation as described in section 4.1.2. From Theorem 4.2, we see that $n^{-1} \sum_{i=1}^n w_i$ is a consistent estimator of $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$. Since $\theta_5 = \theta_1\theta_2 + \rho\sqrt{\left(\theta_3 - \theta_1^2\right)\left(\theta_4 - \theta_2^2\right)}$, it follows that $n^{-1} \sum_{i=1}^n w_i$ is a consistent estimator of $(\theta^{(1)}, \theta_1\theta_2 + \rho\sqrt{\left(\theta_3 - \theta_1^2\right)\left(\theta_4 - \theta_2^2\right)})'$, which enables us to define the following empirical likelihood function

$$L = \prod_i p_i,$$

where p_i, $i = 1, ..., n$, are subject to restrictions

$$p_i \geq 0, \ \sum_i p_i = 1, \ \sum_i p_i g(w_i, \phi) = 0,$$

with

$$g(w_i, \phi) = \left(\tilde{y}_i - \theta_1, \tilde{z}_i - \theta_2, \tilde{y}_i^2 - \theta_3, \tilde{z}_i^2 - \theta_4, \tilde{y}_i\tilde{z}_i - \theta_1\theta_2\rho\left(\left(\theta_3 - \theta_1^2\right)\left(\theta_4 - \theta_2^2\right)\right)^{1/2}\right)', \tag{4.27}$$

$i = 1, ..., n$. Following Qin and Lawless (1994), we obtain the following empirical log-likelihood ratio

$$l(\phi) = \sum_i \log\left\{1 + t'(\phi)g(w_i, \phi)\right\}, \tag{4.28}$$

where $t(\phi)$ satisfies

$$\frac{1}{n}\sum_i \frac{g(w_i, \phi)}{1 + t'(\phi)g(w_i, \phi)} = 0.$$

The empirical likelihood ratio statistic for ρ is given by

$$W(\rho) = 2l(\tilde{\theta}^{(1)}, \rho) - 2l(\tilde{\phi}), \tag{4.29}$$

where $\tilde{\theta}^{(1)}$ minimizes $l(\theta^{(1)}, \rho)$ with respect to $\theta^{(1)}$ for fixed ρ, and $\tilde{\phi} = (\tilde{\theta}^{(1)}, \tilde{\rho})'$ minimizes $l(\phi)$ with respect to ϕ. The asymptotic distribution of $W(\rho)$ is given in Theorem 4.5.

Theorem 4.5 *Suppose that the conditions in Theorem 4.2 are satisfied, then as* $n \to \infty$, *there exists a constant* $\omega_1 > 0$ *such that*

$$W(\rho) \xrightarrow{d} \omega_1\chi_1^2.$$

Using this result, the usual $(1 - \alpha)$-level empirical likelihood ratio confidence interval on ρ is given by

$$\left\{\rho | W(\rho) \leq \omega_1 \chi^2_{\alpha,1}\right\}, \quad (4.30)$$

where $\chi^2_{\alpha,1}$ is the upper α quantile of the chi-square distribution with one degree of freedom. Since the scaled coefficient ω_1 is too complicated to estimate, we use modified bootstrap method to approximate the asymptotic distribution of $W(\rho)$.

4.3.2 Bootstrap Calibrated Confidence Intervals

Let

$$w_{i,b} = (\tilde{y}_{i,b}, \tilde{z}_{i,b}, \tilde{y}^2_{i,b}, \tilde{z}^2_{i,b}, \tilde{y}_{i,b}\tilde{z}_{i,b})',$$

be a vector based on bootstrap imputed data as in (4.19) after applying the joint regression imputation in bootstrap procedure, and let

$$g_b(w_{i,b}, \phi) = g(w_{i,b}, \phi) - \mu_{n,b}, \;\; i = 1, ..., n. \quad (4.31)$$

with the correction $\mu_{n,b}$ given by (4.22). That is

$$\begin{aligned} g_b(w_{i,b}, \phi) \;\; &= \;\; \left(\tilde{y}_{i,b} - \theta_1, \tilde{z}_{i,b} - \theta_2, \tilde{y}^2_{i,b} - \theta_3, \tilde{z}^2_{i,b} - \theta_4, \tilde{y}_{i,b}\tilde{z}_{i,b} + \right. \\ & \left. -\theta_1\theta_2\rho \left(\left(\theta_3 - \theta_1^2\right)\left(\theta_4 - \theta_2^2\right)\right)^{1/2}\right)' - \mu_{n,b}. \end{aligned}$$

The modified bootstrap version of the empirical log-likelihood ratio in (4.28) is thus

$$l_b(\phi) = \sum_i \log\left\{1 + t_b'(\phi) g_b(w_{i,b}, \phi)\right\}, \tag{4.32}$$

where $t_b(\phi)$ satisfies

$$\frac{1}{n}\sum_i \frac{g_b(w_{i,b}, \phi)}{1 + t_b'(\phi) g_b(w_{i,b}, \phi)} = 0.$$

We now define

$$W_b(\rho) = 2l_b(\tilde{\theta}_b^{(1)}, \rho) - 2l_b(\tilde{\phi}_b), \tag{4.33}$$

where $\tilde{\theta}_b^{(1)}$ minimizes $l_b(\theta^{(1)}, \rho)$ with respect to $\theta^{(1)}$ for fixed ρ, and $\tilde{\phi}_b$ minimizes $l_b(\phi)$ with respect to ϕ.

Theorem 4.6 below states that the distribution of $W(\rho)$ can be approximated by its modified bootstrap counterpart $W_b(\hat{\rho})$, where $\hat{\rho}$ is given by (4.15).

Theorem 4.6 *Suppose that the conditions in Theorem 2.1 are satisfied. Then, conditioning on D, as $n \to \infty$,*

$$W_b(\hat{\rho})/\omega_{1b} \xrightarrow{d} \chi_1^2,$$

where $\omega_{1b} \xrightarrow{P} \omega_1$ and ω_1 is defined in Theorem 4.5.

A $(1-\alpha)$-level bootstrap empirical likelihood ratio confidence interval on ρ is given by

$$\left\{\rho | W(\rho) \leq W_{1-\alpha}\right\}, \tag{4.34}$$

where $W_{1-\alpha}$ is the $100\,(1-\alpha)\,\%$ sample percentile of $W_1(\hat{\rho}), ..., W_B(\hat{\rho})$.

4.4 Simulation Study

A simulation study was conducted to examine the performance of the proposed adjusted bootstrap CIs for the correlation coefficient ρ. In particular, we compared the performance of the proposed adjusted bootstrap confidence intervals versus their ordinary (unadjusted) counterparts based on three methods: the bootstrap percentile (BP), its corresponding Z-transformed version (ZPB), and the empirical likelihood (EL). The confidence intervals were examined in terms of their coverage probabilities and their average lengths. We obtained simulations results for samples of size $n = 100$ and $n = 200$, based on 1000 simulation processes each with $B = 2000$ bootstrap samples. The standard errors for simulated coverage of the 95% confidence intervals were approximately 0.014 with 1000 simulation runs. The simulations were programmed in R/S-PLUS.

4.4.1 Data Frame

The data frame was based on the simulation study presented in Shao and Wang (2002). The univariate x was generated from the standard exponential distribution. The variables y and z were generated according to

$$y_i = \beta' x_i + v_i^{1/2} \epsilon_i \text{ and } z_i = \gamma' x_i + u_i^{1/2} \eta_i,$$

for $i = 1, ..., n$ with $\beta = 1$, $\gamma = 1$ and $v_i = u_i = x_i$. The error terms ϵ and η generated independently according to

$$\epsilon_i = \kappa\sigma_i + \delta_i \text{ and } \eta_i = \kappa\sigma_i + \tau_i,$$

with σ_i, δ_i and τ_i all i.i.d. $N(0,1)$ and κ=1. Under this setup, $\rho = 2/3$.

We assumed that (y, z) was missing at random (MAR) given x, that is:

$$P((a,b) = (k,l)|y,z,x) = P((a,b) = (k,l)|x), \text{ for any } k,l = 0,1$$

and that a_i and b_i were independent, and used

$$P(y_i \text{ observed } |x_i) = P(a_i = 1|x_i) = \frac{e^{(t_1+t_2x_i)}}{1+e^{(t_1+t_2x_i)}},$$

and

$$P(z_i \text{ observed } |x_i) = P(b_i = 1|x_i) = \frac{e^{(s_1+s_2x_i)}}{1+e^{(s_1+s_2x_i)}},$$

with constants t_1, t_2, s_1, s_2 set to produce particular response rates for y_i and z_i.

In our simulations, the precision of comparisons among the same test procedures at different settings was achieved by re-using the values of input random numbers, in the sense that the results were correlated by having common observations for each of the simulation runs.

4.4.2 Confidence Intervals

The ordinary versions of the confidence intervals were obtained by ignoring the proposed adjustment $\mu_{n,b}$. In particular, under the bootstrap percentile approach, the 95% confidence intervals were given by $(\hat{\rho} - P_{0.975},\ \hat{\rho} - P_{0.025})$, where P_α was the 100α percentile of the sampling distribution of $(\hat{\rho}_b - E^*(\hat{\rho}))$ for the adjusted method; while for the ordinary method, P_α was based on the sampling distribution of $(\hat{\rho}_b - \hat{\rho})$. Similar approach was used to obtain the Z-transformed analogues of the ordinary BP confidence intervals: to construct the ordinary interval, we used $\hat{\rho}_{Z,b}^{ord} = Z(h(\hat{\theta}_b))$ with h given by (4.14) and Fisher Z-transformation $Z(\rho) := 0.5 \log((1-\rho)/(1+\rho))$; while the proposed adjusted Z-transformed BP interval was based on $\hat{\rho}_{Z,b}^{adj} = Z(h(\hat{\theta}_b)) - \mu_{Z\rho,n,b}$ with the adjustment factor $\mu_{Z\rho,n,b}$ calculated using the chain rule

$$\mu_{Z\rho,n,b} = \left(\left(\left.\frac{\partial h(\theta)}{\partial \theta_1}\right|_{\theta=\hat{\theta}}, \ldots, \left.\frac{\partial h(\theta)}{\partial \theta_5}\right|_{\theta=\hat{\theta}} \right) \left(1 - \hat{\rho}^2\right)^{-1} \right)' \mu_{n,b}.$$

Finally, to find the lower and upper bounds of the 95% EL confidence intervals, we followed the algorithm proposed by Wu (2005). The required cut-off values for the χ_1^2 distribution were based on the 95% bootstrap percentiles of $W_b(\hat{\rho})$, as defined by equations (4.31)-(4.33) for the adjusted EL method; while for the ordinary EL method, $g_b(w_{i,b}, \phi) = g(w_{i,b}, \phi)$ in equation (4.31) since $E^*(\hat{\rho}) - \hat{\rho} = 0$.

4.4.3 Results

Tables 1, 2 and 3 show the simulation results for the 95% confidence intervals on ρ for different response probabilities. We begin by looking at the case of fully-observed data. Under full response, the adjustment factor vanishes, hence there were no differences between the ordinary and modified methods. We note that, with fully observed data, the EL outperformed the BP and ZBP methods in terms of the simulated coverage probabilities for all sample sizes considered. For BP and its Z-transformed version, coverage probability improved as the sample size got larger. However, the EL method always resulted in coverage probability that was closest to the nominal value compared to the BP and ZBP methods. In terms of the average interval length, both the BP and ZBP methods performed better compared to the EL method. The average lengths of the EL confidence intervals improved and were close to the length of the BP (and ZBP) confidence intervals when the larger sample size ($n = 200$) was used. Overall, we can conclude that the EL method performed very well under full response.

When dealing with missing values, the EL outperformed the BP and ZBP methods in terms of the simulated coverage probabilities for all sample sizes considered. We observe that the ordinary empirical likelihood method resulted in high simulated coverage probabilities that were almost as, or even above, the nominal level of 95% (recall that we observed similar overcoverage of the ordinary EL confidence intervals for the mean and distribution function, under fractional imputation, in simulation studies presented in Chapters 2 and 3). This tendency, seemed to be corrected by the adjusted version of the EL intervals, which had coverage close to the nominal

95%. Similarly for BP and ZBP intervals, their adjusted versions resulted in better coverage compared to the ordinary intervals. In terms of the average interval length, the BP confidence intervals were on average shorter compared to the average length of the EL (both ordinary and adjusted). The differences however became smaller when larger sample size was considered. The adjusted EL method resulted in shorter confidence intervals compared to the ordinary EL method. In general, the adjusted EL method had best coverage probabilities for all scenarios considered when dealing with missing data. The average lengths of the adjusted EL confidence intervals were close to the corresponding lengths of the BP and ZBP intervals only when large samples were considered. In every simulation scenario, the adjusted methods led to coverage probabilities that were closer to the nominal 95% compared their ordinary counterparts.

Table 4.1: Bootstrap confidence interval coverage probability and average interval length for the correlation coefficient ρ under full response with sample size n and B=2000 bootstrap repetitions.

	Coverage (%)			Average Length		
n	BP	ZBP	EL	BP	ZBP	EL
100	84.8	90.6	94.1	0.218	0.218	0.244
200	90.8	92.2	94.9	0.218	0.219	0.244

Table 4.2: Bootstrap confidence interval coverage probability for the correlation coefficient ρ under joint regression imputation with sample size n, response probabilities (p_a, p_b) and B=2000 bootstrap repetitions.

(p_a, p_b)	n	OrdBP	AdjBP	OrdZBP	AdjZBP	OrdEL	AdjEL
(.62,.62)	100	77.5	84.9	81.8	89.8	95.5	91.6
	200	80.6	88.5	84.9	90.8	96.7	94.0
(.81,.78)	100	82.3	83.2	86.8	89.0	94.4	93.7
	200	87.8	88.9	90.0	91.2	95.6	94.6

Table 4.3: Average interval length for the correlation coefficient ρ with response probabilities (p_a, p_b), sample size n and B=2000 bootstrap repetitions.

(p_a, p_b)	n	OrdBP	AdjBP	OrdZBP	AdjZBP	OrdEL	AdjEL
(.62,.62)	100	0.392	0.392	0.380	0.388	0.476	0.440
	200	0.276	0.276	0.272	0.274	0.329	0.303
(.81,.78)	100	0.317	0.317	0.315	0.316	0.377	0.370
	200	0.227	0.227	0.226	0.226	0.260	0.254

4.5 Conclusions

In this chapter, we proposed asymptotically correct adjusted bootstrap percentile and empirical likelihood confidence intervals on the correlation coefficient ρ under joint regression imputation. We constructed the adjusted confidence intervals based on the bootstrap data obtained by imitating the process of imputing the original data set in bootstrap resampling. Our simulation study demonstrated that the proposed adjusted method leads to better coverage and improved length of confidence intervals for ρ under joint regression imputation.

4.6 Appendix: Proofs

The proofs are based on the theorems and results stated in the appendix to Chapter 2 (Section 2.6.1) and the following theorem.

Theorem 4.7 (Serfling) *(Serfling, 1980).*

Suppose that the k-dimensional vector X_n *is asymptotically* $N_k(\mu, b_n^2\Sigma)$ *with* Σ *a covariance matrix and* $b_n \to 0$*. Let* $f(x) = (f_1(x)..., f_m(x))$*,* $x = (x_1, ..., x_k)$*, be a vector-valued function for which each component* $f_i(x)$ *is real-valued and has non-zero differential* $f_i(\mu, t)$*,* $t = (t_1, ..., t_k)$*, at* $x = \mu$*. Put* $D = [\partial f_i/\partial x_j|_{x=\mu}]_{m\times k}$*. Then* $f(X_n)$ *is asymptotically* $N_m(f(\mu), b_n^2 D\Sigma D')$*.*

4.6.1 Proof of Theorem 4.1.

We use E^* to denote the probability and expectation with respect to randomness in the imputation procedure. We can decompose $(\hat{\theta} - \theta)$ as follows

$$\sqrt{n}(\hat{\theta} - \theta) = U_n + V_n$$

where

$$V_n = \sqrt{n}\left(E^*(\hat{\theta}) - \theta\right)$$

and

$$U_n = \sqrt{n}\left(\hat{\theta} - E^*(\hat{\theta})\right).$$

Denote $\sigma_\epsilon^2 = var(\epsilon)$, $\sigma_\eta^2 = var(\eta)$, $\sigma_{\epsilon,\eta} = cov(\epsilon,\eta)$, $S_{na} = \frac{1}{n}\sum_{i=1}^n a_i x_i x_i'/v_i$, $T_{na} = \frac{1}{n}\sum_{i=1}^n a_i x_i \epsilon_i / v_i^{1/2}$, $S_{nb} = \frac{1}{n}\sum_{i=1}^n b_i x_i x_i'/u_i$, $T_{nb} = \frac{1}{n}\sum_{i=1}^n b_i x_i \eta_i / u_i^{1/2}$ where $v_i = v(x_i)$ and $u_i = u(x_i)$. It can be shown that $\hat{\beta}_r = \beta + S_{na}^{-1}T_{na}$ and $y_i - \hat{\beta}_r x_i = v_i^{1/2}\epsilon_i - T_{na}' S_{na}^{-1} x_i$. Similarly, $\hat{\gamma} = \gamma + S_{nb}^{-1}T_{nb}$ and $z_i - \hat{\gamma}_r x_i = u_i^{1/2}\eta_i - T_{nb}' S_{nb}^{-1} x_i$. Also, we note that $\hat{\sigma}_\epsilon^2 = \sigma_\epsilon^2 + o_p(1)$, $\hat{\sigma}_\eta^2 = \sigma_\eta^2 + o_p(1)$,and $\hat{\sigma}_{\epsilon,\eta} = \sigma_{\epsilon,\eta} + o_p(1)$.

Using the above properties, and under the assumption that $\tilde{\epsilon}_i^{*\prime}$s, $\tilde{\eta}_i^{*\prime}$s and $(\epsilon_i^*, \eta_i^*)'$s are all generated independently from populations with zero mean, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n E^*(\tilde{y}_i) &= \frac{1}{n}\sum_{i=1}^n a_i y_i + \frac{1}{n}\sum_{i=1}^n (1-a_i) b_i E^*\left(y_i^*\right) + \frac{1}{n}\sum_{i=1}^n (1-a_i)(1-b_i) E^*\left(y_i^*\right) \\
&= \frac{1}{n}\sum_{i=1}^n a_i y_i + \frac{1}{n}\sum_{i=1}^n (1-a_i) b_i E^*\left\{\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i) + v_i^{1/2}\tilde{\epsilon}_i^*\right\} + \\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-a_i)(1-b_i) E^*\left\{\left(\hat{\beta}_r' x_i\right) + v_i^{1/2}\epsilon_i^*\right\} \\
&= \frac{1}{n}\sum_{i=1}^n a_i y_i + \frac{1}{n}\sum_{i=1}^n (1-a_i) b_i \hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i) + \\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-a_i)(1-b_i)\left(\hat{\beta}_r' x_i\right) \\
&= \frac{1}{n}\sum_{i=1}^n a_i y_i + \frac{1}{n}\sum_{i=1}^n (1-a_i) b_i \left\{\frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\eta^2}\eta_i - x_i' \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2} S_{nb}^{-1} T_{nb}\right\} + \\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-a_i)\left\{\beta' x_i + x_i' S_{na}^{-1} T_{na}\right\},
\end{aligned}$$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E^{*}(\tilde{y}_i^2) &= \frac{1}{n}\sum_{i=1}^{n} a_i y_i^2 + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i E^{*}\left(y_i^{*2}\right) + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^{*}\left(y_i^{*2}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} a_i y_i^2 + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i E^{*}\left[\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)^2 + \right. \\
&\quad \left. +2\left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)\left(v_i^{1/2}\tilde{\epsilon}_i^{*}\right) + v_i \tilde{\epsilon}_i^{*2}\right] + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^{*}\left[\left(\hat{\beta}_r' x_i\right)^2 + 2\left(\hat{\beta}_r' x_i\right)\left(v_i^{1/2}\epsilon_i^{*}\right) + v_i \epsilon_i^{*2}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} a_i y_i^2 + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i \left(\left(\hat{\beta}_r' x_i\right)^2 + 2\left(\hat{\beta}_r' x_i\right)\left(\frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right) + \right. \\
&\quad \left. + \left(\frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right)^2 + v_i Var^{*}(\tilde{\epsilon}_i^{*})\right) + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left\{\left(\hat{\beta}_r' x_i\right)^2 + v_i Var^{*}(\epsilon_i^{*})\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n} a_i y_i^2 + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i \left(\left(\hat{\beta}_r' x_i\right)^2 + 2\left(\hat{\beta}_r' x_i\right)\left(\frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}(z_i - \hat{\gamma}_r' x_i)\right) + \right. \\
&\quad \left. + \frac{v_i \hat{\sigma}_{\epsilon,\eta}^2}{u_i \hat{\sigma}_\eta^4}(z_i - \hat{\gamma}_r' x_i)^2 + v_i\left(\hat{\sigma}_\epsilon^2 - \frac{\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\eta^2}\right)\right) + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left\{\left(\hat{\beta}_r' x_i\right)^2 + v_i \hat{\sigma}_\epsilon^2\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n} a_i y_i^2 + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)\left\{\left(\beta' x_i + x_i' S_{na}^{-1} T_{na}\right)^2 + v_i \hat{\sigma}_\epsilon^2\right\} + \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i \frac{\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\eta^2}\left(2\left(\beta' x_i + x_i' S_{na}^{-1} T_{na}\right)\left(\eta_i v_i^{1/2} - T_{nb}' S_{nb}^{-1} x_i v_i^{1/2} u_i^{-1/2}\right) + \right. \\
&\quad \left. + v_i \frac{\hat{\sigma}_{\epsilon,\eta}}{u_i \hat{\sigma}_\eta^2}(u_i^{1/2}\eta_i - T_{nb}' S_{nb}^{-1} x_i)^2 - 1\right).
\end{aligned}$$

and similarly, we can obtain the expressions for $\frac{1}{n}\sum_{i=1}^{n} E^{*}(\tilde{z}_i)$ and $\frac{1}{n}\sum_{i=1}^{n} E^{*}(\tilde{z}_i^2)$.

