STATISTICAL INFERENCES FOR OUTCOME DEPENDENT SAMPLING DESIGN WITH MULTIVARIATE OUTCOMES

Tsui-Shan Lu

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics, Gillings School of Global Public Health.

Chapel Hill
2009

Approved by:

Advisor: Haibo Zhou, Ph.D.
Reader: Gary G. Koch, Ph.D.
Reader: Matthew P. Longnecker, Sc.D.
Reader: Mark A. Weaver, Ph.D.
Reader: Fei Zou, Ph.D.

Abstract

TSUI-SHAN LU: Statistical Inferences for Outcome Dependent Sampling Design with Multivariate Outcomes.
(Under the direction of Haibo Zhou.)

An outcome-dependent sampling (ODS) design has been shown to be a cost-effective sampling scheme. In the ODS design with a continuous outcome variable, one observes the exposure with a probability, maybe unknown, depending on the outcome. In practice, multivariate data arise in many contexts, such as longitudinal data or cluster units. While the ODS design has been an interest in statistical and applied literature, the statistical inference procedures for such design with multivariate cases still remain undeveloped. We develop a general sampling design and inference methods using the ODS under continuous multivariate settings (*Multivariate-ODS*). The standard estimation methods for multivariate data ignoring the *Multivariate-ODS* design will yield biased and inconsistent estimates. Therefore, new statistical methods are needed to reap the benefits of a *Multivariate-ODS* design.

In this dissertation, we propose three commonly occurring ODS sampling strategies and study the new semiparametric methods for estimating regression parameters. We allow a simple random sample (SRS) in all three sampling strategies and the difference is how the supplemental samples are selected. The first design, the *Multivariate-ODS* with a maximum selection criterion, selects the supplemental sample based on whether the maximum value of the outcomes from an individual exceeds a known cutpoint; the second design, the *Multivariate-ODS* with a summation criterion, draws the supplemental

sample based on whether the sums of the outcome values are above a given cutpoint; the third design, the *Multivariate-ODS* with a general criterion, is a more general design where the selection of the supplemental samples is based on each individual's responses, instead of on the aggregate of the outcomes.

Acknowledgments

I would like to thank my advisor, Dr. Haibo Zhou, who not only directs me in the dissertation with extensive and in-depth cultivation, but also encourages me to explore all kinds of research possibilities and appreciate the joy the research brings. I would also like to thank my dissertation committee members. Thank you to Dr. Gary Koch for taking his time and encouraging me to keep moving forward constantly. Thank you to Dr. Matthew Longnecker at the National Institute of Environmental Health Sciences (NIEHS) for helping me with the CPP data set and serving as a contributing member of my committee. Thank you to Dr. Mark Weaver whose dissertation and paper were my best guidebooks to my dissertation. Thank you to Dr. Fei Zou for her many helpful suggestions from her expertise in genetics.

Thank you to Dr. Paul Stewart for providing financial support during the first three and half years of my degree and giving me the opportunity to learn how to be a biostatistician.

Finally, much deserved thanks goes to my family, who support me all the time and are always there for me. I would also like to express my sincere appreciation to all of my unwavering friends for their support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Introduction

The observational epidemiology study designs are often used when investigating the relationships between a disease outcome and an exposure given other characteristics. The commonly used designs include cohort and case-control studies; the former is a study to observe several individual exposures and the individual disease occurrence on the basis of a follow-up period and can end up taking a very long time, whereas the case-control design is retrospective and studying the patients already having a disease to yield more information on risk factors of this group of people that differ from those who are free of disease (Cornfield, 1951). The case-control study in epidemiology or the choice-based sampling in econometrics are examples of an *outcome-dependent sampling* (ODS) design, which is more appealing and increasing the efficiency for studying rare diseases because the researchers can concentrate resources on observations with the greatest amount of information of primary interest (Anderson, 1972). If the observations on exposures and other covariates are easier or cheaper to measure, then the ideal situation is to collect all of the data on every member in a finite population studied. However, this is not always the case due to high cost, limited resources and inefficiency. As a result, the case-control

study is preferred since it can avoid making statistical inferences on the entire population and still achieve the efficiency provided by the selected subsets of observations sampled based on the outcome. The logistic regression method is widely utilized to estimate the adjusted relative risks between a dichotomized response and exposures, are applied to analyze the subsamples of diseased cases and diseased-free controls obtained from an underlying population.

Based on the framework of the case-control study design, one can further enhance efficiency and reduce cost by double sampling for stratification, balancing the numbers of exposed and non-exposed individuals within cases and controls for whom covariate information is ascertained. White (1982) proposed a two-stage stratified design, where data on the response variable and the exposure variable are obtained for a large sample in the first stage and only information on other covariates from a subsample is available in the second stage, with the purpose of studying the association between a rare exposure and a rare disease, sampling a larger proportion of the subjects from the small groups and a smaller proportion from the large groups to achieve the efficiency of the estimates of the parameters of interest. Variations of White's two-stage sampling have been discussed and proposed. For example, Breslow and Cain (1988) considered the preliminary sample to be separate samples of cases and controls drawn from subpopulation of diseased and non-diseased subjects, and developed modified logistic regression for data in a two-stage case-control design. Prentice (1986) considered a case-control within a cohort design for the failure-time data. Other published research making inferences on two-stage case-control studies includes Zhao and Lipsitz (1992), Schill et al. (1993), Wacholder and Weinberg (1994), Lawless, Kableisch and Wild (1999), and Wang and

Zhou (2005 and 2009). Breslow and Holubkov (1997) proposed the method to obtain the full maximum likelihood estimator of logistic regression parameters under the two-stage outcome-dependent sampling with the binary outcome variable. Much work for studying dichotomous outcomes under an ODS setting has been continuously developed.

For studies of investigating the association between an exposure measure and a continuous outcome, a common approach is to dichotomize the outcome or categorize it with several cutpoints and conduct statistical analyses on the categorical outcomes. However, this will result in selection bias since dichotomization of the outcome will induce a loss of efficiency and information and increase the risk for misclassification (Suissa, 1991; Zhou et al., 2002).

For directly using continuous outcome variable without losing information on dichotomization, Zhou et al. (2002) considered a general ODS scheme where an overall simple random sample from the base population (the prospective component) and additional supplement samples drawn from segments of the outcome space of particular interest (the retrospective component) were observed. In other words, the supplemental random samples are chosen depending on the outcome, a case-control-like sampling, from the observations believed to be the most informative. They proposed a semiparametric empirical likelihood inference procedure in which the underlying distribution of covariates is treated as a nuisance parameter and is left unspecified. Weaver and Zhou (2005) developed an estimated likelihood method for continuous outcome under a similar outcome-dependent sampling scheme with the exception that the sampling is independent of a continuous auxiliary covariate. For missing exposure and other important covariates of each member, they proposed a maximum estimated likelihood estimator (MELE)

which is related to the "plug-in" method (Pepe and Fleming, 1991, and Zhou and Pepe, 1995). Under the setting of the ODS samping described by Zhou et al. (2002), Wang and Zhou (2006) considered the model for both the binary outcome and the response variable with more than two categories while the information on the parent cohort is little and the sampling probability is not identifiable, which for example, arises when the percentage of response from each member in the first stage is relatively low. They proposed a semiparametric empirical likelihood-based method with auxiliary covariates that relate to the exposure of interest. The advantage to such ODS design is that the statistical power is improved over the simple random sample design because investigators can oversample sub-populations believed to be influential, and in the meantime the study itself can enhance efficiency by allowing the selection probability of each individual in the ODS sample to depend on the outcome.