Finally,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E^*(\tilde{y}_i\tilde{z}_i) &= \frac{1}{n}\sum_{i=1}^{n} a_i b_i y_i z_i + \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i E^*\left(y_i^*\right) z_i + \frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i)y_i E^*\left(z_i^*\right) + \\
&+\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left(y_i^* z_i^*\right).
\end{aligned}
$$

The second component of the above summation can be written as

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i E^*\left(y_i^*\right) z_i &= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i z_i E^*\left\{\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}\left(z_i - \hat{\gamma}_r x_i\right) + v_i^{1/2}\tilde{\epsilon}_i^*\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i z_i \left(\hat{\beta}_r' x_i + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2}\left(z_i - \hat{\gamma}_r x_i\right)\right).
\end{aligned}
$$

Similarly, the third component is

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i)y_i E^*\left(z_i^*\right) &= \frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i)y_i E^*\left\{\hat{\gamma}_r' x_i + \frac{u_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{v_i^{1/2}\hat{\sigma}_\epsilon^2}\left(y_i - \hat{\beta}_r x_i\right) + u_i^{1/2}\tilde{\eta}_i^*\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i)y_i \left(\hat{\gamma}_r' x_i + \frac{u_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{v_i^{1/2}\hat{\sigma}_\epsilon^2}\left(y_i - \hat{\beta}_r x_i\right)\right),
\end{aligned}
$$

and finally, for the last component, we have

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left(y_i^* z_i^*\right) &= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)E^*\left\{\left(\hat{\beta}_r' x_i\right)\left(\hat{\gamma}_r' x_i\right) + \hat{\beta}_r' x_i u_i^{1/2}\eta_i^* + \right. \\
&\left. + \hat{\gamma}_r' x_i v_i^{1/2}\epsilon_i^* + u_i^{1/2} v_i^{1/2}\eta_i^*\epsilon_i^*\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\left(\hat{\beta}_r' x_i\right)\left(\hat{\gamma}_r' x_i\right) + u_i^{1/2} v_i^{1/2} Cov^*(\eta_i^*, \epsilon_i^*)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\left(\hat{\beta}_r' x_i\right)\left(\hat{\gamma}_r' x_i\right) + u_i^{1/2} v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}\right).
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E^*(\tilde{y}_i\tilde{z}_i) \;=\; & \frac{1}{n}\sum_{i=1}^{n} a_i b_i y_i z_i + \\
& +\frac{1}{n}\sum_{i=1}^{n}(1-a_i)b_i z_i \left\{\beta' x_i + x_i' S_{na}^{-1} T_{na} + \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\eta^2}\eta_i - x_i' \frac{v_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{u_i^{1/2}\hat{\sigma}_\eta^2} S_{nb}^{-1} T_{nb}\right\} + \\
& +\frac{1}{n}\sum_{i=1}^{n} a_i(1-b_i) y_i \left\{\gamma' x_i + x_i' S_{nb}^{-1} T_{nb} + \frac{u_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{\hat{\sigma}_\epsilon^2}\epsilon_i - x_i' \frac{u_i^{1/2}\hat{\sigma}_{\epsilon,\eta}}{v_i^{1/2}\hat{\sigma}_\epsilon^2} S_{na}^{-1} T_{na}\right\} + \\
& +\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i)\left(\beta' x_i + x_i' S_{na}^{-1} T_{na}\right)\left(\gamma' x_i + x_i' S_{nb}^{-1} T_{nb}\right) + \\
& + \left(\frac{1}{n}\sum_{i=1}^{n}(1-a_i)(1-b_i) u_i^{1/2} v_i^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} a_i b_i\right)^{-1}\right) \times \\
& \times \left(\frac{1}{n}\sum_{i=1}^{n} a_i b_i u^{-1/2}(x_i)\left(v_i^{-1/2}\epsilon_i - x_i' S_{na}^{-1} T_{na}\right)\left(u_i^{-1/2}\eta_i - x_i' S_{nb}^{-1} T_{nb}\right)\right),
\end{aligned}
\tag{4.35}
$$

so that all $E^*(\hat{\theta}_1), ..., E^*(\hat{\theta}_5)$ are functions of sums of independent random variables.

In Section 4.2, we showed that the random regression imputation is model-unbiased for the marginal first and second moments as well as for the cross-product moment. Based on this result, we can say that there exist functions f_i with

$$E[f_j(a_i, b_i, x_i, y_i, \epsilon_i, \eta_i)] = 0,\ \ j = 1, ..., 5,$$

such that

$$E^*(\hat{\theta}_j) - \theta_j = \frac{1}{n}\sum_{i=1}^{n} f_j(a_i, b_i, x_i, y_i, \epsilon_i, \eta_i) + o_p(n^{-1/2}).$$

Thus, by CLT, there exists $\Sigma_1 > 0$ such that

$$\sqrt{n}\left(E^*\left(\hat{\theta}\right) - \theta\right) \xrightarrow{d} N(0, \Sigma_1), \tag{4.36}$$

that is $V_n \xrightarrow{d} N(0, \Sigma_1)$.

Next, by CLT, there exists $\Sigma_{2n} = \Sigma_2 + o_p(1)$ such that

$$\sqrt{n}\left(\hat{\theta} - E^*(\hat{\theta})\right) \xrightarrow{d} N(0, \Sigma_{2n}), \tag{4.37}$$

or $U_n \xrightarrow{d} N(0, \Sigma_{2n})$ and by Polya Theorem, $\sup_t |P(\Sigma_{2n}^{-1} U_n \leq t) - \Phi(t)| = o_p(1)$. So by Chen and Rao Theorem, we have $\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, \Sigma_1 + \Sigma_2)$. ■

4.6.2 Proof of Theorem 4.2.

Theorem 4.1 states that there exists a 5×5 matrix $\Sigma > 0$ such that, as $n \to \infty$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma). \tag{4.38}$$

Now, $\hat{\rho} = h(\hat{\theta})$ with the formula for the function h given by 4.14, and therefore by Serfling's Theorem 4.7, $h(\hat{\theta})$ is asymptotically $N(0, c_0' \Sigma c_0)$ where $c_0 = (\partial h(\theta)/\partial\theta_1, ..., \partial h(\theta)/\partial\theta_5)'$.

■

4.6.3 Proof of Theorem 4.3.

We use P_b and E_b to denote the conditional probability and expectation given D, and P_b^* and E_b^* to denote the probability and expectation with respect to the randomness in the bootstrap imputation procedure. We use the same routine as the proof of Theorem 4.1 to prove Theorem 4.3. Denote $S_{na,b} = n^{-1}\sum_{i=1}^n a_{i,b}x_{i,b}x'_{i,b}/v_{i,b}$, $T_{na,b} = n^{-1}\sum_{i=1}^n a_{i,b}x_{i,b}\epsilon_{i,b}/v_{i,b}^{1/2}$, $S_{nb,b} = n^{-1}\sum_{i=1}^n b_{i,b}x_{i,b}x_{i,b}/u_{i,b}$, $T_{nb,b} = n^{-1}\sum_{i=1}^n b_{i,b}x_{i,b}\eta_{i,b}/u_{i,b}^{1/2}$. It can be shown that $\hat\beta_{r,b} = \beta + S_{na,b}^{-1}T_{na,b}$, and $y_{i,b} - \hat\beta'_{r,b}x_{i,b} = v_{i,b}^{1/2}\epsilon_{i,b} - T'_{na,b}S_{na,b}^{-1}x_{i,b}$. Similarly, $\hat\gamma_{r,b} = \gamma + S_{nb,b}^{-1}T_{nb,b}$, and $z_{i,b} - \hat\gamma'_{r,b}x_{i,b} = u_{i,b}^{1/2}\eta_{i,b} - T'_{nb,b}S_{nb,b}^{-1}x_{i,b}$. Also $\hat\sigma^2_{\epsilon,b} = \hat\sigma^2_\epsilon + o_{p_b}(1)$, $\hat\sigma^2_{\eta,b} = \hat\sigma^2_\eta + o_{p_b}(1)$, $\hat\sigma_{\epsilon\eta,b} = \hat\sigma_{\epsilon\eta} + o_{p_b}(1)$.

Let us first look at $\hat\theta_{5,b}$, the fifth component of $\hat\theta_b$. Using the above properties and following the proof of Theorem 4.1, we can show that $n^{-1}\sum_{i=1}^n E_b^*(\tilde y_{i,b}\tilde z_{i,b})$ is a function of sums of independent random variables with respect to P_b. Thus for some integer r, we can write

$$E_b^*(\hat\theta_{5,b}) = f_{5,b}\left(G_{1,b}, ..., G_{r,b}\right), \tag{4.39}$$

where $G_{j,b}$, $j = 1, .., r$, are sums of independent random variables with respect to P_b. Applying Taylor series expansion, we can see that

$$\begin{aligned} f_{5,b}\left(G_{1,b}, ..., G_{r,b}\right) &= f_{5,b}\left(G_1, ..., G_r\right) \\ &+ \left(\left.\frac{\partial f_{5,b}(u_1, ..., u_r)}{\partial u_1}\right|_{u_j=G_j}, ..., \left.\frac{\partial f_{5,b}(u_1, ..., u_r)}{\partial u_r}\right|_{u_j=G_j} \right)' \\ &\times \left(G_{1,b} - G_1, ..., G_{r,b} - G_r\right) + o_{p_b}(n^{-1/2}), \end{aligned} \tag{4.40}$$

where $G_j := E_b(G_{j,b})$, $j = 1, ..., r$. We can show that

$$f_{5,b}\left(G_1, ..., G_r\right) = E^*\left(\hat\theta_5\right). \tag{4.41}$$

Similar expressions can be obtained for $\hat\theta_{j,b}$, $j = 1, ..., 4$ and for $E^*(\hat\theta_{j,b})$, $j = 1, ..., 4$.

Therefore, there exists $\Sigma_{1,b} = \Sigma_1 + o_p(1)$ such that

$$(\Sigma_{1,b})^{-1/2}\sqrt{n}\left(E_b^*\left(\hat{\theta}_b\right) - E^*\left(\hat{\theta}\right)\right) \xrightarrow{d} N(0, I_5). \tag{4.42}$$

By CLT, under P_b^* , there exists $\Sigma_{2n,b} = \Sigma_2 + o_p(1)$ such that

$$(\Sigma_{2n,b})^{-1/2}\sqrt{n}(\hat{\theta}_b - E_b^*\left(\hat{\theta}_b\right)) \xrightarrow{d} N(0, I_5). \tag{4.43}$$

4.6.4 Proof of Theorem 4.4.

Let us recall the vector

$$\mu_{n,b} := E^*(\hat{\theta}) - \hat{\theta} = (\mu_{n,b,1}, \mu_{n,b,2}, \mu_{n,b,3}, \mu_{n,b,4}, \mu_{n,b,5})'. \tag{4.44}$$

Let

$$\mu_{\rho,n,b} = \left(\frac{\partial h(\theta)}{\partial \theta_1}\mid_{\theta=\hat{\theta}}, ..., \frac{\partial h(\theta)}{\partial \theta_5}\mid_{\theta=\hat{\theta}}\right)'\mu_{n,b}, \tag{4.45}$$

where h is given by (4.14) and the derivatives are

$$\begin{aligned}\frac{\partial h(\theta)}{\partial \theta_1} &= -\theta_2\left[(\theta_3 - \theta_1^2)(\theta_4 - \theta_2^2)\right]^{-1/2} \\ &\quad + [\theta_5 - \theta_1\theta_2]\left[(\theta_3 - \theta_1^2)(\theta_4 - \theta_2^2)\right]^{-3/2}\theta_1(\theta_4 - \theta_2^2),\end{aligned}$$

$$\begin{aligned}
\frac{\partial h(\theta)}{\partial \theta_2} &= -\theta_1 \left[(\theta_3 - \theta_1^2)(\theta_4 - \theta_2^2)\right]^{-1/2} \\
&\quad + [\theta_5 - \theta_1\theta_2] \left[(\theta_3 - \theta_1^2)(\theta_4 - \theta_2^2)\right]^{-3/2} \theta_2(\theta_3 - \theta_1^2),
\end{aligned}$$

$$\frac{\partial h(\theta)}{\partial \theta_3} = -\frac{1}{2} [\theta_5 - \theta_1\theta_2] \left[(\theta_3 - \theta_1^2)(\theta_4 - \theta_2^2)\right]^{-3/2} (\theta_4 - \theta_2^2),$$

$$\frac{\partial h(\theta)}{\partial \theta_4} = -\frac{1}{2} [\theta_5 - \theta_1\theta_2] \left[(\theta_3 - \theta_1^2)(\theta_4 - \theta_2^2)\right]^{-3/2} (\theta_3 - \theta_1^2),$$

$$\frac{\partial h(\theta)}{\partial \theta_5} = \left[(\theta_3 - \theta_1^2)(\theta_4 - \theta_2^2)\right]^{-1/2}.$$

Using Taylor's series formula, we can approximate

$$\begin{aligned}
E^*(\hat{\rho}) &\cong h(\hat{\theta}) + \left(\frac{\partial h(\theta)}{\partial \theta_1} |_{\theta=\hat{\theta}}, \ldots, \frac{\partial h(\theta)}{\partial \theta_5} |_{\theta=\hat{\theta}} \right)' \left(E^*(\hat{\theta}) - \hat{\theta} \right) \\
&= \hat{\rho} + \mu_{\rho,n,b}.
\end{aligned}$$

The main result of this theorem can be now verified using Theorem 4.3 and Theorem 4.7. ■

4.6.5 Proof of Theorem 4.5.

Recall that the empirical likelihood ratio statistic for ρ was defined as $W(\rho) = 2l(\tilde{\theta}^{(1)}, \rho) - 2l(\tilde{\phi})$ where $\tilde{\theta}^{(1)}$ minimizes $l(\theta^{(1)}, \rho)$ with respect to $\theta^{(1)}$ for fixed ρ, and $\tilde{\phi}$

minimizes $l(\phi)$ with respect to ϕ. To derive the asymptotic distribution of $W(\rho)$ we use results presented in Qin and Lawless (1994), in particular, proofs of Theorems 1 and 2 and Corollary 5. Let ϕ, $\theta^{(1)}$ and ρ denote the true values. Following Qin and Lawless (1994), we define

$$Q_{1n}(\phi, t) = \frac{1}{n}\sum_i \frac{g(w_i, \phi)}{1 + t'(\phi)g(w_i, \phi)},$$

and

$$Q_{2n}(\phi, t) = \frac{1}{n}\sum_i (1 + t'(\phi)g(w_i, \phi))^{-1}\left(\frac{\partial g(w_i, \phi)}{\partial \phi}\right)' t(\phi).$$

Then $\tilde{\phi}$ and $\tilde{t} = t(\tilde{\phi})$ satisfy $Q_{1n}(\tilde{\phi}, \tilde{t}) = 0$ and $Q_{2n}(\tilde{\phi}, \tilde{t}) = 0$. Further $\tilde{\theta}^{(1)}$ and $\tilde{t} = t(\tilde{\theta}^{(1)}, \rho)$ satisfy $Q_{1n}\left(\left(\tilde{\theta}^{(1)}, \rho\right), \tilde{t}\right) = 0$ and $Q_{2n}\left(\left(\tilde{\theta}^{(1)}, \rho\right), \tilde{t}\right) = 0$.

Taking derivatives about t and ϕ of $Q_{1n}(\phi, 0)$ gives

$$S_{n11} := \frac{\partial Q_{1n}(\phi, 0)}{\partial t'} = -\frac{1}{n}\sum_i g(w_i, \phi)g'(w_i, \phi),$$

and

$$S_{n12} := \frac{\partial Q_{1n}(\phi, 0)}{\partial \phi} = \frac{1}{n}\sum_i \frac{\partial g(w_i, \phi)}{\partial \phi}.$$

We also define

$$S_{n13} := \frac{\partial Q_{2n}(\left(\theta^{(1)}, \rho\right), 0)}{\partial \theta^{(1)}} = \frac{1}{n}\sum_i \frac{\partial g(w_i, \left(\theta^{(1)}, \rho\right))}{\partial \theta^{(1)}}.$$

As given in Qin and Lawless (1994),

$$\begin{aligned} l(\tilde{\phi}) &= \log\left\{1+\tilde{t}'g(w_i,\tilde{\phi})\right\} \\ &= -\frac{n}{2}Q'_{1n}(\phi,0)A_{1n}Q_{1n}(\phi,0)+o_p(1), \end{aligned}$$

where $A_{1n} = S_{n11}^{-1}(I + S_{n12}S_{n22.1}^{-1}S_{n12}^{-1}S_{n11}^{-1})$ with $S_{n22.1}^{-1} = S'_{n12}(S_{n11})^{-1}S_{n12}$. Similarly, for the other component of the empirical likelihood ratio statistic for ρ,

$$\begin{aligned} l(\tilde{\theta}^{(1)},\rho) &= -\frac{n}{2}Q'_{1n}\left(\left(\theta^{(1)},\rho\right),0\right)A_{2n}Q_{1n}\left(\left(\theta^{(1)},\rho\right),0\right)+o_p(1) \\ &= -\frac{n}{2}Q'_{1n}(\phi,0)A_{2n}Q_{1n}(\phi,0)+o_p(1), \end{aligned}$$

where $A_{2n} = S_{n11}^{-1}(I + S_{n13}S_{n33.1}^{-1}S_{n13}^{-1}S_{n11}^{-1})$ with $S_{n33.1}^{-1} = S'_{n13}(S_{n11})^{-1}S_{n13}$. Also note that $A_{1n} - A_{2n} = S_{n11}^{-1}(S_{n12}S_{n22.1}^{-1}S_{n12}^{-1} - S_{n13}S_{n33.1}^{-1}S_{n13}^{-1})S_{n11}^{-1}$.

It can be shown, that there exists S_{11} such that $S_{n11} = S_{11}+o_p(1)$. Let $S_{n12} = S_{12}$ and $S_{n13} = S_{13}$ respectively have the following structures

$$S_{12} = \begin{bmatrix} -I_4 & 0 \\ \\ s_{121} & s_{122} \end{bmatrix} \text{ and } S_{13} = \begin{bmatrix} -I_4 \\ \\ s_{121} \end{bmatrix}.$$

with components

$$s_{121} = \left[-\theta_2+\frac{\rho\theta_1\sigma_z}{\sqrt{\sigma_y\sigma_z}}, -\theta_1+\frac{\rho\theta_2\sigma_y}{\sqrt{\sigma_y\sigma_z}}, -\frac{\rho\sigma z}{2\sqrt{\sigma_y\sigma_z}}, -\frac{\rho\sigma_y}{2\sqrt{\sigma_y\sigma_z}}\right],$$

$$s_{122} = -\sqrt{\sigma_y\sigma_z},$$

where $\sigma_y := (\theta_3 - \theta_1^2)^{1/2}$ and $\sigma_z := (\theta_4 - \theta_2^2)^{1/2}$.

Let $S_{22.1} = S'_{12}(-S_{11})^{-1}S_{12}$, $S_{33.1} = S'_{13}(-S_{11})^{-1}S_{13}$ and define $S = (S_{12}S_{22.1}^{-1}S'_{12} - S_{13}S_{33.1}^{-1}S'_{13})$. Then

$$\begin{aligned}
W(\rho) &= nQ'_{1n}(\phi,0)\,(-S_{11})^{-1}\,S\,(-S_{11})^{-1}\,Q_{1n}(\phi,0) + o_p(1) \\
&= \left(S_{11}^{-1/2}\sqrt{n}Q_{1n}(\phi,0)\right)'\left(S_{11}^{-1/2}SS_{11}^{-1/2}\right)\left(S_{11}^{-1/2}\sqrt{n}Q_{1n}(\phi,0)\right) + o_p(1) \\
&= \left(\Sigma^{-1/2}S_{11}^{-1/2}\sqrt{n}Q_{1n}(\phi,0)\right)'\left(\Sigma^{1/2}S_{11}^{-1/2}SS_{11}^{-1/2}\Sigma^{1/2}\right)\left(\Sigma^{-1/2}S_{11}^{-1/2}\sqrt{n}Q_{1n}(\phi,0)\right) + \\
&\quad + o_p(1).
\end{aligned}$$

Using Theorem 4.1 and the above result, we conclude that $W(\rho)$ is asymptotically a non-standard chi-square variable.

Further, it can be shown that $W(\rho)$ is asymptotically a weighted sum of independent standard chi-square random variables $\sum_{i=1}^{5}\omega_i\chi_{1,i}^2$, where $\omega_1, ..., \omega_5$ are the eigenvalues of $\Sigma^{1/2}S_{11}^{-1/2}SS_{11}^{-1/2}\Sigma^{1/2}$ and $\chi_{1,1}^2, ..., \chi_{1,5}^2$ are independent χ_1^2 random variables (Satorra and Bentler, 1988).

Let n_0 denote the number of non-zero eigenvalues, that is $n_0 = rank(S_{11}^{-1/2}SS_{11}^{-1/2}) = rank(A_1 - A_2)$, where $A_1 = (-S_{11})^{-1/2}S_{12}S_{22.1}^{-1}S'_{12}(-S_{11})^{-1/2}$,

$A_2 = (-S_{11})^{-1/2}S_{13}S_{33.1}^{-1}S'_{13}(-S_{11})^{-1/2}$. Since A_j's are symmetric and idempotent, $rank(A_1) = 5$, and $rank(A_2) = 4$. By the proof of Corollary 5 in Qin and Lawless (1994), $A_1 - A_2 \geq 0$; Thus, there exists an orthogonal matrix P_1 such that $I_5 = P'_1A_1P_1 + P'_1A_2P_1$ (in fact, $A_1 = I_5$ so that we can take $P_1 = I_5$). Further, there exists an orthogonal matrix P_2 such that

$$I_5 = P_2' P_1' A_1 P_1 P_2 + \begin{pmatrix} I_4 & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore, $rank(A_1 - A_2) = rank(P_2' P_1' A_1 P_1 P_2) = 1$, so there is only one non-zero eigenvalue in $\Sigma^{1/2} S_{11}^{-1/2} S S_{11}^{-1/2} \Sigma^{1/2}$. Since $(A_1 - A_2) \geq 0$, $\omega_1 > 0$ and it follows that $W(\rho) \xrightarrow{d} \omega_1 \chi_1^2$ as $n \to \infty$. ■

4.6.6 Proof of Theorem 4.6.

Similarly as in the proof of Theorem 4.5, we define

$$Q_{1bn}(\phi, t_b) = \frac{1}{n} \sum_i \frac{g_b(w_{i,b}, \phi)}{1 + t_b'(\phi) g_b(w_{i,b}, \phi)},$$

and

$$Q_{2bn}(\phi, t_b) = \frac{1}{n} \sum_i (1 + t_b'(\phi) g_b(w_{i,b}, \phi))^{-1} \left(\frac{\partial g_b(w_{i,b}, \phi)}{\partial \phi} \right)' t_b(\phi).$$

Then $\tilde{\phi}_b$ and $\tilde{t}_b = t_b(\tilde{\phi}_b)$ satisfy $Q_{1bn}(\tilde{\phi}_b, \tilde{t}_b) = 0$ and $Q_{2bn}(\tilde{\phi}_b, \tilde{t}_b) = 0$. Further $\tilde{\theta}_b^{(1)}$ and $\tilde{t}_b = t(\tilde{\theta}_b^{(1)}, \hat{\rho})$ satisfy $Q_{1bn}\left(\left(\tilde{\theta}_b^{(1)}, \hat{\rho}\right), \tilde{t}_b\right) = 0$ and $Q_{2bn}\left(\left(\tilde{\theta}_b^{(1)}, \hat{\rho}\right), \tilde{t}_b\right) = 0$.