The methods discussed above were all developed for a univariate continuous outcome variable; that is, only one outcome measurement per subject has been so far considered. In practice, data collected are often in a multivariate form for the response variable: longitudinal in nature where multiple observations for an individual are collected or where studies are conducted on the basis of participating cluster units. We can see that multivariate data arise in many contexts in some examples: in epidemiological cohort studies where the outcomes are recorded for members within families; in animal experiments in which treatments are applied to samples of littermates; in most clinical trials where study subjects are experiencing multiple events. Among these studies, a common feature is that the responses might be correlated. As the field of epidemiology expands and evolves, an increasing number of studies are conducted using the *Multivariate-ODS* design, a further

generalization of the biased sampling, which is built on the idea of the ODS design with aggregate of the responses and allows investigators to concentrate resources on the segments with the greatest amount of information. The related and motivated examples of studies will be given in the following section. The robust and efficient statistical method accounting for the *Multivariate-ODS* setting, however, is still underdeveloped. Therefore, new and efficient development of statistical inference procedure is needed in order to take advantage of data sets under the *Multivariate-ODS* design.

In this dissertation, we propose to develop statistical inferences on regression models under a *Multivariate-ODS* design. We will show that if the outcome-dependent nature is correctly accounted for, then we can develop more efficient and powerful estimators. Then we can investigate the sampling strategies under the *Multivariate-ODS* framework that will indeed lead to more cost-effective studies. The underlying distributions of covariates will be modeled nonparametrically using the empirical likelihood methods. A novelty of the proposed methods is that one will be able to make inferences on the regression parameters without postulating any of the distributions for the covariates by combining a nonparametric component with a parametric regression model. We will use simulated data to evaluate our proposed estimators and compare their efficiency with those of other naive and existed methods. The proposed method will be also applied to analyze the data sets presented in the next section.

1.2 Studies With Multivariate-ODS Design Schemes

We are motivated by the following studies, which illustrate how each design involves a *Multivariate-ODS* scheme with continuous outcomes.

1.2.1 The Collaborative Perinatal Project

The Collaborative Perinatal Project (CPP) is a prospective cohort study designed to identify determinants of neurodevelopmental deficits in children (Niswander and Gordon, 1972; Gray et al., 2000). Nearly 56,000 pregnant women were recruited into the CPP study from 1959 through 1966 at any one of 12 study centers across the United States (Baltimore, Maryland; Boston, Massachusetts; Buffalo, New York; Memphis, Tennessee; Minneapolis, Minnesota; New Orleans, Louisiana; New York, New York (2 hospitals); Philadelphia, Pennsylvania; Portland, Oregon; Providence, Rhode Island; and Richmond, Virginia). Women were enrolled, usually at their first prenatal visit; it resulted in 55,908 pregnancies (9,161 women contributed multiple pregnancies to the study). Data were collected on the mothers at each prenatal visit and at delivery and when the children were 24 hours, 4 and 8 months, and 1, 3, 4, 7, and 8 years. Among all the measures, we are interested in audiometric evaluation, which was done when the children were approximately 8 years old. Longnecker et al. (2004) studied the association in humans between maternal third trimester serum polychlorinated biphenyls (PCBs) levels and audiometry results in offsprings at approximately 8 years old. They defined sensorineural hearing loss (SNHL) as a hearing threshold $\geq$ 13.3 dB based on the average across both ears at 1000, 2000, and 4000 Hz, in conjunction with no evidence of conductive

hearing loss, which was defined by the air-bone difference in hearing threshold being $\geq$ 10 dB, based on the average across both ears as well. The sample selected by the study investigators in the analysis is indeed in a *Multivariate-ODS* setting: 726 having an 8-year audiometric evaluation of 1200 subjects selected at random from the underlying population and a supplemental sample of 200 eligible children randomly selected from the 440 children whose 8-year audiometric evaluation showed SNHL. In other words, we can investigate such data by developing a regression model for multiple continuous outcomes under a *Multivariate-ODS* design with two components: an overall random sample of the population and one supplemental random sample taken from subjects who are defined as having hearing loss, in order to achieve greater efficiency than a completely simple random sample or simply dichotomizing the continuous outcome.