Taking derivatives about t_b and ϕ of $Q_{1bn}(\phi, 0)$ gives

$$S_{n11} := \frac{\partial Q_{1bn}(\phi, 0)}{\partial t_b'} = -\frac{1}{n} \sum_i g_b(w_{ib}, \phi) g_b'(w_{ib}, \phi),$$

and

$$S_{n12} := \frac{\partial Q_{1bn}(\phi, 0)}{\partial \phi} = \frac{1}{n} \sum_i \frac{\partial g_b(w_{ib}, \phi_b)}{\partial \phi_b}.$$

We also define

$$S_{n13} := \frac{\partial Q_{2bn}\left(\left(\theta_b^{(1)}, \hat{\rho}\right), 0\right)}{\partial \theta_b^{(1)}} = \frac{1}{n}\sum_i \frac{\partial g_b\left(w_{i_b}, \left(\theta_b^{(1)}, \hat{\rho}\right)\right)}{\partial \theta_b^{(1)}}.$$

Let $\hat{\phi} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5)$, following the proof of Theorem 4.5,

$$\begin{aligned} l_b(\tilde{\phi}_b) &= \log\left\{1 + \tilde{t}' g_b(w_{i_b}, \tilde{\phi}_b)\right\} \\ &= -\frac{n}{2} Q'_{1bn}(\hat{\phi}, 0) A_{1n} Q_{1bn}(\hat{\phi}, 0) + o_p(1), \end{aligned}$$

where $A_{1n} = S_{n11}^{-1}(I + S_{n12}S_{n22.1}^{-1}S_{n12}^{-1}S_{n11}^{-1})$ with $S_{n22.1}^{-1} = S'_{n12}(S_{n11})^{-1}S_{n12}$, and

$$\begin{aligned} l(\tilde{\theta}_b^{(1)}, \hat{\rho}) &= -\frac{n}{2} Q'_{1bn}\left(\left(\hat{\theta}^{(1)}, \hat{\rho}\right), 0\right) A_{2n} Q_{1bn}\left(\left(\hat{\theta}^{(1)}, \hat{\rho}\right), 0\right) + o_p(1) \\ &= -\frac{n}{2} Q'_{1bn}(\hat{\phi}, 0) A_{2n} Q_{1bn}(\hat{\phi}, 0) + o_p(1), \end{aligned}$$

where $A_{2n} = S_{n11}^{-1}(I + S_{n13}S_{n33.1}^{-1}S_{n13}^{-1}S_{n11}^{-1})$ with $S_{n33.1}^{-1} = S'_{n13}(S_{n11})^{-1}S_{n13}$.

Let $S_{22.1} = S'_{12}(-S_{11})^{-1}S_{12}$, $S_{33.1} = S'_{13}(-S_{11})^{-1}S_{13}$ and define $S = (S_{12}S_{22.1}^{-1}S'_{12} - S_{13}S_{33.1}^{-1}S'_{13})$ with components defined similarly as in Theorem 4.5. Then we can write

$$\begin{aligned} W_b(\hat{\rho}) &= nQ'_{1bn}(\hat{\phi}, 0)\left(-S_{11}\right)^{-1} S \left(-S_{11}\right)^{-1} Q_{1bn}(\hat{\phi}, 0) + o_p(1) \\ &= \left(S_{11}^{-1/2}\sqrt{n}Q_{1bn}(\hat{\phi}, 0)\right)' \left(S_{11}^{-1/2} S S_{11}^{-1/2}\right) S_{11}^{-1/2}\sqrt{n}Q_{1bn}(\hat{\phi}, 0) + o_p(1), \end{aligned}$$

which, combined with Theorem 4.3, proves Theorem 4.6. ■

Chapter 5

Goodness-of-Fit for Incomplete Longitudinal Binary Data

5.1 Introduction

Logistic regression is often used to model the relationship between longitudinal outcomes and predictor variables; once a model has been fit, its adequacy is examined by a goodness-of-fit test. In the presence of missing data, with incompleteness either due to early withdrawal of some subjects (dropout pattern) or temporary unavailability (intermittent pattern), the existing goodness-of-fit tests provide valid inference under the missing completely at random (MCAR) assumption. However, the restrictive MCAR mechanism does not usually happen with real data. The missing at random (MAR) setting, that allows missingness to depend on observed variables, is much more practical assumption for most longitudinal studies. In this chapter, we describe the approach to estimate parameters of a marginal logistic regression model for longitudinal binary data, with MAR observations, using a model-based approach similar to the weighted generalized estimating equations (WGEE) method of Robins et al. (1995) extended by Preisser et al. (2000). We then present a goodness-of-fit test to assess the adequacy of the fitted model that builds on the score test concept presented by Horton et al. (1999) and can be applied to the case of longitudinal data with MAR observations. We draw a comparison between the proposed goodness-of-fit method, which incorporates the estimation of the missingness model parameters, and the existing ordinary method that ignores the missingness process (Horton et al. 1999).

This chapter is organised as follows. In Section 5.2, we review Liang and Zeger's (1986) concept of generalized estimating equations under the MCAR assumption and outline the weighted generalized estimating equations (WGEE) method to analyse MAR data. In Section 5.3, we describe the score goodness-of-fit test for fully-observed longitudinal data based on the Hosmer and Lemeshow's (1980) approach. The proposed goodness-of-fit method is presented in Section 5.4. A small simulation study to assess the proposed method is described in Section 5.5.

5.2 Parameter Estimation

5.2.1 Notation

Suppose that data consist of time sequences of measurements on several individuals and is of the form $\{(y_{it}, x_{it}), i = 1, 2, ..., n, t = 1, ..., T\}$ where T denotes the, common to all n individuals, set of all observation times. Let $y_i = (y_{i1}, ..., y_{iT})'$ be the vector of responses for individual i, with $y_{it} = 1$ if subject i has the characteristic of interest at time t, $1 \leq t \leq T$, otherwise $y_{it} = 0$. The y_{it}'s are assumed to be correlated within, but independent across, subjects. Let $x_i = (x_{i1}, ..., x_{iT})\prime$ denote the corresponding covariate matrix with x_{it} being a $b \times 1$ covariate vector at time t. To describe the relationship between the binary response y_{it} and the covariates x_{it},we consider a marginal logistic linear regression model

$$\text{logit}(p_{it}) = x_{it}'\beta, \tag{5.1}$$

with Bernoulli marginal density of response $f\left(y_{it}|x_{it}\right) = p_{it}^{y_{it}}\left(1 - p_{it}\right)^{1-y_{it}}$, $E\left(y_{it}|x_{it}, \beta\right) = p_{it}$ and $Var(y_{it}) = p_{it}\left(1 - p_{it}\right)$ where β is a $b \times 1$ vector of unknown regression coefficients, and $p_i = (p_{i1}, ..., p_{iT})'$ is a vector of true event probabilities.

In the case of incomplete data, we mark the observed data by superscript "0", that is, y_i^0, x_i^0 and p_i^0 denote respectively the observed responses, covariates and event probabilities. For example, under the monotone missingness pattern, $y_i^0 = (y_{i1}, ..., y_{iT_i})'$ represent the observed responses, $x_i^0 = (x_{i1}, ..., x_{iT_i})'$ is the corresponding observed covariate matrix, and $p_i^0 = (p_{i1}, ..., p_{iT_i})'$ is the vector of true event probabilities where T_i, $1 \leq T_i \leq T$, denotes the number of times an individual i is observed until a drop-out occurs.

5.2.2 Generalized Estimating Equations

An estimator of the vector of model parameters β can be obtained as a solution to the generalized estimating equations (GEE). GEEs, first introduced by Liang and

Zeger (1986), represent an extension to the generalized linear model to accommodate correlated data. The GEE method yields consistent estimators of the model parameters β, provided that the model for the marginal means for the outcomes is correct, even if the correlation structure is misspecified. However, in the case of incomplete data, this approach is valid only under the strong assumption that data are missing completely at random (MCAR), that is, with nonresponse independent of both observed and unobserved $y_i's$ given x (Liang and Zeger, 1986; Robins et al., 1995).

Generally, in the GEE method, we relate the marginal expectation of responses p_{ij} to a linear combination of the covariates via a known link function $h(p_{it}) = x'_{it}\beta$ and describe the marginal variance of y_{it} as a function of the marginal mean $p_{it}(\beta)$, that is $Var(y_{it}) = v_{it}\phi$ where $v_{it} = v(p_{it})$ is a known function. The scale parameter ϕ is possibly unknown; for binary data $\phi = 1$ and $\phi \geq 1$ for binomial or count data. Commonly used link functions include:

- Identity link: $h(p_{it}) = p_{it}$ and $v_{it} = 1$ for a normally distributed response vector,
- Logit link: $h(p_{it}) = \log(p_{it}/(1 - p_{it}))$ and $v_{it} = p_{it}(1 - p_{it})$ for a binary response,
- Log link: $h(p_{it}) = \log(p_{it})$ and $v_{it} = p_{it}$ for a Poisson response.

Under MCAR with monotone pattern, the estimator of β can be obtained as the solution of the generalized estimating equation

$$\sum_{i=1}^{n} U_i^0 = \sum_{i=1}^{n} D_i^{0\prime} \left(V_i^0\right)^{-1}(y_i^0 - p_i^0) = 0, \tag{5.2}$$

where $D_i^0 = \partial p_i^0 / \partial\beta$ and V_i^0 is the working variance-covariance matrix for y_i^0. We assume that $V_i^0 = A_i^{1/2} H_i(\eta) A_i^{1/2}$ where $A_i = diag(v_{i1}, ..., v_{iT_i})$ and $H_i(\eta)$ is a chosen, not necessarily correctly specified, $T_i \times T_i$ working correlation matrix for each y_i that may depend on a vector of unknown parameters η assumed to be the same for all subjects. The $(t_1, t_2)^{th}$ element of $H_i(\eta)$, $H_{i(t_1,t_2)}(\eta)$, is the known, hypothesized, or estimated correlation between the observations y_{it_1} and y_{it_2} on subject i. Examples of the working correlation structures include

- Independence structure: $H_{i(t_1,t_2)}(\eta) = I$,
- Exchangeable structure with equal correlations : $H_{i(t_1,t_2)}(\eta) = \rho$,
- AR(1) structure: $H_{i(t_1,t_2)}(\eta) = \rho^{|t_1 - t_2|}$, and
- Unspecified (or unstructured): $H_{i(t_1,t_2)}(\eta) = \rho_{t_1 t_2}$.

The estimator of β is obtained by replacing η by a consistent estimator $\hat{\eta}$, and solving (5.2) for $\hat{\beta}$ iteratively. Under mild regularity conditions and MCAR mechanism, Liang and Zeger (1986) showed that $\hat{\beta}$ is a consistent estimator for β and $\sqrt{n}\left(\hat{\beta} - \beta\right)$ is asymptotically multivariate normal with zero mean and covariance matrix $\Sigma = \lim_{n\to\infty} n\, \Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}$ where

$$\Sigma_0 = \sum_{i=1}^{n} D_i^{0\prime}\left(V_i^0\right)^{-1} D_i^0 \text{ and } \Sigma_1 = \sum_{i=1}^{n} D_i^{0\prime}\left(V_i^0\right)^{-1} Cov(Y_i^0)\left(V_i^0\right)^{-1} D_i^0. \tag{5.3}$$

Replacing β and η by their consistent estimators $\hat{\beta}$ and $\hat{\eta}$, and $Cov(Y_i^0)$ by $(y_i^0 - p_i^0)(y_i^0 - p_i^0)'$ gives a consistent sandwich estimator $\hat{\Sigma}$ of Σ, even if the working correlation matrices $H_i(\eta)$ are misspecified.

5.2.3 Weighted Estimating Equations

We now consider the MAR setting in the sense that missingness depends on the observed outcomes but not on the unobserved measurements. Studies indicate that GEE approach may yield biased estimates if data are not MCAR (Laird, 1988; Liang and Zeger, 1986). Robins et al. (1995) extended the GEE method to data with dropouts, under a less rigorous assumption of missing at random (MAR), by introducing a class of weighted estimating equations (WGEE) that result in consistent and asymptotically normal estimators of β. Generally, their approach follows the classical Horvitz-Thompson method since the observations are weighted inversely proportional to their probability of being observed. The weights are obtained from a model for the missing data process and must be correctly specified so that the resulting estimators are consistent.

In a longitudinal study, we distinguish between two patterns for missingness: monotone and intermittent. Monotone missingness, or so-called dropout, means that if an observation is missing then all subsequent observations are also missing for a given individual. on the contrary, intermittent pattern means that either missing, or observed, response may be present at any time for a given individual in the data file.

5.2.3.1 Model for Monotone Missingness

Let R_{it} be the response indicator for y_{it}, that is $R_{it} = 1$ if y_{it} is observed or $R_{it} = 0$ otherwise. We assume that, at $t = 1$, all data are observed ($R_{i1} = 1$) and form a vector of response indicators $R_i = (R_{i1}, R_{i2}, ..., R_{iT})'$ for subject i. Further, we let $\breve{R}_{it} = (R_{i1}, ..., R_{i(t-1)})'$ represent individual's response history prior to time t and define

$$\lambda_{it}(\alpha) = P(R_{it} = 1 | \breve{R}_{it}, y_i^0, x_i^0, \alpha), \tag{5.4}$$

where the superscript 0 denotes observed data, α is a vector of unknown parameters and $\lambda_{i1} = 1$. For the case of data with drop-outs only, individual i is observed up to time T_i and so $\breve{R}_{it}$ can be replaced by $R_{i(t-1)} = 1$ since all $R_{i1}, ..., R_{i(t-2)} = 1$ w.p. 1 for $t = 1, ..., T_i$. That is,

$$\lambda_{it}(\alpha) = P(R_{it} = 1 | R_{i(t-1)} = 1, y_{i1}, ..., y_{i(t-1)}, x_i^0, \alpha), \tag{5.5}$$

with $\lambda_{it}(\alpha)$ taking values in $(0, 1]$.

Let $z_{it}^0 = (y_{it}^0, x_{it}^0)$ represent the observed data for individual $i = 1, ..., n$ at time $t = 1, ..., T_i$. For $t \geq 2$, the estimator $\hat{\lambda}_{it}$ is obtained by fitting a logistic model

$$\text{logit}\{\lambda_{it}(\alpha)\} = z_{it}^0 \alpha, \tag{5.6}$$

and the partial log likelihood for the i^{th} subject is

$$\sum_t R_{i(t-1)} \log\{\lambda_{it}(\alpha)^{R_{it}} (1 - \lambda_{it}(\alpha))^{1-R_{it}}\}. \tag{5.7}$$

Differentiating (5.7) with respect to α gives the i^{th} score component

$$u_{i\alpha}(\alpha) = \sum_t R_{i(t-1)} z_{it}^0 \left(R_{it} - \lambda_{it}(\alpha)\right), \tag{5.8}$$

and the summation over all individuals i results in an estimating equation

$$u_\alpha(\alpha) = \sum_i u_{i\alpha}(\alpha) = 0, \tag{5.9}$$

which we solve for $\widehat{\alpha}$ and consequently, obtain $\hat{\lambda}_{it} = \left(1 + \exp\{-z_{it}^0 \widehat{\alpha}\}\right)^{-1}$.

We define the weights as the inverse of the unconditional probability of an individual i being observed at time t. Thus, the estimator $\hat{w}_{it}$ of the weight w_{it} is

$$\hat{w}_{it} = \prod_{l=1}^{t} \hat{\lambda}_{il}^{-1}. \tag{5.10}$$

5.2.3.2 Model for Intermittent Missingness

The assumption that the missing data pattern is monotone is very restrictive as in practice subjects could come back to a study after a missed visit. To extend the WGEE technique, from the drop-out only assumption, to intermittent missing data, Preisser et al. (2000) used $\lambda_{it}(\alpha) = P(R_{it} = 1 | \breve{R}_{it}, z_i^0, \alpha)$ where $\breve{R}_{it} = \left(R_{i1}, ..., R_{i(t-1)}\right)$ represents individual's response history. Suppose that $P(R_{i1} = 1) = 1$ and that missingness at time t is independent of the response history prior to time $t-1$ that is $P(R_{it} = 1 | \breve{R}_{it}, z_i^0, \alpha) = P(R_{it} = 1 | R_{i(t-1)}, z_i^0, \alpha)$, also define

$$\lambda_{it}^{(r)} = P(R_{it} = 1 | R_{i(t-1)} = r, z_i^0, \alpha), \ r = 0, 1, \ t = 3, ..., T.$$

The marginal probabilities of being observed are given by

$$\begin{aligned} w_{i1}^{-1} &= 1 \\ w_{i2}^{-1} &= \lambda_{i2} \\ w_{it}^{-1} &= \lambda_{it}^{(1)} w_{i(t-1)}^{-1} + \lambda_{it}^{(0)} \left(1 - w_{i(t-1)}^{-1}\right), \ t = 3, ..., T. \end{aligned} \tag{5.11}$$

To obtain an estimator of λ_{it}, we fit a logistic regression model

$$\text{logit}\, \lambda_{it}(\alpha) = v_{it}\alpha, \tag{5.12}$$

where ν_{it} is a function of $\breve{R}_{it}$, y_i^0 and x_i, $t = 2, ..., T$. The parameter vector α can be estimated by considering an estimating equation that involves specification of the working correlation structure. To simplify the calculations, we use the working independence matrix I. The estimating equation is given by (5.28) and details are discussed in Section 5.4.1.

5.2.3.3 Estimation

Consistent estimators of parameters β in the marginal mean model (5.1) can now be obtained by incorporating the weights and solving the following weighted estimating equation

$$u_\beta(\beta, \alpha) = \sum_i u_{i\beta}(\beta, \alpha) = \sum_i D_i'(x_i, \beta) V_i^{-1} W_i(\alpha)\left(y_i - p_i(\beta)\right) = 0, \tag{5.13}$$

where $D_i(x_i, \beta) = (\partial p_i / \partial \beta)$, V_i is a $T \times T$ working covariance matrix for y_i and $W_i(\alpha) = diag\{R_{i1}w_{i1}, ..., R_{iT}w_{iT}\}$ is a diagonal matrix of occasion-specific weights. The matrix $W_i(\alpha)$ incorporates the dropout process through the response indicators R_{it} which set i^{th} subject's weight at time t to: w_{it} if y_{it} is observed, and 0 if y_{it} is missing. Let us denote $u_{i\alpha} = u_{i\alpha}(\alpha)$, $u_{i\beta} = u_{i\beta}(\beta, \alpha)$, we will use this shortened notation in the remainder of this section. Under correctly specified models for the marginal means and the dropout process, there exists a unique solution $\hat{\beta}$ to equation (5.13). Robins et al. (1995) derived the following asymptotic equations for $\hat{\alpha}$ and $\hat{\beta}$ based on Taylor expansions

$$n^{1/2}(\hat{\alpha} - \alpha) = -n^{1/2} E\left(\sum \frac{\partial u_{i\alpha}}{\partial \alpha'}\right)^{-1} u_{i\alpha} + o_p(1), \tag{5.14}$$

$$n^{1/2}\left(\hat{\beta} - \beta\right) = -E\left(\frac{\partial u_{i\beta}}{\partial \beta'}\right)^{-1}\left(n^{-1/2}\sum_i u_{i\beta} + E\left(\frac{\partial u_{i\beta}}{\partial \alpha'}\right)^{-1} n^{1/2}(\hat{\alpha} - \alpha)\right) + o_p(1),$$

and, since $E(\partial u_{i\beta}/\partial \alpha') = -E(u_{i\beta}u'_{i\alpha})$ and $E(\partial u_{i\alpha}/\partial \alpha') = -Var(u_{i\alpha})$, they obtained

$$n^{-1/2}\left(\hat{\beta} - \beta\right) = -\Gamma^{-1} n^{-1/2} \sum_i Q_i + o_p(1), \tag{5.15}$$

where $\Gamma = E[\partial u_{i\beta}/\partial \beta']$ and $Q_i = u_{i\beta} - E[u_{i\beta}u'_{i\alpha}](E[u_{i\alpha}u'_{i\alpha}])^{-1} u_{i\alpha}$. From (5.15), law of large numbers and the central limit theorem, it follows that $\hat{\beta}$ is a consistent estimate of β and $n^{1/2}\left(\hat{\beta} - \beta\right)$ has an asymptotic normal distribution with mean zero and asymptotic variance $\Gamma^{-1}E[Q_iQ'_i](\Gamma^{-1})'$. The estimator of the asymptotic variance of $\hat{\beta}$ is given by

$$\left(\left(\sum_i D_i' V_i^{-1} W_i D_i\right)^{-1}\right)' \left(\sum_i q_i q_i'\right) \left(\sum_i D_i' V_i^{-1} W_i D_i\right)^{-1}, \tag{5.16}$$

where $q_i = u_{i\beta} - \left(\sum_i u_{i\beta} u_{i\alpha}'\right) \left(\sum_i u_{i\alpha} u_{i\alpha}'\right)^{-1} u_{i\alpha}$ and parameters β and α are replaced by their estimators $\hat{\beta}$ and $\hat{\alpha}$ (Preisser et al., 2000).

Given that the dropout model is specified correctly, WGEE does not require correct specification of the correlation structure to estimate β and $Var(\hat{\beta})$ consistently. However, if the missingness model is misspecified, $\hat{\beta}$ may be biased. As shown by Robins et al. (1995), if some observations have very small predicted probabilities of being observed, they will have inappropriately large weights and hence, large influence on the analysis which may imply problems. Preisser et al. (2000) suggest checking $(\max\{w_{iT}\})$ and further scrutiny of data points.

5.3 Goodness-of-Fit Test for Complete Data

Goodness-of-fit test are used to determine whether a fitted model adequately describes observed data and are based on an assessment of the fitted model's overall departure from the observed data. Hosmer and Lemeshow (1980) proposed a goodness-of-fit test based on grouping subjects according to their event probabilities from the logistic regression model. The estimated probabilities are ordered, and then separated into groups of approximately equal size; usually ten groups are recommended. For each group, we calculate the observed and expected number of events and use the Pearson's chi-square statistics to compare the counts. We begin by describing the Hosmer-Lemeshow (1980) goodness-of-fit method for cross-sectional data, and then present its extension to longitudinal data (Horton et al., 1999).

5.3.1 Cross-Sectional Data

Consider a marginal logistic regression model for cross-sectional data ($T = 1$, therefore the subscript t is omitted from the notation). Let $y_i \sim$Bin(p_i) and

$$\text{logit}(p_i) = x_i'\beta, \; i = 1, ..., n, \tag{5.17}$$

where β is a vector of regression parameters corresponding to a vector of p covariates $x_i = (x_{i1}, ..., x_{ip})'$. Let $\hat{\beta}$ denote the maximum likelihood estimator of β,the estimated event probabilities are

$$\hat{p}_i = p_i(\hat{\beta}) = \frac{\exp(x_i'\hat{\beta})}{1 + \exp(x_i'\hat{\beta})}. \tag{5.18}$$

Following Hosmer and Lemeshow's (1980) approach, we form $G = 10$ groups, approximately equal in size, based on deciles of risk determined by the estimated probabilities $\hat{p}_i$. Let $\hat{p}_{[1]} \leq \hat{p}_{[2]} \leq ... \leq \hat{p}_{[n]}$ represent the ordered values of $\hat{p}_i$,corresponding to subjects $i = 1, ..., n$. We construct the first group out of the first $n/10$ $\hat{p}_{[j]}$s with $j = 1, ..., n/10$, the second group out of the next $n/10$ $\hat{p}_{[j]}$s with $j = (n/10+1), ..., (2n/10)$, and so on, until 10 groups are formed. Subjects in the same group are considered similar in that they have similar estimated probabilities.

Given the partition of the data, we define group indicators: $I_{ig} = 1$ if $\hat{p}_i$ is in group g, or otherwise $I_{ig} = 0$, $g = 1, ..., G - 1$, and form the following alternative model by including the additional $(G-1)$ covariates $\gamma = [\gamma_1, ..., \gamma_{G-1}]'$ in (5.17), that is

$$\text{logit}(p_i) = x_i'\beta \ + \sum_{g=1}^{G-1} I_{ig}\gamma_g. \tag{5.19}$$

Even though I_{ig} is based on the random quantities $\hat{p}_i$, asymptotically, the partition can be considered as based on the true p_i, and therefore one can treat I_{ig} as a fixed covariate (Moore and Spruill, 1975). If the model (5.17) is specified correctly, then $\gamma = 0$ in (5.19). To test whether these additional covariates are significant (or we have evidence for lack of fit), we form the null hypothesis $H_0 : \gamma = 0$ and typically use the Pearson, Wald or Score test statistic.

- The Pearson test statistics is given by

$$X_P^2 = \sum_{g=1}^{G} \frac{(o_g - e_g)^2}{e_g(1 - e_g/n_g)},$$

  where $o_g = \sum_{i=1}^{n_j} y_{ig}$ and $e_g = \sum_{i=1}^{n_j} \hat{p}_{ig}$ respectively represent the observed, and expected, number of outcomes in the g^{th} decile, n_g is the size of the group, and the subscript ig refers to the i^{th} unit within the g^{th} decile group, $g = 1, ..., 10$. The distribution of X_P^2 was approximated by simulation studies as a chi-square with $G-2$ degrees of freedom under the model (5.17) (Hosmer and Lemeshow, 1980).

- The Wald statistic is given by

$$W = (o - e)'S^{-1}(o - e),$$

  where $o = (o_1, ..., o_G)'$, $e = (e_1, ..., e_G)$ and S is a consistent estimator of the covariance matrix of $o - e$ that can be obtained using Taylor series approximation (Graubart et al., 1997). Under the null hypothesis, W follows asymptotically a chi-square distribution with $G-1$ degrees of freedom. That is, we reject the fit of the logistic model at the α level when $W > \chi^2_{G-1,1-\alpha}$.

- The score test statistic for testing $H_0 : \gamma = 0$ is

$$X^2 = u(\tilde{\beta}, 0)' \left[(var(u(\beta,\gamma)))^{-1} \right]_{(\beta=\tilde{\beta},\gamma=0)} u(\tilde{\beta}, 0), \tag{5.20}$$

where

$$u(\beta,\gamma) = \begin{bmatrix} u_1(\beta,\gamma) \\ u_2(\beta,\gamma) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} (\partial p_i/\partial\beta)' V_i^{-1}(y_i - p_i) \\ (\partial p_i/\partial\gamma)' V_i^{-1}(y_i - p_i) \end{bmatrix}$$

is the score vector for model (5.19), $\tilde{\beta}$ is the estimate of β under H_0, and

$$\left[var(u(\tilde{\beta},\gamma)) \right]_{(\beta=\tilde{\beta},\gamma=0)} = -du(\beta,\gamma)/d\left[\beta,\gamma\right], \tag{5.21}$$

evaluated at $(\beta = \tilde{\beta}, \gamma = 0)$ (Parzen and Lipsitz, 1999). The fit of the logistic model is rejected at level α when $X^2 > \chi^2_{G-1,1-\alpha}$.

In general, the advantage of the Hosmer-Lemeshow tests is that they are based on intuitively appealing groupings of estimated probabilities; while the disadvantage is that, as confirmed by our simulation trials, the test statistic depends on the choice of cutpoints that define the groups. In the reminder of the chapter, we focus on the score test statistic and present the extension to the longitudinal data case in the next section. The advantages and disadvantages of the score test statistic are discussed in the appendix (Section 5.7.2).

5.3.2 Longitudinal Data

Suppose we want to determine if the mean in a marginal logistic regression model with repeated measures is correctly specified as

$$logit(p_{it}) = x_{it}'\beta,\ i = 1, ..., n,\ t = 1, ..., T. \tag{5.22}$$

Following Horton et al. (1999) who extended the Hosmer-Lemeshow theory to the case of longitudinal data, we form $G = 10$ groups of maximum $nT/10$ observations $(y_{it}, x_{it})'s$ based on deciles of risk derived from the corresponding values of $\hat{p}_{it}$, where $\text{logit}(\hat{p}_{it}) = x_{it}'\hat{\beta}$, and $\hat{\beta}$ is the GEE estimator of β. Because of possible ties in predicted risks, the total number of subjects within decile group may vary; it could also happen that a particular subject may belong to different decile groups at different times, and so the group-variable can be considered as a time-varying covariate.

Let us define the $(G-1)$ group indicators $I_{itg} = 1$ if $\hat{p}_{it}$ is in group g and $I_{itg} = 0$ otherwise, $g = 1, .., G-1$. Similarly, as in the case of cross-sectional data, to test goodness-of-fit for model (5.22), we treat I_{itg} as a fixed covariate and form an alternative model

$$logit(p_{it}) = x_{it}'\beta + \gamma_1 I_{it1} + ... + \gamma_{G-1} I_{itG-1}. \tag{5.23}$$

Let $\gamma = (\gamma_1, ..., \gamma_{G-1})$, a test of the fit of model (5.22) is equivalent to the test of

$$H_0 : \gamma = 0,$$

which can be conducted using the quasi-score statistic within the GEE framework. The score vector under the alternative $H_1 : \gamma \neq 0$ is

$$u(\beta,\gamma) = \begin{bmatrix} u_\beta(\beta,\gamma) \\ u_\gamma(\beta,\gamma) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} D'_{\beta i} V_i^{-1}(y_i - p_i(\beta,\gamma)) \\ D'_{\gamma i} V_i^{-1}(y_i - p_i(\beta,\gamma)) \end{bmatrix}, \tag{5.24}$$

where $D_{\beta i} = \dfrac{\partial p_i(\beta,\gamma)}{\partial \beta}$, $D_{\gamma i} = \dfrac{\partial p_i(\beta,\gamma)}{\partial \gamma}$, $p_i = (p_{i1}, p_{i2}, ..., p_{iTi})$, $y_i = (y_{i1}, y_{i2}, ..., y_{iT_i})$ and $V_i = A_i^{1/2} H_i(\eta) A_i^{1/2}$ with $A_i = diag(p_{it}(1-p_{it}))$ and a working correlation matrix $H_i(\eta)$ that depends on an unknown parameter vector η.

Let $\tilde{\beta}$ be the estimate of β under H_0 obtained by solving $u_\beta(\tilde{\beta}, 0) = 0$. The general score test statistic for for testing $H_0 : \gamma = 0$ is based on large sample distribution of $u_\gamma(\tilde{\beta}, 0)$. It is given by

$$X^2 = u_\gamma(\tilde{\beta}, 0)' \left(\widehat{var}(u_\gamma(\tilde{\beta}, 0))\right)^{-1} u_\gamma(\tilde{\beta}, 0), \tag{5.25}$$

and is asymptotically distributed as χ^2_{G-1} under H_0 (Horton et al. 1999).

5.4 Goodness-of-Fit Test in the Presence of Missing Response Data

We now propose a goodness-of-fit test for longitudinal data with missing observations. The approach is based on combining Horton's goodness-of-fit method (Horton et al., 1999) with theWGEE approach to estimate parameters in the marginal mean model in the presence of missing responses (Robins et al., 1995 and Preisser et al., 2000).

5.4.1 Models for Response Data and Missingness Process

We consider a logit marginal model (5.22) for binary responses y_{it} and obtain the estimators of parameters following closely the method presented by Preisser et al. (2000). We first determine a set of weights, through a model for missingness, and then apply these weights in the estimation of parameters procedure. As before, to indicate availability of data, we let R_{it} be the response indicator for y_{it} ($R_{it} = 1$ if y_{it} is observed or otherwise $R_{it} = 0$), so that the vector $R_i = (R_{i1}, R_{i2}, ..., R_{iT})'$ contains information on the completeness of the response for individual $i = 1, ..., n$. The probability of observing a response from individual i at time t, conditional on the individual being observed at time $t - 1$, is

$$\lambda_{it}(\alpha) = P(R_{it} = 1|\breve{R}_{it}, y_i^0, x_i, \alpha), \tag{5.26}$$

where, for individual i, $\breve{R}_{it} = (R_{i1}, ..., R_{i(t-1)})'$ contains information on the response history, y_i^0 denotes the observed values of $y_i = (y_{i1}, ..., y_{iT})$, $x_i = (x_{i1}, ..., x_{iT})$ and α is a vector of unknown parameters. To obtain an estimator of λ_{it}, we fit a logistic regression model

$$\text{logit}\, \lambda_{it}(\alpha) = v_{it}\alpha, \tag{5.27}$$

where ν_{it} is a function of $\breve{R}_{it}$, y_i^0 and x_i, $t = 2, ..., T$.

The parameter vector α of the missingness model (5.27) can be estimated by considering an estimating equation

$$u_\alpha(\alpha) = \sum_{i=1}^{n} u_{i\alpha}(\alpha) = \sum_{i=1}^{n} D_i^{*\prime} V_i^{*-1}(R_i - \lambda_i(\alpha)) = 0, \tag{5.28}$$

where $D_i^* = \partial\lambda_i(\alpha)/\partial\alpha$, $\lambda_i = (\lambda_{i2}, ..., \lambda_{iT})$, $R_i = (R_{i2}, ..., R_{iT})$, and $V_i^* = A_i^{*1/2} H_i A_i^{*1/2} = diag(\lambda_{it}(1-\lambda_{it}))$ under working independence matrix $H_i = I$. Notice that under the independence assumption, the estimating equation corresponds to the likelihood function for the binary indicators of missing data. We also note that unless $H_i = I$, V_i^* would involve covariates from the observed and unobserved occasions, which we assumed to be known.

Under the MAR mechanism, for $t \geq 2$,the marginal probability of observing a response from individual i at time t is

$$\begin{aligned} \pi_{it} &= P(R_{it} = 1|y_i, x_i) = P(R_{it} = 1|y_i^0, x_i) & (5.29) \\ &= \sum_{\breve{R}_{it}} \lambda_{it} \prod_{l=2}^{t-1} \lambda_{il}^{r_{il}} (1-\lambda_{il})^{1-r_{il}}, & (5.30) \end{aligned}$$

since $P(R_{it} = 1|y_i^0, x_i) = \sum_{\breve{R}_{it}} P(R_{it} = 1|\breve{R}_{it}, y_i^0, x_i) \times P(R_{i(t-1)} = r_{i-1}|\breve{R}_{i(t-1)}, y_i^0, x_i) \times P(R_{i(t-2)} = r_{i(t-2)}|\breve{R}_{i(t-2)}, y_i^0, x_i) \times ... \times P(R_{i2} = r_{i2}|\breve{R}_{i2}, y_i^0, x_i) \times P(R_{i1} = r_{i1}|\breve{R}_{i1}, y_i^0, x_i)$ where $r_{it} \in \{0, 1\}$ and the summation is taken over all the possible values of the response history $\breve{R}_{it}$.

The weights are then determined by the equation

$$w_{it}^{-1} = \prod_{t} \pi_{it}, \tag{5.31}$$

and the WGEE based on the observed responses can be formulated as follows

$$u_{\beta}(\beta, \alpha) = \sum_{i=1}^{n} u_{i\beta}(\beta, \alpha) = \sum_{i=1}^{n} D_i' V_i^{-1} W_i(\alpha)(y_i - p_i(\beta)) = 0, \tag{5.32}$$

where $W_i(\alpha) = diag\,(R_{it} w_{it}, t = 1, ..., T)$, $V_i = A_i^{1/2} H_i A_i^{1/2} = diag(p_{it}(\beta)(1 - p_{it}(\beta)))$ under a $(T \times T)$ working independence correlation matrix $H_i = I$, and $D_i = \partial p_i(\beta)/\partial \beta'$.

A Newton-Raphson iterative method can be applied to solve (5.28) and (5.32) simultaneously using initial values $\beta^{(0)}, \alpha^{(0)}$. The iterations are given by

$$\begin{aligned} \begin{pmatrix} \hat{\beta}^{(m+1)} \\ \hat{\alpha}^{(m+1)} \end{pmatrix} &= \begin{pmatrix} \hat{\beta}^{(m)} \\ \hat{\alpha}^{(m)} \end{pmatrix} + \\ &\quad - \left. \begin{pmatrix} \sum_{i=1}^{n} \frac{\partial}{\partial \beta} u_{i\beta}(\beta, \alpha) & \sum_{i=1}^{n} \frac{\partial}{\partial \alpha} u_{i\beta}(\beta, \alpha) \\ 0 & \sum_{i=1}^{n} \frac{\partial}{\partial \alpha} u_{i\alpha}(\alpha) \end{pmatrix} \right|_{(\hat{\beta}^{(m)}, \hat{\alpha}^{(m)})}^{-1} \\ &\quad \times \left. \begin{pmatrix} \sum_{i=1}^{n} u_{i\beta}(\beta, \alpha) \\ \sum_{i=1}^{n} u_{i\alpha}(\alpha) \end{pmatrix} \right|_{(\hat{\beta}^{(m)}, \hat{\alpha}^{(m)})}, \end{aligned}$$

where the superscript (m) denotes the iteration step. When an assumed level of convergence is reached, we obtain $\hat{\beta}$ and $\hat{\alpha}$.

Note that the solution can also be achieved by applying a two-stage iterative procedure, where at the first step, we obtain

$$\hat{\alpha}^{(m+1)} = \hat{\alpha}^{(m)} - \left[\sum_{i=1}^{n} \frac{\partial}{\partial \alpha'} u_{i\alpha}\left(\hat{\alpha}^{(m)}\right)\right]^{-1} \sum_{i=1}^{n} u_{i\alpha}(\hat{\alpha}^{(m)}),$$

which is then used to calculate

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} - \left[\sum_{i=1}^{n} \frac{\partial}{\partial(\beta,\alpha)'} u_{i\beta}\left(\hat{\beta}^{(m)}, \hat{\alpha}^{(m+1)}\right)\right]^{-1} \sum_{i=1}^{n} u_{i\beta}(\hat{\beta}^{(m)}, \hat{\alpha}^{(m+1)}).$$

The two procedures converge to the same limit under mild regularity conditions. Let $\theta = (\beta', \alpha')$ and denote $H_i(\theta) = (u_{i\beta}(\beta,\alpha), u_{i\alpha}(\alpha))'$. Since $E\left[H_i(\theta)\right] = 0$, given that the response and missingness models are specified correctly and under standard regularity conditions, by Theorem 3.4 of Newey and McFadden (1993), there exist a unique solution $\hat{\theta}$ to the equation $\sum_i H_i(\theta) = 0$ w.p. 1 that satisfies

$$n^{1/2}(\hat{\theta} - \theta) = -\left(E\left[\partial H_i(\theta)/\partial\theta'\right]\right)^{-1} n^{-1/2} \sum_i H_i(\theta) + o_p(1). \tag{5.33}$$

and for the estimator $\hat{\beta}$ we obtain (5.15). Then by the Central Limit Theorem, the asymptotic distribution for $n^{1/2}(\hat{\beta} - \beta)$ is normal with mean 0 and asymptotic variance given by (5.16).

5.4.2 Goodness-of-fit Test

We will now establish a method to determine if the marginal model (5.22) is a good fit following Horton's approach for fully observed repeated binary data (Horton et al., 1999). Similarly to the case with fully-observed data (Section 5.3.2), we first estimate the probabilities $\hat{p}_{it} = p_{it}(\hat{\beta})$, and consider $G = 10$ decile groups to form the alternative model

$$H_1 : \text{logit}(p_{it}) = x'_{it}\beta + \gamma_1 I_{it1} + ... + \gamma_{G-1} I_{itG-1} \tag{5.34}$$

where $I_{itg} = 1$ if $\hat{p}_{it}$ is in group g or $I_{itg} = 0$ otherwise. Here however, we use the WGEE estimator $\hat{\beta}$, which also involves estimation of the missingness model parameter $\hat{\alpha}$, to obtain $\hat{p}_{it}$. Note that the nine group indicators I_{itg} are formed based on $\hat{p}'_{it}s$ corresponding to observed data. The score vector under $H_1 : \gamma \neq 0$, where $\gamma = [\gamma_1, ..., \gamma_{G-1}]'$, incorporates (5.32), (5.28) and $u_\gamma(\beta, \alpha, \gamma) = D'_{\gamma i} V_i^{-1} W_i(\alpha)(y_i - p_i(\beta, \gamma))$, where $D_{\gamma i} = \partial p_i(\beta, \gamma)/\partial \gamma'$ and V_i defined as in (5.32). A score vector for the goodness-of-fit problem that incorporates the missingness mechanism is

$$\begin{aligned} u(\beta, \alpha, \gamma) &= \begin{pmatrix} u_\beta(\beta, \alpha, \gamma) \\ u_\gamma(\beta, \alpha, \gamma) \\ u_\alpha(\alpha) \end{pmatrix} \\ &= \sum_{i=1}^{n} \begin{pmatrix} D'_{\beta i} V_i^{-1} W_i(\alpha)(y_i - p_i(\beta, \gamma)) \\ D'_{\gamma i} V_i^{-1} W_i(\alpha)(y_i - p_i(\beta, \gamma)) \\ D_i^{*\prime} V_i^{*-1}(R_i - \lambda_i) \end{pmatrix}. \end{aligned}$$

Note that under H_0, the first row of $D_{\beta i}$ is given by $\text{logit}(x'_{it}\beta)(1, x_{i11}, ..., x_{i1p})$, and the the first row of $D_{\gamma i}$ is given by $\text{logit}(x'_{it}\beta)(I_{i11}, ..., I_{i1G-1})$. Further, let us denote the combined vector of data and missingness models parameters by θ, that is $\theta = (\beta', \alpha')'$, and split $u(\beta, \alpha, \gamma)$ into two parts $u_1(\theta, \gamma) = [u_\beta(\theta, \gamma)', u_\alpha(\alpha)']'$ and $u_2(\theta, \gamma) = u_\gamma(\theta, \gamma)$. The score test statistic for testing $H_0 : \gamma = 0$ is $X^2 = u(\hat{\theta}, 0)'\widehat{var}\left(u(\hat{\theta}, 0)\right)^{-1} u(\hat{\theta}, 0)$ but since $\hat{\theta} = (\hat{\beta}', \hat{\alpha}')'$ is obtained by solving the estimating equations $u_1(\hat{\theta}, 0) = 0$, the score test for testing $H_0 : \gamma = 0$ is, in fact, based on the score $u_2(\hat{\theta}, 0)$ and is given by

$$X^2 = u_2(\hat{\theta}, 0)' \left(var(u_2(\hat{\theta}, 0))\right)^{-1} u_2(\hat{\theta}, 0), \tag{5.35}$$

and it asymptotically follows χ^2_{G-1} distribution under H_0. To show this result, we expand $u_1(\hat{\theta}, 0)$ around the true value θ using the Taylor's series approximation $u_1(\hat{\theta}, 0) \simeq u_1(\theta, 0) + (\partial u_1(\theta, 0)/\partial\theta)\left(\hat{\theta} - \theta\right)$. Since $u_1(\hat{\theta}, 0) = 0$, we obtain

$$\left(\hat{\theta} - \theta\right) \simeq -\left(\frac{\partial u_1(\theta, 0)}{\partial\theta}\right)^{-1} u_1(\theta, 0). \tag{5.36}$$

Similarly, we can approximate $u_2(\hat{\theta}, 0) \simeq u_2(\theta, 0) + (\partial u_2(\theta, 0)/\partial\theta)\left(\hat{\theta} - \theta\right)$ and then substitute (5.36) for $\left(\hat{\theta} - \theta\right)$ to obtain

$$u_2(\hat{\theta}, 0) \simeq \left(I_{G-1}, -\left(\frac{\partial u_2(\theta, 0)}{\partial\theta}\right)\left(\frac{\partial u_1(\theta, 0)}{\partial\theta}\right)^{-1}\right)(u_2(\theta, 0)', u_1(\theta, 0)')',$$

where I_{G-1} is a $(G-1) \times (G-1)$ identity matrix. Let us replace $\partial u_2(\theta, 0)/\partial\theta$ and $\partial u_1(\theta, 0)/\partial\theta$ by their expected values and denote

$$A = \left(I_{G-1}, -E\left(\frac{\partial u_2(\theta,0)}{\partial\theta}\right)\left(E\left(\frac{\partial u_1(\theta,0)}{\partial\theta}\right)\right)^{-1}\right),$$

so that we can write $u_2(\hat{\theta}, 0) \simeq A\left(u_2(\theta,0)', u_1(\theta,0)'\right)'$. It follows that, under H_0, $u_2(\hat{\theta}, 0)$ is asymptotically normal with mean 0. Also

$$\begin{aligned} var(u_2(\hat{\theta},0)) &= var\left(A\left(u_2(\theta,0)', u_1(\theta,0)'\right)'\right) \\ &= AE\left(\left(u_2(\theta,0)', u_1(\theta,0)'\right)'\left(u_2(\theta,0)', u_1(\theta,0)'\right)\right)A', \end{aligned}$$

since $E\left(\left(u_2(\theta,0)', u_1(\theta,0)'\right)'\right) = 0$ when the response and missing data models are specified correctly. Therefore, X^2 is asymptotically chi-square under H_0.

Finally, we obtain the variance estimate, as presented in Horton et al. (1999), by replacing $E\left(\partial u_2(\theta,0)/\partial\theta\right)$ by $E\left(\partial u_2(\theta,\gamma)/\partial\theta\right)|_{\theta=\hat{\theta},\gamma=0}$, substituting $E\left(\partial u_1(\theta,\gamma)/\partial\theta\right)|_{\theta=\hat{\theta},\gamma=0}$ for $E\left(\partial u_1(\theta,0)/\partial\theta\right)$ in A, and $n^{-1}\sum_i(u_i)$ for $E(u)$. That is

$$\widehat{var}(u_2(\hat{\theta},0)) = \widehat{A}\left(n^{-1}\sum_i\left(u_{i2}(\hat{\theta},0)', u_{i1}(\hat{\theta},0)'\right)'\left(u_{i2}(\hat{\theta},0)', u_{i1}(\hat{\theta},0)'\right)\right)\widehat{A}',$$

where

$$\hat{A} = \left(I, -\sum_i\left(\frac{\partial u_{2i}(\theta,\gamma)}{\partial\theta}\right)\Bigg|_{\theta=\hat{\theta},\gamma=0}\left(\sum_i\left(\frac{\partial u_{1i}(\theta,\gamma)}{\partial\theta}\right)\Bigg|_{\theta=\hat{\theta},\gamma=0}\right)^{-1}\right).$$

5.5 Simulation Study

5.5.1 Set-up

In this section, we draw a comparison between the proposed goodness-of-fit method, which incorporates the estimation of the missingness model parameters, and a simpler method that ignores the missingness process, as if the data were MCAR. We will refer to this method as "Horton's" method. The data and the missingness models used in this simulation study are slight modifications of the set-up to analyse trends in cigarette smoking, supported by the data from the Coronary Artery Risk Development in Young Adults (CARDIA) study, in Preisser et al. (2000).

5.5.1.1 Data Model

We restrict our data to three measurements per subject ($T = 3$) and assume the following marginal model for binary responses y_{it} :

$$\text{logit}\,(p_{it}) = \beta_0 + \beta_1(t-1) + x_{it}\beta_2,\ i = 1, ..., 500,\ t = 1, 2, 3, \tag{5.37}$$

with the true parameter values set to $\beta_0 = -0.5$, $\beta_1 = 0.1$ and $\beta_2 = 0.3$ and the covariates x_{it} generated independently from standard uniform $U(0, 1)$ random variables resulting in $E\,(p_i) = (0.41, 0.44, 0.46)$, which indicates a moderate increase in the prevalence of an event. We obtain the correlated binary responses (y_{i1}, y_{i2}, y_{i3}) using a Bahadur model with $\rho_{i12} = \rho_{i13} = \rho_{i23} = 0.3$, and a third-order association parameter $\rho_{i123} = 0.6$. The choice of a method to obtain correlated binary responses is discussed in Section 5.7.1 of the appendix.

5.5.1.2 Missingness Model

We assume that data are fully observed at time $t = 1$, that is $R_{i0} = 1$ for all individuals i, and consider the intermittent missingness pattern that allows for $R_{i2} = 0$ and $R_{i3} = 1$. The missingness process at time points $t = 2$ and $t = 3$ is expressed as a logistic regression:

$$\text{logit}\,(\lambda_{it}) = \alpha_0 + \alpha_1 I_{\{t=3\}} + \alpha_2(2y_{i1}\text{-}1) + (1\text{-}2R_{it-1})\alpha_3 + (1\text{-}2R_{it-1})(2y_{i1}\text{-}1)\alpha_4, \quad (5.38)$$

where R_{it} denotes availability of data, $I_{\{t=3\}}$ is 1 at $t = 3$ and 0 at $t = 1$ and $t = 2$; that is, logit$(\lambda_{i2}) = \alpha_0$+$\alpha_2(2y_{i1}$-1)-$\alpha_3$-$(2y_{i1}$-1)$\alpha_4$ for $t = 2$, and logit$(\lambda_{i3}) = \alpha_0$+$\alpha_1$+$\alpha_2(2y_{i1}$-1)+(1-2$R_{i2})\alpha_3$+(1-2$R_{i2})(2y_{i1}$-1)$\alpha_4$ for $t = 3$.

We obtain R_{i2} and R_{i3} for 25 sets of simulations corresponding to combinations of parameter values $(\alpha_0, \alpha_0, \alpha_3, \alpha_3, \alpha_4)$ which follow closely the setup of the simulation study in Preisser et al. (2000). The parameter values and resulting average response rates at $t = 2$ and $t = 3$ (denoted respectively by $\bar{R}_2$ and $\bar{R}_3$) are listed in Table 5.1.

For all the observations at $t = 1$ we set the weights to 1, then at $t = 2$,

$$w_{i2}^{-1} = \lambda_{i2},$$

and at $t = 3$,

$$w_{i3}^{-1} = \lambda_{i3}^{(1)}\lambda_{i2} + \lambda_{i3}^{(0)}(1 - \lambda_{i2}),$$

where $\lambda_{i3}^{(j)} = P(R_{it} = 1|R_{i2} = j, x_i, y_i, \alpha)$, $j = 0, 1$, $i = 1, ..., n, .$

Let us define $f_{i3\alpha}^{(j)}$ =logit($\lambda_{i3}^{(j)}$) that is $f_{i3\alpha}^{(1)} = \alpha_0$+$\alpha_1$+$\alpha_2(2y_{i1}$-1)-$\alpha_3$-$(2y_{i1}$-1)$\alpha_4$, and

Table 5.1: Missingness model parameters and resulting average response rates.

Scenario	α_0	α_1	α_2	α_3	α_4	$\bar{R}_2$	$\bar{R}_3$
1	1	-0.1	0	0	0	75%	73%
2	1	-0.1	-0.1	0	0	75%	75%
3	1	-0.1	-0.2	0	0	75%	75%
4	0	-0.1	0	-1.4	0	82%	67%
5	0	-0.1	-0.1	-1.4	0	82%	69%
6	0	-0.1	-0.2	-1.4	0	82%	69%
7	1	-0.1	0	0	0.5	74%	76%
8	1	-0.1	-0.1	0	0.5	74%	76%
9	1	-0.1	-0.2	0	0.5	73%	78%
10	0	-0.1	0	-1.4	0.5	80%	71%
11	0	-0.1	-0.1	-1.4	0.5	79%	72%
12	0	-0.1	-0.2	-1.4	0.5	79%	71%
13	1	-0.1	0	0	1	73%	80%
14	1	-0.1	-0.1	0	1	72%	79%
15	1	-0.1	-0.2	0	1	72%	81%
16	0	-0.1	0	-1.4	1	77%	72%
17	0	-0.1	-0.1	-1.4	1	77%	72%
18	0	-0.1	-0.2	-1.4	1	77%	72%
19	1	-0.1	0	0	1.5	71%	81%
20	1	-0.1	-0.1	0	1.5	71%	82%
21	1	-0.1	-0.2	0	1.5	69%	84%
22	0	-0.1	0	-1.4	1.5	76%	75%
23	0	-0.1	-0.1	-1.4	1.5	76%	74%
24	0	-0.1	-0.2	-1.4	1.5	75%	74%
25	0	-0.1	0	-0.4	1.4	63%	70%

$f_{i3\alpha}^{(0)} = \alpha_0+\alpha_1+\alpha_2(2y_{i1}\text{-}1)+\alpha_3+(2y_{i1}\text{-}1)\alpha_4$. Note that since we only have two possibilities for the response history $(r_{i1}, r_{i2}) = (1, 0)$ or $(r_{i1}, r_{i2}) = (1, 1)$, $w_{i3}^{-1} = \sum_{\bar{R}_{i3}} \lambda_{i3}\lambda_{i2}^{r_{i2}}(1-\lambda_{i2})^{1-r_{i2}}$ agrees with equations (5.29) and (5.31). That is, for $t = 1, 2, 3$, the intermittent missingness model (5.38) is a special case of the model we proposed in Section 5.4.1.

5.5.2 Parameter Estimation

In the simulations, we use independence working covariance structure to simplify the computations. Let θ denote the true parameter vector (β, α), and similarly we construct the estimates vector $\hat{\theta} = (\hat{\beta}, \hat{\alpha})$. We obtain $\hat{\theta}$ using a joint iterative equation given by

$$\theta^{(m+1)} = \theta^{(m)} - \begin{pmatrix} \sum_i \dfrac{\partial u_{i\beta}(\beta,\alpha)}{\partial \beta} & \sum_i \dfrac{\partial u_{i\beta}(\beta,\alpha)}{\partial \alpha} \\ 0 & \sum_i \dfrac{\partial u_{i\alpha}(\alpha)}{\partial \alpha} \end{pmatrix}^{-1} \begin{pmatrix} \sum_i u_{i\beta}(\beta,\alpha) \\ \sum_i u_{i\alpha}(\alpha) \end{pmatrix}\Bigg|_{\theta^{(m)}},$$

at iteration m. Let $f_{it\beta} = \text{logit}(p_{it})$ and $f_{it\alpha} = \text{logit}(\lambda_{it})$. Noting that $\partial p_{it}/\partial\beta = p_{it}(1-p_{it})\partial f_{it\beta}/\partial\beta$, it can be shown that

$$\begin{aligned} u_{i\beta_j}(\beta,\alpha) &= (y_{i1}-p_{i1})\frac{\partial f_{i1\beta}}{\partial \beta_j} + (y_{i2}-p_{i2})\,R_{i2}w_{i2}(\alpha)\frac{\partial f_{i2\beta}}{\partial \beta_j} + \\ &\quad + (y_{i3}-p_{i3})\,R_{i3}w_{i3}(\alpha)\frac{\partial f_{i3\beta}}{\partial \beta_j}, \end{aligned}$$

and similarly,

$$u_{i\alpha_l}(\beta,\alpha) = (R_{i2}-\lambda_{i2})\frac{\partial f_{i2\alpha}}{\partial \alpha_l} + (R_{i3}-\lambda_{i3})\frac{\partial f_{i3\alpha}}{\partial \alpha_l}.$$

The derivatives of the score vector are given by

$$\begin{aligned}\frac{\partial u_{i\beta j}}{\partial \beta_k} &= -p_{i1}(1-p_{i1})\frac{\partial f_{i1\beta}}{\partial \beta_j}\frac{\partial f_{i1\beta}}{\partial \beta_k} - p_{i2}(1-p_{i2})R_{i2}w_{i2}(\alpha)\frac{\partial f_{i2\beta}}{\partial \beta_j}\frac{\partial f_{i2\beta}}{\partial \beta_k} + \\ & -p_{i3}(1-p_{i3})R_{i3}w_{i3}(\alpha)\frac{\partial f_{i3\beta}}{\partial \beta_j}\frac{\partial f_{i3\beta}}{\partial \beta_k},\end{aligned}$$

and,

$$\frac{\partial u_{i\alpha_l}}{\partial \alpha_m} = -\lambda_{i2}(1-\lambda_{i2})\frac{\partial f_{i2\alpha}}{\partial \alpha_l}\frac{\partial f_{i2\alpha}}{\partial \alpha_m} - \lambda_{i3}(1-\lambda_{i3})\frac{\partial f_{i3\alpha}}{\partial \alpha_l}\frac{\partial f_{i3\alpha}}{\partial \alpha_m},$$

with $j,k = 0,1,2$ and $l,m = 0,1,..,4$. Also

$$\begin{aligned}\frac{\partial u_{i\beta j}}{\partial \alpha_l} &= (y_{i2}-p_{i2})\,R_{i2}\frac{\partial f_{i2\beta}}{\partial \beta_j}\frac{\partial w_{i2}(\alpha)}{\partial \alpha_l} + \\ & + (y_{i3}-p_{i3})\,R_{i3}\frac{\partial f_{i3\beta}}{\partial \beta_j}\frac{\partial w_{i3}(\alpha)}{\partial \alpha_l}.\end{aligned}$$

Note that

$$\frac{\partial w_{i2}(\alpha)}{\partial \alpha_l} = -(\exp(f_{it\alpha}))^{-1}\frac{\partial f_{it\alpha}}{\partial \alpha_l},$$

and

$$\begin{aligned}\frac{\partial w_{i3}(\alpha)}{\partial \alpha_l} &= -w_{i3}^2(\alpha)\left(\lambda_{i3}^{(1)}\lambda_{i2}\left[(1-\lambda_{i3}^{(1)})\partial f_{i3\alpha}^{(1)}/\partial \alpha_l + (1-\lambda_{i2})\partial f_{i2\alpha}/\partial \alpha_l\right] + \right.\\ & \left. +\lambda_{i3}^{(0)}\,(1-\lambda_{i2})\left[(1-\lambda_{i3}^{(0)})\partial f_{i3\alpha}^{(0)}/\partial \alpha - \lambda_{i2}\partial f_{i2\alpha}/\partial \alpha_l\right]\right).\end{aligned}$$

5.5.3 Goodness-of-Fit Test

5.5.3.1 Proposed Method

To test the goodness-of-fit of model (5.37), we consider the alternative model

$$\log it(p_{it}) = \beta_0 + \beta_1(t-1) + x_i\beta_2 + \gamma_1 I_{it1} + ... + \gamma_9 I_{it9}, \tag{5.39}$$

where, for each observation i, the group indicators I_{itg} are formed on the basis of the estimated probabilities $\hat{p}_{it} = p_{it}(\hat{\beta})$ for $t = 1, 2, 3$ and $g = 1, ..., 9$. If model (5.37) is appropriate then $\gamma_1 = ... = \gamma_9 = 0$ in (5.39) and a test of the fit of the model is equivalent to a test of $H_0 : \gamma_1 = ... = \gamma_9 = 0$. The score test statistic is estimated by

$$X^2 = u_2(\hat{\beta}, \hat{\alpha}, 0)' \left(\widehat{var}(u_2(\hat{\beta}, \hat{\alpha}, 0))\right)^{-1} u_2(\hat{\beta}, \hat{\alpha}, 0),$$

where $u_2(\hat{\beta}, \hat{\alpha}, 0) = \sum_i u_{i\gamma}(\hat{\beta}, \hat{\alpha}, 0)$ with

$$\begin{aligned} u_{i\gamma}(\beta, \alpha, \gamma) &= D'_{\gamma i} V_i^{-1} W_i(\alpha)(y_i - p_i) \\ &= (y_{i1} - p_{i1}) \frac{\partial f_{i1\gamma}}{\partial \gamma} + (y_{i2} - p_{i2}) R_{i2} w_{i2}(\alpha) \frac{\partial f_{i2\gamma}}{\partial \gamma} + \\ &\quad + (y_{i3} - p_{i3}) R_{i3} w_{i3}(\alpha) \frac{\partial f_{i3\gamma}}{\partial \gamma}, \end{aligned}$$

$f_{i1\gamma} = \beta_0 + \beta_1(t-1) + x_i\beta_2 + \gamma_1 I_{it1} + ... + \gamma_9 I_{it9}$ and $\partial f_{it\gamma}/\partial\gamma_g = I_{itg}$, $g = 1, ..., 9$. The variance estimate is obtained using

$$\widehat{var}(u_2(\hat{\theta},0)) = \widehat{A}\left(n^{-1}\sum_i \begin{pmatrix} u_{i\gamma}(\hat{\theta},0) \\ u_{i\theta}(\hat{\theta},0) \end{pmatrix} \begin{pmatrix} u_{i\gamma}(\hat{\theta},0) \\ u_{i\theta}(\hat{\theta},0) \end{pmatrix}'\right)\widehat{A}', \tag{5.40}$$

where $\theta = (\beta, \alpha)$, $u_{i\theta} = (u_{i\beta}, u_{i\alpha})'$,

$$\hat{A} = \left(I, -\sum_i \left(\frac{\partial u_{i\gamma}(\hat{\theta},0)}{\partial \theta}\right)\left(\sum_i \left(\frac{\partial u_{i\theta}(\hat{\theta},0)}{\partial \theta}\right)\right)^{-1}\right),$$

and I represents a 9×9 identity matrix. As shown before,

$$\left(\sum_i \left(\frac{\partial u_{i\theta}(\hat{\theta},0)}{\partial \theta}\right)\right)^{-1} = \begin{pmatrix} \sum_i \frac{\partial u_{i\beta}(\hat{\beta},\hat{\alpha},0)}{\partial \beta} & \sum_i \frac{\partial u_{i\beta}(\hat{\beta},\hat{\alpha},0)}{\partial \alpha} \\ 0 & \sum_i \frac{\partial u_{i\alpha}(\hat{\alpha})}{\partial \alpha} \end{pmatrix}^{-1}.$$

Finally,

$$\sum_i \left(\frac{\partial u_{i\gamma}(\hat{\theta},0)}{\partial \theta}\right) = \begin{pmatrix} \sum_i \frac{\partial u_{i\gamma}(\hat{\beta},\hat{\alpha},0)}{\partial \beta} & \sum_i \frac{\partial u_{i\gamma}(\hat{\beta},\hat{\alpha},0)}{\partial \alpha} \end{pmatrix},$$

with

$$\begin{aligned} \frac{\partial u_{i\gamma g}}{\partial \beta_k} &= -p_{i1}(1-p_{i1})\frac{\partial f_{i1\gamma}}{\partial \gamma_g}\frac{\partial f_{i1\gamma}}{\partial \beta_k} - p_{i2}(1-p_{i2})R_{i2}w_{i2}(\alpha)\frac{f_{i2\gamma}}{\partial \gamma_g}\frac{f_{i2\gamma}}{\partial \beta_k} + \\ & -p_{i3}(1-p_{i3})R_{i3}w_{i3}(\alpha)\frac{f_{i3\gamma}}{\partial \gamma_g}\frac{f_{i3\gamma}}{\partial \beta_k}, \end{aligned}$$

and

$$\frac{\partial u_{i\gamma g}}{\partial \alpha_l} = (y_{i2} - p_{i2})R_{i2}\frac{f_{i2\gamma}}{\partial \gamma_g}\frac{\partial w_{i2}(\alpha)}{\partial \alpha_l} + (y_{i3} - p_{i3})R_{i3}\frac{f_{i3\gamma}}{\partial \gamma_g}\frac{\partial w_{i3}(\alpha)}{\partial \alpha_l}.$$

5.5.3.2 Horton's Method

The set-up for simulation of the Horton's method is based on the models used for the proposed approach. In particular, to obtain the results based on the method that ignores missingness process, we removed all estimating equations involving the missingness model parameters α while keeping the missingness indicators R_1, R_2, R_3 obtained in the same way as for the proposed method, and set all weights w_i to 1, $i = 1, 2, 3$. That is, $\hat{\beta}$ was obtained using a joint iterative equation given by

$$\beta^{(h+1)} = \beta^{(h)} - \left(\sum_i \frac{\partial u_{i\beta}(\beta)}{\partial \beta} \right)^{-1} \left(\sum_i u_{i\beta}(\beta) \right) \Bigg|_{\beta^{(h)}},$$

where

$$u_{i\beta_j}(\beta) = (y_{i1} - p_{i1}) \frac{\partial f_{i1\beta}}{\partial \beta_j} + (y_{i2} - p_{i2}) R_{i2} \frac{\partial f_{i2\beta}}{\partial \beta_j} + (y_{i3} - p_{i3}) R_{i3} \frac{\partial f_{i3\beta}}{\partial \beta_j}.$$

To test goodness-of-fit, we used Horton's approach for fully observed repeated binary data (Horton et al., 1999) described in Section 5.3. That is,

$u(\beta, \gamma) = \sum_{i=1}^{n} \begin{pmatrix} u_{i\beta}(\beta, \gamma) \\ u_{i\gamma}(\beta, \gamma) \end{pmatrix}$ with $u_{i\beta_j}(\beta)$ as shown above and

$$u_{i\gamma}(\beta, \gamma) = (y_{i1} - p_{i1}) \frac{\partial f_{i1\gamma}}{\partial \gamma} + (y_{i2} - p_{i2}) R_{i2} \frac{\partial f_{i2\gamma}}{\partial \gamma} + (y_{i3} - p_{i3}) R_{i3} \frac{\partial f_{i3\gamma}}{\partial \gamma}.$$

5.5.4 Results

The results are based on random samples of n =500 individuals and averaged over K=2000 independent simulations.

For each parameter θ, we obtain percent relative bias defined as

$$PB(\hat{\theta}) = \frac{1}{K}\sum_{k=1}^{K}\frac{\left(\hat{\theta}_k - \theta\right)}{\theta} * 100\%,$$

and estimated mean square error

$$MSE(\hat{\theta}) = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\theta}_k - \theta\right)^2,$$

where, in both equations, $\hat{\theta}_k$ represents a parameter estimate based on data from k^{th} simulation, with $k = 1, ..., 2000$.

To assess the goodness-of-fit procedure, the size of the test is estimated as the percentage rejection of $H_0 : \gamma = 0$, under H_0, according to the significance level of 0.05 based on K simulations.

Tables 5.2 and 5.3 show the estimated relative bias and standard error for the parameters from the data model (5.37) for the 25 missingness model scenarios described in Table 5.1. Table 5.4 shows the estimated size of the proposed goodness-of-fit test. The results are also presented graphically as a scatter graph and boxplots in Figures 5.1 and 5.2.

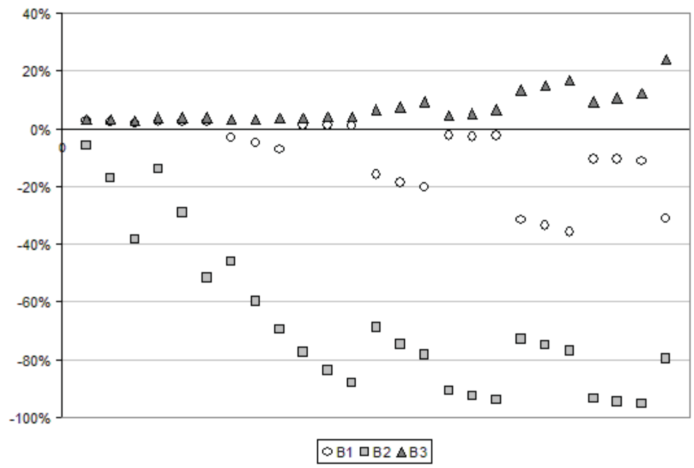

Figure 5.1: Percent change of MSE between Horton's and Proposed methods for three model parameters under each missingness model scenario.

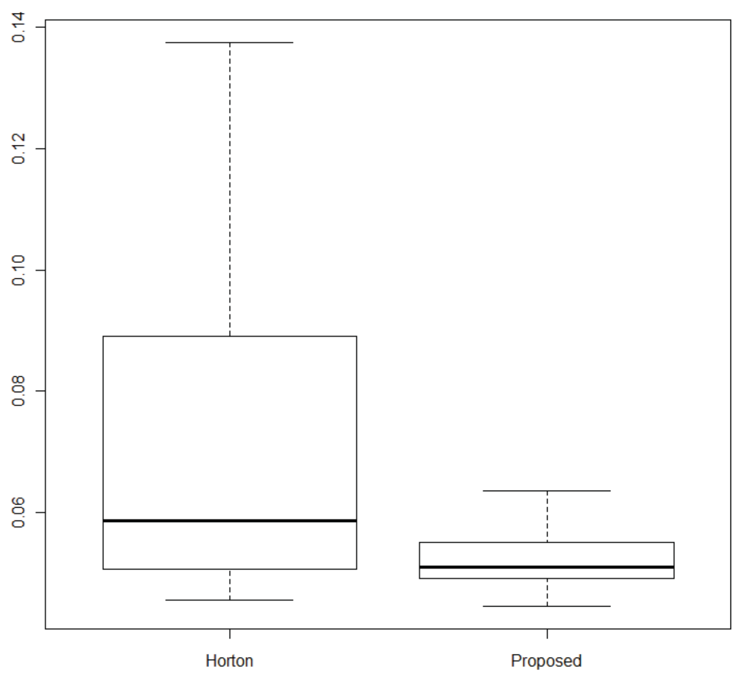

Figure 5.2: Box plot of percent of times the p-value for the goodness of fit test was below 0.05 for each missingness model scenario.

Table 5.2: Percent relative bias for data model parameters for each missingness model scenario.

	$RB(\hat{\beta}_0)$		$RB(\hat{\beta}_1)$		$RB(\hat{\beta}_2)$	
Scenario	Horton's	Proposed	Horton's	Proposed	Horton's	Proposed
1	0.10	0.19	-0.20	0.43	-1.04	-0.82
2	0.85	0.14	-17.68	0.76	-1.07	-0.98
3	1.64	0.15	-35.50	0.87	-0.90	-0.93
4	-0.17	-0.08	-0.44	0.40	-1.81	-1.56
5	0.00	0.04	-22.93	0.67	-1.55	-1.40
6	0.04	0.02	-45.52	0.83	-1.59	-1.57
7	6.44	0.07	-42.39	1.01	-1.00	-1.35
8	7.30	-0.02	-57.53	1.10	-1.06	-1.63
9	8.43	0.15	-71.86	1.05	-0.36	-1.08
10	1.81	-0.10	-87.00	0.42	-1.23	-1.72
11	1.98	-0.16	-109.36	0.37	-1.08	-1.89
12	2.19	-0.17	-131.81	0.12	-0.85	-1.97
13	13.38	0.06	-68.38	0.77	-1.16	-2.11
14	14.61	0.12	-78.36	0.72	-0.80	-1.89
15	15.74	0.07	-86.85	0.85	-0.92	-2.27
16	5.39	-0.25	-154.71	-0.05	-0.34	-2.11
17	5.83	-0.23	-174.88	0.16	0.03	-2.09
18	6.25	-0.26	-195.50	0.69	0.18	-2.52
19	21.09	0.11	-73.84	0.64	-0.73	-2.21
20	22.41	0.18	-77.87	0.84	-0.54	-2.18
21	23.70	0.24	-81.54	0.95	-0.30	-1.92
22	11.39	0.12	-186.39	0.10	0.83	-1.66
23	11.99	0.04	-203.23	0.18	0.89	-2.14
24	12.67	0.03	-219.85	0.15	1.15	-2.29
25	23.69	0.53	-95.60	1.24	0.49	-1.18

Table 5.3: Mean squared errors (MSEŒ100) for data model parameters for each missingness model scenario.

	$MSE(\hat{\beta}_0)$		$MSE(\hat{\beta}_1)$		$MSE(\hat{\beta}_2)$	
Scenario	Horton's	Proposed	Horton's	Proposed	Horton's	Proposed
1	1.82	1.87	0.25	0.24	4.11	4.25
2	1.81	1.86	0.29	0.24	4.09	4.22
3	1.81	1.85	0.38	0.24	4.07	4.20
4	1.80	1.85	0.27	0.23	4.07	4.23
5	1.80	1.84	0.33	0.23	4.06	4.21
6	1.81	1.85	0.48	0.23	4.12	4.28
7	1.92	1.86	0.44	0.24	4.05	4.18
8	1.95	1.86	0.59	0.24	4.09	4.23
9	1.99	1.85	0.78	0.24	4.10	4.25
10	1.82	1.84	1.05	0.24	4.08	4.23
11	1.82	1.84	1.49	0.24	4.08	4.25
12	1.84	1.85	2.04	0.24	4.12	4.30
13	2.33	1.96	0.72	0.22	4.21	4.49
14	2.41	1.96	0.86	0.22	4.25	4.58
15	2.52	2.01	1.01	0.22	4.28	4.67
16	1.91	1.86	2.70	0.24	4.10	4.28
17	1.92	1.87	3.37	0.25	4.13	4.35
18	1.96	1.91	4.15	0.26	4.20	4.48
19	3.00	2.06	0.79	0.21	4.20	4.76
20	3.17	2.10	0.85	0.21	4.24	4.88
21	3.31	2.12	0.91	0.21	4.21	4.91
22	2.24	2.01	3.77	0.24	4.36	4.77
23	2.31	2.06	4.44	0.24	4.42	4.89
24	2.37	2.10	5.15	0.25	4.47	5.02
25	3.38	2.33	1.21	0.25	4.55	5.64

Table 5.4: Size of the goodness of fit test estimated by the percentage rejection of H_0, under H_0, according to the significance level of 0.05

Scenario	Horton's	Proposed
1	0.046	0.051
2	0.054	0.054
3	0.048	0.051
4	0.056	0.063
5	0.046	0.050
6	0.048	0.051
7	0.059	0.046
8	0.065	0.049
9	0.064	0.048
10	0.049	0.056
11	0.048	0.051
12	0.051	0.055
13	0.085	0.053
14	0.088	0.049
15	0.099	0.058
16	0.058	0.055
17	0.056	0.049
18	0.056	0.053
19	0.128	0.064
20	0.130	0.058
21	0.138	0.053
22	0.075	0.045
23	0.089	0.048
24	0.100	0.046
25	0.130	0.062

5.6 Discussion

We proposed a goodness-of-fit test that incorporates the estimation of the missingness model parameters and can be applied to the case of longitudinal data with MAR observations. We assessed the performance of the proposed method versus the ordinary method, that ignores the missingness process (Horton et al. 1999), through the simulation study, under MAR data and examined various intermittent missingness model scenarios.

In terms of the estimation of the data model parameters β_0, β_1 and β_2, we can say that the proposed method performed well. Compared to Horton's method, the resulting mean squared errors and the percent relative biases were considerably smaller for most scenarios; and for the few cases when they were not smaller, the difference, or percent change, between the two methods was not large. It should be noted that while Horton's method resulted in very large percent relative biases of $\hat{\beta}_1$ for some scenarios, the proposed method produced small biases each time. Finally, the proposed goodness-of-fit test had much smaller distortion in test size compared to Horton's test. For scenarios where response rates decreased with time, test sizes were close to the nominal level for both methods. On the other hand, for scenarios with response rates larger at $t = 3$ compared to $t = 2$, Horton's method performed poorly in terms of test size, while the test size based on the proposed method was always reasonable and closer to the nominal level.

Based on these results, we can conclude that the proposed goodness-of-fit method that incorporates the missingness process should be used when dealing with MAR of intermittent pattern longitudinal data.

5.7 Appendix

5.7.1 Choice of a method for generating independent binary outcomes

There exist several methods for generating longitudinally correlated binary data, for example, we refer to Bahadur (1961), Kanter (1975), Prentice (1988), Oman and Zucker (2001), and Qaqish (2003). Farrell and Rogers-Stewart (2008) as well as a recent working paper by Preisser and Qaqish (2012) offer a critical review of the various methods that have been established for obtaining longitudinally correlated binary data. Use of any method is subject to general restrictions on marginal means and pairwise marginal correlations that cannot be violated, and so, from the definition of probability, it follows that

$$\max\left(-\varphi_{it_1}\varphi_{it_2}, -\left(\varphi_{it_1}\varphi_{it_2}\right)^{-1}\right) \leq \rho_{it_1t_2} \leq \min\left(\varphi_{it_1}/\varphi_{it_2}, \varphi_{it_2}/\varphi_{it_1}\right) \tag{5.41}$$

where $\rho_{it_1t_2} = corr(y_{it_1}, y_{it_2})$ and $\varphi_{it} = \sqrt{p_{it}/(1-p_{it})}$. Other restrictions on the ranges of these parameters may be imposed by a specific data-generating method or result from the requirement that the correlation matrix is positive definite (Preisser and Qaqish, 2012).

Prior to our simulation study, we generated two sets of correlated binary responses (y_{i1}, y_{i2}, y_{i3}), using the Bahadur's model (1961) and the method of Emrich and Piedmonte (1991), and compared the performance of the proposed WGEE method to estimate parameters for model (5.37) under the two settings. We assumed a constant correlation $\rho_{it_1t_2} = \rho$, and a third order association parameter $\rho_{it_1t_2t_3}$ =0 for both methods.

The Bahadur's representation is often employed in textbook examples or small simulation studies when T is not large and all coefficients of order three and higher can be ignored, however, it becomes computationally difficult otherwise. Assuming three measurements per subject, the model can be expressed as

$$f(y_i) = \left(\prod_i p_{it}^{y_{it}} \, (1\text{-}p_{it})^{1-y_{it}}\right) \times (1+\rho_{i12}z_{i1}z_{i2}+\rho_{i13}z_{i1}z_{i3}+\rho_{i23}z_{i2}z_{i3}+\rho_{i123}z_{i1}z_{i2}z_{i3})$$

with the marginal probabilities $p_{it} = E(y_{ij}) = P(y_{ij} = 1)$ and standardized deviations $z_{it} = (y_{it} - p_{it})/\sqrt{p_{it}(1-p_{it})}$. The Emrich and Piedmonte (EP) technique for obtaining correlated binary data is based on the multivariate probit model that generates correlated standard normal variables, which are then dichotomized. The algorithm requires specification of only the vector of marginal means and a matrix of pairwise correlations of the multivariate binary distribution and allows for specification of unequal means, arbitrary correlation matrix and negative correlations. We used the R programming package, called *mvtBinaryEP*, to implement the EP method.

We assessed the results in terms of the average estimates and relative biases based on the response data generated by the Bahadur and EP methods. The results were generally good and we could not conclude that either of the methods performed consistently better or worse. However, since the EP-based simulations took significantly more computer time to process compared to the Bahadur's approach, we decided to use the Bahadur's representation in our simulation study.

5.7.2 Goodness-of-Fit Test Deficiencies

It is known that the Hosmer-Lemeshow goodness-of-fit test has some disadvantages, Hosmer et al. (1997) studied the performance of the tests under different settings. The major deficiencies are that the test statistic is sensitive to the choice of cutpoints that define the groups and the number of groups considered. Studies indicate that a goodness-of-fit test statistic based on less than 6 groups usually results in very low power and could falsely indicate that the model is good. Hosmer et al. (1997) used different statistical packages to fit the same data set and obtained the same estimated model parameters, but different p-values ranging from 0.02 to 0.16. Kuss (2002) pointed out that observations belonging to the same group may have considerably different covariate values. This was addressed by Pulkstenis and Robinson (2002) who proposed a two-stage procedure that requires categorical and continuous covariates in the model.

In our study, we have chosen the score statistic for the proposed goodness-of-fit test because it is based only on $\hat{\beta}$ under the null model, while for example, the Wald statistic would require the estimate of γ under the alternative model. Horton et al. (1999) compared the score to the Wald statistic and made remarks about its better small-sample performance as well its availability when the algorithm in the alternative model does not coverage. In our simulation study, we applied the same strategy for obtaining the cut-off points for all cases; therefore, while the results could be different if other algorithms or software packages were used, they are consistent and comparable within the scope of the simulation study.

Chapter 6

Mixed Discrete and Continuous Outcomes Model with Missing Data

6.1 Introduction

Multiple outcomes of mixed, discrete and continuous, nature are commonly collected to assess impacts of various interventions including medical treatments or government policies. For example, in the study on quality of care for schizophrenia patients (Dickey et al., 2003) the binary outcome identified patients on medication, and the continuous outcome was a self-reported quality of interpersonal interactions between a patient and a clinician. A common approach to study data with multiple outcomes is to model each outcome separately as a function of covariates of interest. However, since multiple outcomes may be correlated, proper joint analysis based on multivariate methods should be considered. The challenges in modeling mixed outcomes simultaneously are to obtain the joint probability directly and to estimate the intra-subject correlation. To overcome these challenges, several modeling strategies have been proposed in the literature and can be broadly divided into three classes. The first, conditional modeling, approach is to factorize the likelihood as a product of marginal and conditional distributions. This modeling class avoids direct specification of the joint distribution and has been extended to accommodate covariates (Cox and Wermuth, 1992) and clustered data (Fitzmaurice and Laird, 1995; Regan and Catalano, 1999). Extensions to higher dimensions involve assumptions on large covariance structures and high order associations which may be problematic and are a drawback for this class of models. The second class of methods uses latent variables to model the dependence structure of mixed discrete and continuous outcomes data (Sammel, et al., 1997). A disadvantage of this method is its non-robustness to misspecification of the covariance (Sammel et al., 1999). Finally, the third approach is based on an extension of Liang and Zeger's (1986) generalized estimating equations (GEE) (Prentice and Zhao, 1991). The advantage of the GEE method is its lack of distributional assumptions and robustness to misspecification of the correlation between outcomes which leads to less efficient but more robust estimates. Teixeira-Pinto and Normand (2009) provide a summary of these approaches and implement the GEE approach.

Incomplete data can arise due to missing measurements (or in case of longitudinal studies, presence of time-varying covariates) and cause bias or lead to inefficient analyses. Statistical methods that address missingness, when both categorical and continuous random variables are involved, include imputation, likelihood, and weighting approaches. Fitzmaurice and Laird (1997) considered a multivariate extension of the model proposed by Fitzmaurice and Laird (1995) and demonstrated large gains in efficiency from a multivariate approach. They also proposed the EM-algorithm to fit the extension of the general location model in the presence of missing data. Teixeira-Pinto and Normand (2011) extended Robins et al.'s (1995) weighted GEE to multiple mixed outcomes for missing at random (MAR) data.

In this chapter, we study the properties of the conditional mixed discrete and continuous outcomes model and apply the likelihood method to MAR data. Specifically, we compare the performance of estimation based on a joint model for the mixed outcomes with estimation based on modeling the binary and continuous outcomes separately when all data are observed, and under a scenario with binary data missing at random. This chapter is organized as follows: in Section 6.2, we present the conditional model in the context of fully-observed data; in Section 6.3, we describe the likelihood method and extend it to the case with missing data in Section 6.4. A small simulation study for cross-sectional data is described in Section 6.5 with conclusions in Section 6.6.

6.2 Mixed Outcomes Model

Suppose data consist of mixed binary and continuous outcomes. For each subject $i = 1, ..., n$, we consider a set of B binary response variables $Y_i = (Y_{1i}, ..., Y_{Bi})'$ with covariates x_{1i} and a set of C continuous response variables $Z_i = (Z_{1i}, ..., Z_{Ci})'$ with covariates x_{2i}. Note that a set of covariates that is common to the binary and continuous outcomes could also be considered. We begin by defining the marginal means and covariance matrices for the vector of binary outcomes Y_i and the vector of continuous outcomes Z_i. We then develop the joint model based on factorization of the joint distribution into a marginal component for discrete outcomes, and a conditional distribution for continuous outcomes, given the discrete outcomes.

6.2.1 Marginal Model for Binary Outcomes

We assume that each of the binary responses, Y_{bi} $(b = 1, ..., B)$, follows a Bernoulli distribution with success probability $p_{bi} = E(Y_{bi}|x_{1i}) = P(Y_{bi} = 1|x_{1i})$ and use a marginal logistic regression model with parameters β_b to describe the relationship between Y_{bi} and the covariates x_{1i}, that is,

$$\text{logit}(p_{bi}) = \log\left(\frac{p_{bi}}{1 - p_{bi}}\right) = x'_{1i}\beta_b, \tag{6.1}$$

where β_b is a vector of unknown regression coefficients which may be different for each binary outcome b, and p_{bi} is derived from the marginal density $f(y_{bi}|x_{1i}) = p_{bi}^{y_{bi}}(1 - p_{bi})^{1-y_{bi}}$. We use Bahadur's expansion (Bahadur, 1961) to represent the joint distribution of binary outcomes $Y_i = (Y_{1i}, ..., Y_{Bi})'$, that is

$$\begin{aligned} f(y_i|x_{1i}) &= \left(\prod_{b=1}^{B} p_{bi}^{y_{bi}}(1 - p_{bi})^{1-y_{bi}}\right) \\ &\quad \left(1 + \sum_{j<k} \rho_{jk,i} v_{ji} v_{ki} + \sum_{j<k<l} \rho_{jkl,i} v_{ji} v_{ki} v_{li} + ... + \rho_{1...B,i} v_{1i} ... v_{Bi}\right), \end{aligned}$$

where

$$v_{ji} = (y_{ji} - p_{ji}) \{p_{ji} (1 - p_{ji})\}^{-0.5},$$

$$\begin{aligned} \rho_{jk,i} &= corr(Y_{ji}, Y_{ki}) = \frac{E\left(\left(Y_{ji} - p_{ji}\right)\left(Y_{ki} - p_{ki}\right)|x_{1i}\right)}{\{p_{ji}(1-p_{ji})\,p_{ki}(1-p_{ki})\}^{1/2}} \\ &= E(v_{ji}v_{ki}), \end{aligned}$$

$\rho_{1...B,i} = E(v_{1i}...v_{Bi})$, $j < k < l = 1, ..., B$, and $i = 1, ..., n$.

Let us denote the expectation of Y_i by

$$\mu_{1i}(\beta) = E(Y_i|x_{1i}) = (p_{1i}, ..., p_{Bi})' \tag{6.2}$$

which is a $B \times 1$ vector with elements $\mu_{1bi} = p_{bi}$, $b = 1, ..., B$, and the $B \times B$ covariance matrix of Y_i by

$$\Sigma_{1i}(\tau) = Var(Y_i|x_{1i}), \tag{6.3}$$

with τ being the association parameters of Y_i.

6.2.2 Marginal Model for Continuous Outcomes

We assume that the continuous outcomes $Z_i = (Z_{1i}, ..., Z_{Ci})'$ are distributed according to the multivariate normal density

$$Z_i \sim MVN(\mu_{2i}, \Sigma_{2i}(\lambda)), \tag{6.4}$$

where $\mu_{2i} = E(Z_i|x_{2i})$ is the expectation of Z_i which is a $C \times 1$ vector with elements

$$\mu_{2ic}(\alpha) = x'_{2i}\alpha_c, \tag{6.5}$$

where α_c is a vector of unknown regression coefficients, which may be different for each binary outcome $c = 1, ..., C$, and $\Sigma_{2i}(\lambda) = Var(Z_i|x_{2i})$ is a $C \times C$ covariance matrix of Z_i defined as a function of association parameters λ.

6.2.3 Factorization Model for Binary and Continuous Outcomes

For the remainder of this chapter, we assume that the same covariate vector x_i predicts both Y_i and Z_i. Following the work of Fitzmaurice and Laird (1997), to account for the association between the responses, we express the joint distribution $f(y_i, z_i|x_i)$ as a product of a marginal distribution for the binary outcomes Y_i and a conditional distribution for the continuous outcomes Z_i given Y_i

$$f(y_i, z_i|x_i) = f(y_i|x_i)f(z_i|y_i, x_i). \tag{6.6}$$

Let

$$\Gamma = \Gamma(\gamma) = Cov(Y_i, Z_i|x_i) \tag{6.7}$$

denote the $(C \times B)$ matrix of association parameters between Y_i and Z_i. Large absolute values of components of Γ indicate strong correlation. If Γ=0 the binary and continuous outcomes are independent. We model the conditional density $f(z_i|y_i, x_i)$

as a multivariate normal, that is for $i = 1, ..., n$, $c = 1, ..., C$,

$$f(z_i|y_i, x_i) = \frac{|\Sigma_{21i}|^{-0.5} \exp\left\{-0.5(z_i - \mu_{21i})'\Sigma_{21i}^{-1}(z_i - \mu_{21i})\right\}}{(2\pi)^{C/2}},$$

where μ_{21i} is a vector of conditional means, and Σ_{21i} denotes the conditional variance, which are respectively given by

$$\mu_{21i} = E\left(Z_i|Y_i, x_i\right) = \mu_{2i} + \Gamma(Y_i - \mu_{1i}), \tag{6.8}$$

and

$$\begin{aligned} \Sigma_{21i} &= Var\left(Z_i|Y_i, x_i\right) \\ &= Var\left(Z_i|x_i\right) - Var\left(E\left(Z_i|Y_i, x_i\right)\right) \\ &= \Sigma_{2i} - \Gamma\Sigma_{1i}\Gamma' \end{aligned} \tag{6.9}$$

where $|\Sigma_{21i}|$ denotes the determinant of $\Sigma_{21i}, i = 1, ..., n$. Note that under this setting we have

$$E\left(Z_i|x_i\right) = E(E\left(Z_i|Y_i, x_i\right)) = \mu_{2i} + \Gamma E(Y_i - \mu_{1i}) = \mu_{2i}, \tag{6.10}$$

and

$$\begin{aligned} Var\left(Z_i|x_i\right) &= E(Var\left(Z_i|Y_i, x_i\right)) + Var(E\left(Z_i|Y_i, x_i\right)) \\ &= \Sigma_{2i} - \Gamma\Sigma_{1i}\Gamma' + \Gamma Var(Y_i)\Gamma' = \Sigma_{2i}. \end{aligned} \tag{6.11}$$

Hence, for the ordered discrete and continuous data $(Y_{1i}, ..., Y_{Bi}, Z_{1i}, ..., Z_{Ci})$, $i =$

$1, .., n$, we can construct the following matrix of association parameters τ_{jk}, λ_{lm}, and γ_{jl} with $j, k = 1, ..., B$ and $l, m = 1, ..., C$

$$\begin{bmatrix} \Sigma_{1i} & \Gamma_i' \\ \Gamma_i & \Sigma_{2i} \end{bmatrix} = \begin{bmatrix} \tau_{11,i} & \cdots & \tau_{1B,i} & \gamma_{11,i} & \cdots & \gamma_{1C,i} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tau_{B1,i} & \cdots & \tau_{BB,i} & \gamma_{B1,i} & \cdots & \gamma_{BC,i} \\ \gamma_{11,i} & \cdots & \gamma_{1B,i} & \lambda_{11,i} & \cdots & \lambda_{1C,i} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{C1,i} & \cdots & \gamma_{CB,i} & \lambda_{C1,i} & \cdots & \lambda_{CC,i} \end{bmatrix}. \tag{6.12}$$

Let $\beta' = (\beta_1, ..., \beta_B)$, $\alpha' = (\alpha_1, ..., \alpha_C)$, $\tau_i' = (\tau_{jk} : j, k = 1, ..., B)$, $\lambda_i' = (\lambda_{lm} : l, m = 1, ..., C)$, and $\gamma_i' = (\gamma_{jl} : j = 1, ..., B, l = 1, ..., C)$.The goal of this work is to estimate the regression parameters (β', α') and the association parameters $(\tau_i', \lambda_i', \gamma_i'), i = 1, ..., n$. The novelty of this model is the extension of Fitzmaurice and Laird (1995) model with the inclusion of Bahadur's model and the conditional covariance matrix Σ_{21i} that is compatible with the marginal model for Z_i.

6.3 Likelihood Equation and Parameter Estimation for Complete Data

For illustrative purposes, we consider a simple case of a response vector consisting of two binary ($B = 2$) and two continuous outcomes ($C = 2$). The (4×1) vector of mixed response variables for subject i can be written as $(Y_i, Z_i)' = (Y_{1i}, Y_{2i}, Z_{1i}, Z_{2i})'$,

$i = 1, ..., n$. Assuming single covariates x_i, we define

$$\mu_{1i} = \begin{bmatrix} p_{1i} \\ p_{2i} \end{bmatrix} = \begin{bmatrix} \exp\left(\beta_{10} + x_i\beta_{11}\right) / \left(1 + \exp\left(\beta_{10} + x_i\beta_{11}\right)\right) \\ \exp\left(\beta_{20} + x_i\beta_{21}\right) / \left(1 + \exp\left(\beta_{20} + x_i\beta_{21}\right)\right) \end{bmatrix}, \tag{6.13}$$

$$\mu_{2i} = \begin{bmatrix} \alpha_{10} + x_i\alpha_{11} \\ \alpha_{20} + x_i\alpha_{21} \end{bmatrix}, \tag{6.14}$$

and we have the following matrix of association parameters

$$\begin{bmatrix} \Sigma_{1i} & \Gamma_i' \\ \Gamma_i & \Sigma_2 \end{bmatrix} = \begin{bmatrix} \tau_{11,i} & \tau_{12,i} & \gamma_{11} & \gamma_{12} \\ \tau_{21,i} & \tau_{22,i} & \gamma_{21} & \gamma_{22} \\ \gamma_{11} & \gamma_{12} & \lambda_{11} & \lambda_{12} \\ \gamma_{21} & \gamma_{22} & \lambda_{21} & \lambda_{22} \end{bmatrix} \tag{6.15}$$

For simplicity, we assume that $\Gamma_i = \Gamma$ and $\Sigma_{2i} = \Sigma_2$ for all subjects $i = 1, ..., n$. Note that Σ_{1i} is not necessarily the same as Σ_{1j} for $i \neq j$ since $\tau_{11,i} = p_{1i}\left(1 - p_{1i}\right), \tau_{22,i} = p_{2i}\left(1 - p_{2i}\right)$ and $\tau_{12,i} = \tau_{21,i} = \rho\sqrt{\tau_{11,i}\tau_{22,i}}$ are all functions of p_{1i} and p_{2i} which are generated independently for each i. We also assume that $\lambda_{12} = \lambda_{21} = \zeta\sqrt{\lambda_{11}\lambda_{22}}$.

Let

$$\theta = \left(\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \rho, \lambda_{11}, \zeta, \lambda_{22}, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}\right)' \tag{6.16}$$

be the vector containing all the association and regression parameters. We write the factorization model (6.6) for $(y_i, z_i) = (y_{1i}, y_{2i}, z_{1i}, z_{2i})$ as a product of two-outcome

Bahadur representation, and a conditional bivariate-normal distribution as follows

$$\begin{aligned} f(y_i, z_i|x_i, \theta) &= f(y_{1i}, y_{2i}|x_i) f(z_{1i}, z_{2i}|y_{1i}, y_{2i}, x_i) \\ &= \left[p_{i1}^{y_{1i}}(1-p_{i1})^{1-y_{1i}} p_{i2}^{y_{2i}}(1-p_{i2})^{1-y_{2i}}\left(1+\rho v_{1i}v_{2i}\right)\right] \\ &\quad \times \left[(2\pi)^{-1}|\Sigma_{21i}|^{-0.5}\exp\left(-0.5(z_i-\mu_{21i})'\Sigma_{21i}^{-1}(z_i-\mu_{21i})\right)\right], \end{aligned}$$

where $\rho = \rho_{12}$, $\Sigma_{21i} = \Sigma_{2i} - \Gamma_i\Sigma_{1i}\Gamma_i'$ is a (2×2) conditional variance, $\mu_{21i} = \mu_{2i} + \Gamma_i(Y_i - \mu_{1i})$ is a (2×1) vector of conditional means. The likelihood equation is

$$L(\theta|y, z, x) = \prod_{i=1}^{n} f(y_i|x_i) f(z_i|y_i, x_i), \tag{6.17}$$

where $y = (y_1, .., y_n)$ with $y_i = (y_{1i}, y_{2i})$, similarly $z = (z_1, .., z_n)$ with $z_i = (z_{1i}, z_{2i})$, $i = 1, ..., n$, also $x = (x_1, .., x_n)$. The corresponding log-likelihood function is

$$\begin{aligned} l(\theta|y, z, x) &= \sum_i \log f(y_i|x_i) + \sum_i \log f(z_i|y_i, x_i) \qquad (6.18) \\ &= \sum_i \log\left\{p_{1i}^{y_{1i}}(1-p_{1i})^{1-y_{1i}} p_{2i}^{y_{2i}}(1-p_{2i})^{1-y_{2i}}\left(1+\rho v_{1i}v_{2i}\right)\right\} \\ &\quad + \sum_i \log\left\{(2\pi)^{-1}|\Sigma_{21i}|^{-0.5}\exp\left(-0.5(z_i-\mu_{21i})'\Sigma_{21i}^{-1}(z_i-\mu_{21i})\right)\right\}. \end{aligned}$$

The maximum likelihood estimates of the regression and association parameters θ, defined by (6.16), can be obtained using the following Newton-Raphson iterative procedure

$$\theta^{(k+1)} = \theta^{(k)} - \left[\frac{\partial^2 l}{\partial\theta^2}\right]^{-1}\Bigg|_{(\theta^{(k)})} \left(\frac{\partial l}{\partial\theta}\right)\Bigg|_{(\theta^{(k)})}, \tag{6.19}$$

where $\partial l/\partial\theta$ is a 16×1 vector of first order derivatives, $\left[\partial^2 l/\partial\theta^2\right]$ is a 16×16 Hessian matrix of second order derivatives, and k denotes the iteration number.

Note that all the parameters need to be processed simultaneously. Let us denote

$$\begin{aligned} h_i &= \rho v_{1i} v_{2i} \\ &= \frac{E\left(\left(y_{1i}-p_{1i}\right)\left(y_{2i}-p_{2i}\right)\right)}{\left\{p_{1i}\left(1-p_{1i}\right)p_{2i}\left(1-p_{2i}\right)\right\}}\left(y_{1i}-p_{1i}\right)\left(y_{2i}-p_{2i}\right) \\ &= \frac{\tau_{12i}}{\tau_{11i}\tau_{22i}}\left(y_{1i}-p_{1i}\right)\left(y_{2i}-p_{2i}\right), \end{aligned}$$

so that we can write

$$\begin{aligned} l(\theta|y,z,x) &= \sum_i [y_{1i}\log p_{1i} + (1-y_{1i})\log(1-p_{1i}) + y_{2i}\log p_{2i} + \\ &\quad + (1-y_{2i})\log(1-p_{2i}) + \log\left(1+h_i\right) + \\ &\quad - \log(2\pi) - 0.5\log\left(|\Sigma_{21i}|\right) - 0.5(Z_i-\mu_{21i})'\Sigma_{21i}^{-1}(Z_i-\mu_{21i})]. \end{aligned} \tag{6.20}$$

Let us denote $\dfrac{\partial l\left(\theta|y,z,x,w\right)}{\partial\theta} = \sum_i \dfrac{\partial l_i}{\partial\theta}$. The structure of partial derivatives with respect to the regression parameters is as shown below

$$
\begin{aligned}
\frac{\partial l_i}{\partial \beta_{jk}} &= (y_{ji} - p_{ji})\frac{\partial b_{ji}}{\partial \beta_{jk}} + \frac{1}{1+h_i}\frac{\partial h_i}{\partial \beta_{jk}} - 0.5\mathrm{tr}\left(\Sigma_{21i}^{-1}\frac{\partial \Sigma_{21i}}{\partial \beta_{jk}}\right) \\
&+ (z_i - \mu_{21i})'\Sigma_{21i}^{-1}\frac{\partial \mu_{21i}}{\partial \beta_{jk}} - 0.5(z_i - \mu_{21i})'\frac{\partial \Sigma_{21i}^{-1}}{\partial \beta_{jk}}(z_i - \mu_{21i}),
\end{aligned}
$$

where $j = 1, 2$, $k = 0, 1$ and $b_{ji} = \beta_{j0} + x_i\beta_{j1}$. Also,

$$
\frac{\partial l_i}{\partial \alpha_{jk}} = (z_i - \mu_{21i})'\Sigma_{21i}^{-1}\frac{\partial \mu_{21i}}{\partial \alpha_{jk}}.
$$

The partial derivatives for the association parameters are structured as follows

$$
\frac{\partial l_i}{\partial \rho} = \frac{1}{1+h_i}\frac{\partial h_i}{\partial \rho} - 0.5\mathrm{tr}\left(\Sigma_{21i}^{-1}\frac{\partial \Sigma_{21i}}{\partial \rho}\right) - 0.5\left(z_i - \mu_{21i}\right)'\frac{\partial \Sigma_{21i}^{-1}}{\partial \rho}(z_i - \mu_{21i}),
$$

$$
\frac{\partial l_i}{\partial \lambda_{lm}} = -0.5\mathrm{tr}\left(\Sigma_{21i}^{-1}\frac{\partial \Sigma_{21i}}{\partial \lambda_{lm}}\right) - 0.5(z_i - \mu_{21i})'\frac{\partial \Sigma_{21i}^{-1}}{\partial \lambda_{lm}}(z_i - \mu_{21i}),
$$

and

$$
\frac{\partial l_i}{\partial \gamma_{jl}} = -0.5\mathrm{tr}\left(\Sigma_{21i}^{-1}\frac{\partial \Sigma_{21i}}{\gamma_{jl}}\right) + (Z_i - \mu_{21i})\Sigma_{21i}^{-1}\frac{\partial \mu_{21i}}{\partial \gamma_{jl}} - 0.5(Z_i - \mu_{21i})'\frac{\partial \Sigma_{21i}^{-1}}{\partial \gamma_{jl}}(Z_i - \mu_{21i}),
$$

where $j, l, m = 1, 2$.

The detailed derivations of parameter-level derivatives and matrix differentiations are straightforward. The second order derivatives can be obtained in a similar manner and involve compound expressions which are not shown here. Note that the difficulty of this computation lies mainly in the high chance of making errors when deriving and recording these lengthy expressions.

6.4 Likelihood Equation and Parameter Estimation for Missing Data

In this section we extend the proposed factorization model to handle missing at random (MAR) data using maximum likelihood method. We assume that missing data can occur only among the binary variables Y_i, while the continuous variables Z_i and the covariates x_i are fully observed. Under MAR setting, the missingness process depends only on the observed data, and the MLE of θ in (6.16), can be obtained by maximizing the observed data likelihood obtained by summing over the missing data. Suppose the binary data can be partitioned into its observed (denoted by superscript o) and missing (denoted by superscript m) components: $Y_i = (Y_i^o, Y_i^m)$. The maximum likelihood estimate of θ is the vector $\hat{\theta}$ that maximizes

$$L(\theta|y^o, z, x) \propto \prod_{i=1}^{n} \sum_{y^m} f(y_i|x_i, \theta) f(z_i|y_i, x_i, \theta). \tag{6.21}$$

To be more explicit, we revert to the simple case of the outcome vector consisting of two binary and two continuous variables defined in Section 6.2. Let $R_{bi} = 1$ when Y_{bi} is observed, and $R_{bi} = 0$ when Y_{bi} is missing, for $b = 1, 2$ and $i = 1, ..., n$. Note that with two binomial variables, we have four possible missingness scenarios for each individual i: $(r_{1i}, r_{2i}) = (1, 1)$ when data are observed, $(r_{1i}, r_{2i}) = (1, 0)$ or $(r_{1i}, r_{2i}) = (0, 1)$ when one binary outcome is missing and the other observed, and

$(r_{1i}, r_{2i}) = (0, 0)$ when both binary outcomes are missing.

To account for the missing data, we expand the likelihood equation (6.17) by summing over all possible values y_{2i}, $y_{1i} = 0, 1$ of the binary variables with missing data. That is, the MAR data likelihood, can be written as

$$L(\theta|y^o, z, x, w) = \prod_{i=1}^{n} L_{1i} \times L_{2i} \times L_{3i} \times L_{4i} \tag{6.22}$$

where, respectively with the four missingness scenarios,

$$L_{1i} = (f(y_i|x_i)f(z_i|y_i, x_i))^{r_{1i}r_{2i}},$$

$$L_{2i} = \left(\sum_{y_{i1}} f(y_i|x_i)f(z_i|y_i, x_i)\right)^{(1-r_{1i})r_{2i}},$$

$$L_{3i} = \left(\sum_{y_{i2}} f(y_i|x_i)f(z_i|y_i, x_i)\right)^{r_{1i}(1-r_{2i})},$$

$$L_{4i} = \left(\sum_{y_{i1}} \sum_{y_{i2}} f(y_i|x_i)f(z_i|y_i, x_i)\right)^{(1-r_{1i})(1-r_{2i})}.$$

The Newton-Raphson algorithm can be used to obtain the maximum likelihood estimates of θ with data missing at random. Note that with large number of outcomes and complex data models, the implementation of this approach may be extremely computational intensive.

6.5 Simulation Study

A small simulation study was conducted to investigate if joint modelling improves efficiency of the estimators. We compared the joint model versus a separate model in terms of bias, mean squared error and standard error of the estimators of θ under the scenarios of fully observed and MAR binary data.

We generated data based on the factorization model described in Section 6.2, with $B = 2$ and $C = 2$.

The following values were used for regression parameters $(\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}) = (-0.4, 0.1, -0.5, 0.25)$ and $(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})' = (1.0, 1.0, 0.5, 0.25)$. Further, we set $\rho = 0.2$ and the continuous data association parameters to $(\lambda_{11}, \zeta, \lambda_{22})'$=(1.0,0.5,1.0)$'$. As to the association between binary and continuous outcomes, we considered different values for Γ to test different strengths of association (0.2 for weak, 0.5 for medium and 0.9 for strong) and but reported the results based on $\gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22} = 0.5$ since the different values had no notable effect on the performance of the joint model. The covariates corresponding to binary outcomes x_{1i} were generated independently from the standard uniform $U(0, 1)$ distribution and those corresponding to continuous outcomes x_{2i} were generated independently from $N(0, 1)$. Note that we considered a data model with different set of covariates for the binary and continuous outcomes as suggested by Teixeira-Pinto and Normand (2009) who showed that when the same covariate was associated with the binary and continuous outcomes, their multivariate models produced estimates with MSEs identical to the univariate model that ignored correlation between the outcomes. However, when the outcomes shared a different set of covariates, the efficiency gain was higher.

Under the complete-data setting, the models (6.1), (6.4), and (6.6) were all fitted based on the same sets of generated data.

In order to test the models under the MAR scenario, we modeled the missingness processes as logistic regressions, that is, $\pi_{ji} = P(R_{ji} = 1|z_i, x_{1i}, x_{2i})$, $j = 1, 2$, with $\text{logit}(\pi_{1i}) = 1+x_{1i}+z_{1i}-1.5z_{2i}$, and $\text{logit}(\pi_{2i}) = 1+0.5x_{1i}+z_{1i}-2z_{2i}$. The parameters were chosen to obtain approximately 20% of missing data for Y_1 and 30% of missing data for Y_2.

We considered sample sizes $n = 200$ and $n = 500$, and obtained results from $K = 1000$ simulations. The simulations were programmed in R with function OPTIM based on quasi-Newton optimization method used to obtain the MLEs. The results were reported in terms of the mean square error

$$MSE(\hat{\eta}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\eta}_k - \eta)^2,$$

the average standard error, $SE(\hat{\eta})$, obtained from the Hessian matrix, and the percent relative bias defined as

$$RB(\hat{\eta}) = \frac{1}{K}\sum_{k=1}^{K}\frac{(\hat{\eta}_k - \eta)}{\eta} \times 100\%.$$

In the above equations, $\hat{\eta}_k$ represents a parameter estimate based on data from k^{th} simulation, with $k = 1, ..., K$.

6.5.1 Results

Tables 6.1 and 6.2 display the mean square error (MSE), standard error (SE) percent relative bias (RB) for model parameters under respectively, fully-observed and MAR data scenarios. The resulting estimates of the binary parameters are also presented graphically as box plots in Figures 6.1 and 6.2.

We observe that both sets of results, those based on the fully-observed and those based on MAR data, were consistent in terms of MSE and SE. Under full response, the $MSEs$ and SEs for the binary parameter estimates generated from the joint model were only slightly better compared to their counterparts generated from the separate model. Under the MAR data scenario, the joint model performed notably better as compared to the separate model, especially with larger sample size. For example the MSE for $\hat{\beta}_{20}$ was 0.191 under the separate model, versus 0.046 under the joint model.

In terms of the random bias, we observed some unusual patterns. For example, under the separate model with fully-observed data, $\hat{\beta}_{11}$ had a large RB of 21.46 with sample of size 200, which got reduced to 6.65 when sample of 500 was used. In this case, we could say that the large bias was caused by a small sample. However, this explanation does not apply to the MAR data scenario, as for example, $\hat{\beta}_{20}$ generated RB of 74.01 for $n = 200$ and RB of 74.40 for $n = 500$. Generally, the random bias was more stable for scenarios with sample size $n = 500$. Also overall, the joint model led to smaller RB as compared to the separate model.

As visualized in Figures 6.1 and 6.2, the range of estimates produced in the simulations by the joint model was narrower compared to estimates based on the separate model. Under full response, the overall patterns of resulting estimates were similar for the two models; however, in the presence of MAR data, the joint model performed better, which is clearly shown by the box plots of $\hat{\beta}_{20}$.

Table 6.1: Mean squared error (MSE), standard error (SE) and relative bias (RB) with fully observed data, for sample sizes n, averaged over 1000 simulations.

Est.	n=200						n=500					
	MSE		SE		RB		MSE		SE		RB	
	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint
$\hat{\beta}_{10}$	0.089	0.089	0.290	0.278	0.94	1.17	0.032	0.030	0.183	0.175	1.28	1.02
$\hat{\beta}_{11}$	0.266	0.243	0.502	0.472	21.46	23.71	0.097	0.086	0.316	0.297	6.65	5.87
$\hat{\beta}_{20}$	0.091	0.084	0.292	0.279	0.72	0.85	0.034	0.030	0.184	0.176	0.15	-0.26
$\hat{\beta}_{21}$	0.256	0.227	0.504	0.474	3.10	3.68	0.102	0.086	0.317	0.298	3.00	0.89
$\hat{\alpha}_{10}$	0.005	0.005	0.070	0.070	0.49	-0.78	0.002	0.002	0.044	0.044	-0.16	0.10
$\hat{\alpha}_{11}$	0.005	0.005	0.070	0.070	0.50	0.39	0.002	0.002	0.045	0.045	0.01	-0.08
$\hat{\alpha}_{20}$	0.005	0.004	0.071	0.065	0.16	0.03	0.002	0.002	0.045	0.041	-0.04	0.15
$\hat{\alpha}_{21}$	0.005	0.005	0.070	0.070	0.76	0.70	0.002	0.002	0.045	0.045	0.26	0.20
$\widehat{\rho}$	0.005	0.004	0.071	0.065	-0.50	-1.28	0.002	0.002	0.045	0.041	-1.03	-0.24
$\hat{\lambda}_{11}$	0.009	0.010	0.099	0.098	-1.35	-1.27	0.004	0.004	0.063	0.063	-0.35	-0.33
$\widehat{\zeta}$	0.008	0.009	0.099	0.098	-1.26	-1.20	0.004	0.004	0.063	0.063	-0.33	-0.26
$\hat{\lambda}_{22}$	0.003	0.003	0.053	0.053	-0.50	-0.51	0.001	0.001	0.034	0.033	-0.01	0.05

Table 6.2: Mean squared error (MSE), standard error (SE) and relative bias (RB) with data missing at random for sample sizes n, averaged over 1000 simulations.

Est.	n=200						n=500					
	MSE		SE		RB		MSE		SE		RB	
	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint
$\hat{\beta}_{10}$	0.124	0.120	0.336	0.326	-4.61	1.74	0.045	0.043	0.211	0.205	-2.24	1.48
$\hat{\beta}_{11}$	0.351	0.315	0.571	0.540	15.50	28.51	0.130	0.117	0.358	0.339	14.82	11.87
$\hat{\beta}_{20}$	0.291	0.141	0.376	0.352	74.01	0.59	0.191	0.046	0.235	0.221	74.40	0.70
$\hat{\beta}_{21}$	0.438	0.349	0.635	0.578	29.16	4.68	0.163	0.128	0.397	0.362	38.22	6.19
$\hat{\alpha}_{10}$	0.008	0.008	0.093	0.091	-1.81	-1.12	0.003	0.003	0.059	0.057	-1.83	-0.33
$\hat{\alpha}_{11}$	0.005	0.005	0.070	0.070	0.53	0.39	0.002	0.002	0.045	0.045	0.01	0.05
$\hat{\alpha}_{20}$	0.005	0.005	0.071	0.067	0.16	0.15	0.002	0.002	0.045	0.042	-0.04	0.02
$\hat{\alpha}_{21}$	0.005	0.005	0.070	0.070	0.79	0.69	0.002	0.002	0.045	0.045	0.29	0.22
$\widehat{\rho}$	0.005	0.004	0.071	0.067	-0.51	-0.82	0.002	0.002	0.045	0.042	-0.97	-0.36
$\hat{\lambda}_{11}$	0.009	0.010	0.099	0.098	-1.34	-1.29	0.004	0.004	0.063	0.063	-0.36	-0.37
$\widehat{\zeta}$	0.008	0.009	0.099	0.098	-1.28	-1.21	0.004	0.004	0.063	0.063	-0.35	-0.38
$\hat{\lambda}_{22}$	0.003	0.003	0.053	0.053	-0.50	-0.53	0.001	0.001	0.033	0.033	-0.02	0.04

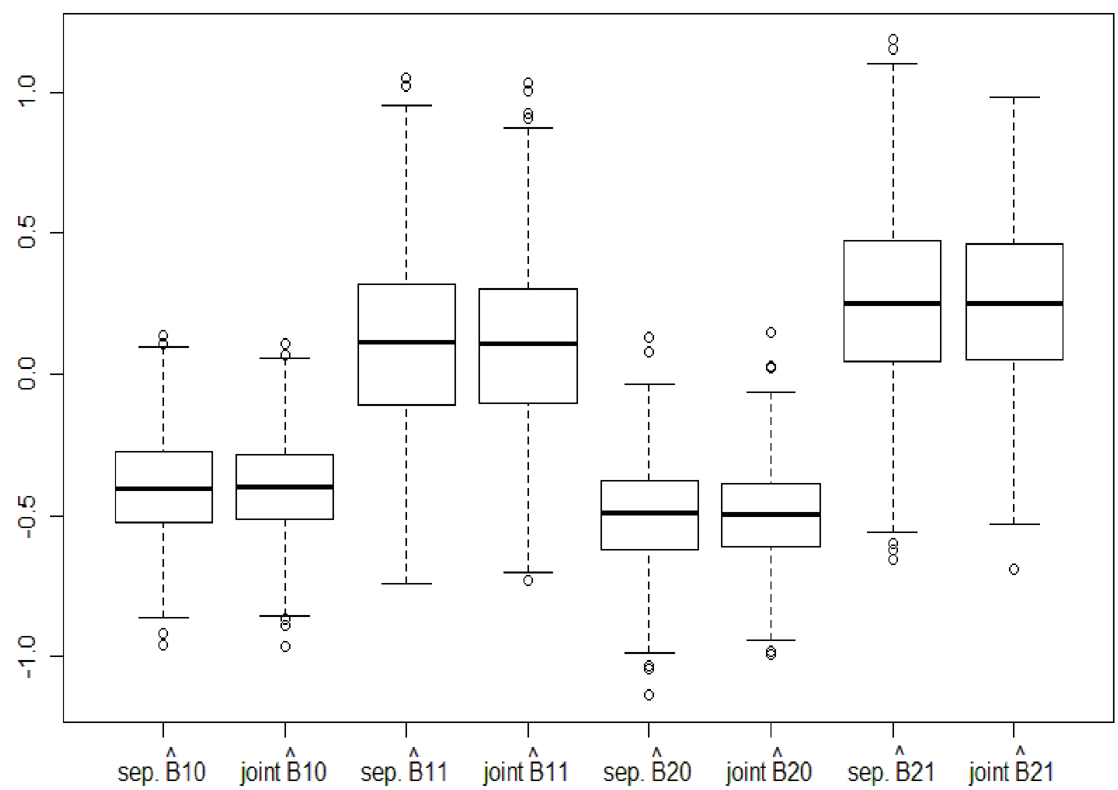

Figure 6.1: Box plot of the estimates of data model parameters under separate and joint models with fully-observed data and sample size 500.

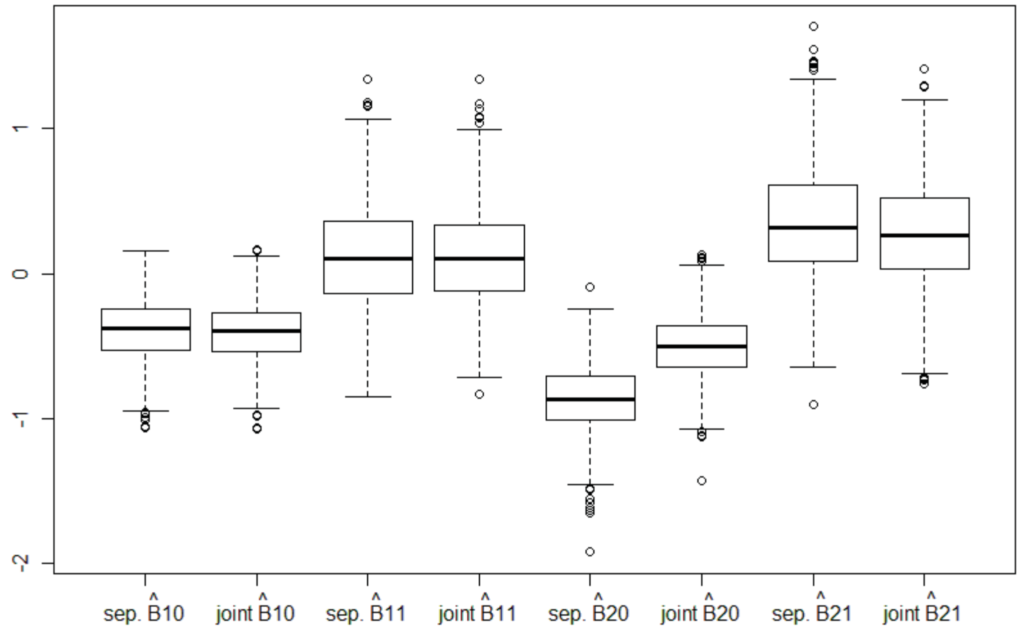

Figure 6.2: Box plot of the estimates of data model parameters under separate and joint models with MAR data and sample size 500.

6.6 Discussion

In this chapter, we studied the properties of the conditional mixed discrete and continuous outcomes model. We compared the performance of estimation based on a joint model for the mixed outcomes with estimation based on a model based on modeling the binary and continuous outcomes separately under two scenarios: one with fully-observed data, and the other with binary data missing at random. Results from the simulation study suggest that when data was fully observed, the point estimates of the parameters were very similar for the separate and joint models; However, with MAR data, the joint model performed better in the estimation of binary parameters.

A disadvantage of the joint modeling method is that it requires significant computational efforts, both in terms of time needed to submit the procedures in R/SPLUS, as well as high chance for making errors when deriving and recording compound expressions for derivatives.

Chapter 7

Conclusion

In this thesis, we have studied various aspects of statistical inference in the presence of missing data. In Chapters 2, 3 and 4, we obtained asymptotically correct normal approximation and empirical likelihood confidence intervals, and proposed adjustments to the Shao and Sitter's (1996) bootstrap confidence intervals under imputation for missing data. We began by focusing on inference about the mean $\mu = E(Y)$, and showed the corresponding results for the distribution function $F(y)$, for given y. In Chapter 2, we investigated asymptotic properties of the imputed estimators under the basic i.i.d. scenario and considered the fractional imputation method (Kim and Fuller, 2004). This theory was extended to a more practical case with imputation classes in Chapter 3. In Chapter 4, we considered a bivariate parameter with possible nonresponse in both variables, and constructed confidence intervals on the correlation coefficient ρ under joint regression imputation (Shao and Wang, 2002). The structure for developing the theory was similar for the three chapters: first, we studied asymptotic properties of imputed estimators and proposed asymptotically valid bootstrap percentile confidence intervals on the parameters of interest under imputation; then we established limiting distributions of the empirical likelihood ratio statistics and proposed asymptotically valid bootstrap calibrated empirical likelihood confidence intervals. We showed that the adjusted bootstrap estimators should be used when we use the bootstrap data obtained by imitating the process of imputing the original data set. Results of simulation studies demonstrated that, in general, the proposed method led to better coverage and improved efficiency of bootstrap percentile and bootstrap calibrated empirical likelihood confidence intervals. For future research, it would be of interest to extend the results on the correlation coefficient, presented in Chapter 4, to multiple imputation classes, based on the method shown in Chapter 3. Further, all results could be extended to complex surveys by adapting the pseudo-EL approach (Wu and Rao, 2006). Finally, extensions to different imputation methods could be examined since there are some disadvantages to the imputation methods that we chose for our research. For example, under fractional imputation, the confidence intervals require identification flags on the imputed values present in data file, which in practice may be difficult to obtain due to confidentiality reasons (Qin et al., 2008); Under joint regression imputation, additional variability due to the random selection of residuals may lead to inefficient estimators. Chauvet and Haziza (2011) proposed a balanced joint random regression imputation that preserves the coefficient of correlation between two variables and eliminates imputation variance arising from random selection of residuals. It would be of interest to extend our work on correlation coefficient by studying the balanced joint random regression imputation.

In Chapter 5, we established a goodness-of-fit test that can be applied to the case of longitudinal data with MAR observations by combining the concepts of the weighted generalized estimating equations (WGEE) (Robins et al., 1995) and the score test (Hosmer and Lemeshow, 1980, Horton et al., 1999). We compared the proposed goodness-of-fit method, which incorporates the estimation of the missingness model parameters, versus the ordinary method that ignores the missingness process (Horton et al. 1999). Our simulation study showed that the proposed goodness-of-fit method should be used when dealing with intermittent missingness. The results were obtained under an assumption that subjects had equal chance of participating in the longitudinal study, or that longitudinal survey data was generated by simple random sampling with negligible sampling fraction; For future research, we would like to adapt the proposed goodness-of-fit method to complex survey data, which may involve dependencies both within and among subjects, using design weights and based on the work of Roberts et al. (2009), Rao et al. (1998), Binder (1993) and Rao and Wu (1988).

In Chapter 6, we studied the conditional model for a mixture of correlated discrete and continuous outcomes and applied the likelihood method to MAR data. We studied the performance of the joint model compared to modeling the binary and continuous outcomes separately. When all data were observed, adopting the mixed model did not lead to notable improvements. On the contrary, under a scenario with binary data missing at random, the joint model performed significantly better. For future research, an application of the proposed method to real data would be of interest. We would also like to consider extending the method to the case of longitudinal data. The major challenge to this method is its computational complexity, which increases with the number of outcomes of interest. In case of longitudinal data, adapting the weighted generalized estimating equations approach (Robins et al., 1995) would allow to relax distributional assumptions and simplify the computations but could lead to a loss in efficiency.

Bibliography

[1] Allison, P. D. (2012). Handling missing data by maximum likelihood. Statistical Horizons, Haverford, PA, USA.

[2] Archer K. J., Lemeshow S. and Hosmer D. W. (2007). Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design. Computational Statistics and Data Analysis. 51(9): 4450-4464.

[3] Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items, In: Studies in Item Analysis and Prediction. Standford Mathematical Studies in the Social Sciences. Solomon, H. (Ed.); Vol 6, 158–68.

[4] Billingsley, P. (1995). Probability and measure. 4th ed., John Willey & Sons, New York.

[5] Binder, D. A. (1993). On the variances of asymptotically normal estimators from complex surveys. International Statistical Review 51, 279-292.

[6] Binder, D. A., Roberts G.R. (2003). Design-based and model-based methods for estimating model parameters, in Analysis of Survey Data. Chambers, R. L. and Skinner, C. J. editors, Wiley, Chichester.

[7] Brick, J. M. and Kalton, G. (1996). Handling missing data in survey research. Statistical Methods in Medical Research, 5, 215 - 238.

[8] Chauvet, G. & Haziza, D. (2012). Fully efficient estimation of coefficients of correlation in the presence of imputed survey data. The Canadian Journal of Statistics, 40, 124-149.

[9] Chen, J. and Rao, J.N.K. (2007). Asymptotic normality under two-phase sampling designs. Statistica Sinica, 17, 1047-1064.

[10] Chen, J., Rao, J. N. K. and Sitter, R. R. (2000). Efficient random imputation for missing data in complex surveys. Statistica Sinica, 10, 1153-1169.

[11] Chen, J., Sitter, R.R. and Wu, C. (2002). Using Empirical Likelihood Method to Obtain Range Restricted Weights in Regression Estimators for Surveys. Biometrika, 89, 230-237.

[12] Chen, Y. and Shao, J. (1999). Inference with survey data imputed by hot deck when imputed values are nonidentifiable. Statist. Sinica. 9, 361-384.

[13] Deville, J. C. and Sarndal, C. E. (1994). Variance estimation for the regression imputed Horvitz-Thompson estimator. Journal of Official Statistics. 10, 381-394.

[14] Dickey, B., Normand, S. L. T., Hermann, R. C., Eisen, S. V., Cortes, D. E., Cleary, P. D. and Ware, N. (2003). Guideline recommendations for treatment of schizophrenia: the impact of managed care, Arch. Gen. Psychiatry, 60, 340–8.

[15] Diggle, P., Heagerty, K. Y., Liang K-Y and Zeger S. L. (2002). Analysis of Longitudinal Data, 2nd edn. Oxford. Oxford University Press.

[16] Diggle P. and Kenward M. F. G. (1994). Informative drop-out in longitudinal data analysis. Appl. Statist, 43, No. 1: 49-93.

[17] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics, 7, 1–26.

[18] Emrich, L.J. and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. Amer. Stat., 45, 302–304.

[19] Fang F., Hong Q. and Shao J. (2009). A pseudo empirical likelihood approach for stratified samples with nonresponse. Annals of Statistics, Vol 37, No. 1, 371-393.

[20] Farrell, P. J. and Rogers-Stewart, K. (2008). Methods for generating longitudinally correlated binary data. International Statistical Review, 76, 1, 28–38.

[21] Fitzmaurice G. M., Davidian M., Verbeke G. and Molenberghs, G. (eds) (2008) Longitudinal Data Analysis: A Handbook of Modern Statistical Methods.

[22] Fitzmaurice G. M. and Laird N. M. (1997). Regression models for mixed discrete and continuous responses with potentially missing values. Biometrics. 53:110–122.

[23] Fitzmaurice, G. M., Molenberghs, G. and Lipsitz, S. R. (1995). Regression models for longitudinal binary responses with informative dropouts. J. Roy. Statist. Soc. Ser. B 57 691–704.

[24] Francisco, C. A. and Fuller, W. A. (1991). Quantile estimation with a complex survey design. Ann. Statist. 19, 454-469.

[25] Graubard, B. I., Korn, E. L., Midthune, D. (1997). Testing goodness-of-fit for logistic regression with survey data. In: Proceedings of the Section on Survey Research Methods. American Statistical Association.

[26] Gupta A. D. (2008). Asymptotic theory of statistics and probability. Springer.

[27] Hartley, H. O. and Rao, J. N. K. (1968). A new estimation theory for sample surveys. Biometrika. 55, 547-557.

[28] Hartley, H. O., and Rao, J. N. K. (1969). A new estimation theory for sample surveys II. In New Developments in Survey Sampling. New York: Wiley Inter-Science.

[29] Haziza, D. and Beaumont, J-F. (2007). On the construction of imputation classes in surveys. International Statistical Review, 75, 25-43.

[30] Horton, N.J., Bebchuk J. D., Jones, C. L., Lipsitz, S. R., Catalano, P. J., Zahner, G. E. P. and Fitzmaurice, G. M. (1999). Goodness-of-fit for GEE: an example with mental health service utilization. Statist. Medicine 18, 213–222.

[31] Hosmer, D. W., Hosmer, T. L. E., Cessie S. and Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. Statist. Medicine 16, 965–980.

[32] Hosmer, D. W. and Lemeshow, S. (1980). Goodness-of-fit tests for the multiple logistic regression model. Comm. Statist. Theory Meth. A A9 (10) 1043–1069.

[33] Kalton, G. and Kasprzyk, D. (1986). The treatment of missing servey data. Survey Methodology. 12. 1-16.

[34] Kalton, G. and Kish, L. (1984). Some efficient random imputation methods. Communications in Statistics: Series A, 13, 1919-1939.

[35] Kanter, M. (1975). Auto-regression for discrete processes model. J. Appl. Prob., 12, 371–375.

[36] Kim, J. K. and Fuller, W. A. (2004). Fractional hot deck imputation. Biometrika, 91, 559-578.

[37] Kott, P. (2012). Why one should incorporate the design weights when adjusting for unit nonresponse using response homogeneity groups. Survey Methodology, Vol. 38, No. 1, pp. 95-99.

[38] Kuss, O. (2002). Global goodness-of-fit test in logistic regression with sparse data. Statist. Med. 21: 3789-3801.

[39] Laird, N. M., (1988). Missing data in longitudinal studies. Statistics in Medicine 7, 305-315.

[40] Lessler, J. T. and Kalsbeek, W. D. (1992). Nonsampling error in surveys. Wiley, New York.

[41] Liang K. and Zeger S. (1986). Longitudinal data analysis using generalized linear models. Biometrika. 73(1):13–22.

[42] Little, R. J. A. and Rubin, D. B. (2002). Statistical analysis with missing data. 2nd edition. John Wiley & Sons, New York.

[43] McCullagh P. and Nelder JA (1983). Quasi likelihood functions. Annals of Statistics, 11, 59-67.

[44] Moore, D. S. and Spruill M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. Ann. Statist. Volume 3, Number 3, 599-616.

[45] Newey, W. K. and McFadden, D. (1993). Estimation in large samples. Handbook of Economics, Vol. 4.

[46] Oman, S. D. and Zucker, D. M. (2001). Modelling and generating correlated binary variables. Biometrika, 88, 287–290.

[47] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 75, 237–249.

[48] Owen, A. B. (1990). Empirical likelihood ratio confidence regions. Ann. Statist. 18, 90-120.

[49] Owen, A. B. (2001). Empirical likelihood. Chapman and Hall, New York.

[50] Parzen M. and Lipsitz S. R., (1999). A global goodness-of-fit statistic for Cox regression models. Biometrics 55, 580-584.

[51] Preisser, J. S., Galecki, A.T., Lohman, K.K. and Wagenknecht, L.E. (2000). Analysis of smoking trends with incomplete repeated responses. Journal of the American Statistical Association, 95, 1021-31.

[52] Preisser, J. S., Lohman K. K. and Rathouz P. J. (2002). Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. Statist. Med. 2002; 21:3035–3054.

[53] Preisser, J. S. and Qaqishy, B. F., (2012). A comparison of methods for generating correlated binary variates with specified marginal means and correlations (working paper). The Berkeley Electronic Press.

[54] Prentice, R. L., (1988). Correlated binary regression with covariates specific to each binary observation, Biometrics, 44, 1033–48.

[55] Pulkstenis E. and Robinson T. J. (2002). Two goodness-of-fit tests for logistic regression modles with continuous covariates. Statist. Med. 21: 79-93.

[56] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. Ann. Statist., 22, 300–325.

[57] Qin, J. and Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. Canadian Journal of Statistics, 23-2, 145–159.

[58] Qin, Y, Rao, J. N. K. and Ren, Q. (2008). Confidence intervals for marginal parameters under imputation for item nonresponse. Journal of Statistical Planning and Inference, 138, 2283-2302.

[59] Qin, Y, Rao, J. N. K. and Ren, Q. (2008). Confidence intervals for marginal parameters under fractional linear regression imputation for missing data. Journal of Multivariate Analysis, 99, issue 6, 1232-1259.

[60] Qin, Y., Rao, J. N. K. and Wu, C. (2010). Empirical likelihood confidence intervals for the Gini measure of income inequality. Economic Modelling. 27. 1429-1435.

[61] Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex surveys: chisquared tests for goodness for fit and independence in two-way tables. Journal of the American Statistical Association, 76, 221-230.

[62] Rao, J. N. K. (1996). On variance estimation with imputed survey data. J. Amer. Statist. Assoc. 91, 499-520.

[63] Rao, J. N. K. (1998). Marginal models for repeated observation: Inference with survey data. American Statistical Association Proceedings of the Survey Research Methods Section, 76- 82.

[64] Rao, J. N. K., Scott, A. J. and Skinner, C. J. (1998). Quasi-score tests with survey data. Statistica Sinica, 8, 1059-1070.

[65] Rao, J. N. K. and Wu, C. (1988). Re-sampling inference with complex survey data. JASA, March Vol. 83 No. 401 pp. 231-241.

[66] Regan, M. M. and Catalano, P. J. (1999). Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology. Biometrics, 55, 760–768.

[67] Roberts, G., Ren, Q. and Rao, J. N. K. (2009). Using marginal mean models for data from longitudinal surveys with a complex design: some advances in methods. In Methodology of Longitudinal Survey, P. Lynn (Ed.), John Wiley & Sons, New York, pp. 351–366.

[68] Robins, J., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes under the presence of missing data. Journal of the American Statistical Association, 90,106-121.

[69] Rubin, D. B. (1976). Inference and missing data. Biometrika 63, 581-592.

[70] Rubin, D. B. (1978). Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. Proceedings of the Survey Research Methods Section of the American Statistical Association, 20- 34.

[71] Sammel, M. D., Lin, X and Ryan, L. M. (1999). Multivariate linear mixed models for multiple outcomes. Statistics in Medicine, 18: 2479-2492.

[72] Sammel, M. D., Ryan, L. M. and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. Journal of the Royal Statistical Society, Series B: Methodological, 59, 667–678.

[73] Satorra, A., and Bentler, P. M. (1986). Some robustness properties of goodness of t statistics in covariance structure analysis. ASA Proceedings.

[74] Satorra, A., and Bentler, P. M. (1988). Scaling corrections for chisquare statistics in covariance structure analysis. ASA Proceedings.

[75] Serfling, R. J. (1980). Approximation theorems of mathematical statistics. John Wiley & Sons, New York.

[76] Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap. New York: Springer.

[77] Shao, J. and Sitter, R. (1996). Bootstrap for imputed survey data. Journal of American Statistical Association, 91, 1278-1288.

[78] Shao, J. and Wang, H. (2002). Sample correlation coefficients based on survey data under regression imputation. J. Amer. Statist. Assoc. 97, 545-552.

[79] Sinha S. K., Laird N. M. and Fitzmaurice G. M. (2010). Multivariate logistic regression with incomplete covariate and auxiliary information. Journal of Multivariate Analysis, 101: 2389-2397

[80] Skinner, C. J. and Rao, J. N. K. (2002). Jackknife variance estimation for multivariate statistics under hot-deck imputation from common donors. Journal of Statistical Planning and Inference. 102. 149-167.

[81] Teixeira-Pinto, A. and Normand, S. (2009). Correlated bivariate continuous and binary outcomes: issues and applications. Stat Med. June 15; 28(13): 175.

[82] Teixeira-Pinto, A. and Normand, S. (2011). Missing data in regression models for non-commensurate multiple outcomes. Revstat, 9.1: 37–55.

[83] Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. J. Amer. Statist. Assoc. 70, 865-871.

[84] Wang, D. and Chen, S. X. (2009). Empirical likelihood for estimating equation with missing values. Ann. Statist., 37, 490–517.

[85] Wu, C. (2005). Algorithms and R codes for the pseudo empirical likelihood method in survey sampling. Survey Methodology. Vol. 31, No. 2. pp. 239-243.

[86] Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuaous outcomes. Biometrics. Vol. 42, No 1: 121-130.