1.2.2 The Family Heart Study

The Family Heart Study (FHS) (Higgins et al., 1996 and Liao et al., 1997) is a population-based, multi-center study designed to identify and evaluate the genetic and nongenetic determinants of coronary heart disease (CHD), atherosclerosis, and cardiovascular risk factors. Individuals and families were recruited in two phases from three ongoing parent cohort studies: the Forsyth, North Carolina, and Minneapolis, Minnesota cohorts of the Atherosclerosis Risk in Communities (ARIC) study, the Framingham Heart Study, and the Utah Family Thee Study in Salt Lake City. In phase I (June 1993 to July 1995), 3168 probands, 6283 parents (3140 fathers and 3143 mothers), 2834 current spouses, 12140 siblings, and 10902 children in the probands' families were recruited

to the FHS. A simple random sample of approximately 500 probands from each study site and another 500 probands with a high family risk score for CHD were sampled to characterize personal histories of CHD and related conditions. A family risk score was calculated using reported (observed) number of CHD events in first-degree relatives and the numbers expected, defined as the sum of the probabilities for the individual family members. To be eligible in the phase II, families should have two or more CHD events and risk scores of 0.5 or higher. This resulted in 588 randomly selected families and 657 families with the highest risk scores.

Liao et al. (1997) used logistic regression models to estimate the adjusted prevalence of proband stroke, based on the data from phase I of the FHS. The estimate obtained in their report could only capture the plausible risk factors for proband stroke status, but indeed ignore the data set analyzed was not a random sample and moreover, only a small number of strokes was present. To better take advantage of this huge data and establish a relationship between familial stroke history and other determinants, we can develop a model in a *Multivariate-ODS* design, in which the risk scores of father and son from each family are considered as outcome variables at the same time, to make the most use of all available data. We will revisit the Family Heart Study in the later Chapter.

1.3 Notation and Data Structure

1.3.1 Study Population and Model

Suppose we have a base population, each subject having multiple responses and corresponding covariates, such as exposure of interest and other characteristics observed or

measured, for which we denote $\boldsymbol{Y}$ as a vector of responses and $\boldsymbol{X}$ as the corresponding covariate vector. Suppose these realizations are from the joint density of $(\boldsymbol{Y}, \boldsymbol{X})$ that can be written as $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X})$, $(\boldsymbol{Y}, \boldsymbol{X}) \in \mathbb{Y} \times \mathbb{X}$, where $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$ is the conditional density function for $\boldsymbol{Y}$ given $\boldsymbol{X}$, $\boldsymbol{\theta}$ is a vector of the regression coefficients of interest, and $g_{\boldsymbol{X}}(\boldsymbol{X})$ is the marginal density of $\boldsymbol{X}$, which is independent of $\boldsymbol{\theta}$, and $\mathbb{Y}$ and $\mathbb{X}$ are the spaces of $\boldsymbol{Y}$ and $\boldsymbol{X}$, respectively. In the next section, we will first review the ODS design with only one response variable for each subject. Then the *Multivariate-ODS* schemes with difference selection criteria how the supplemental samples are obtained will be described thereafter.

1.3.2 The ODS Design for Univariate Outcome Variable

Let Y be a one-dimension continuous outcome variable. Assume that the domain of Y is partitioned into K mutually exclusive intervals by the fixed constants, $-\infty = a_0 < a_1 < \ldots < a_{K-1} < a_K = \infty$. The kth interval is denoted as $C_k = (a_{k-1}, a_k]$, $k = 1, \ldots, K$. Zhou et al. (2002) discussed a general ODS design allowing study investigators to obtain an overall *simple random sample* (SRS) of size n_0 and some *supplemental samples* of size n_k for the kth interval. The data structure for their design have two component:

(i) SRS Component: $\{Y_{0i}, \boldsymbol{X}_{0i}\}, \quad i = 1, \cdots, n_0;$

(ii) Supplemental Component: $\{Y_{ki},\ \boldsymbol{X}_{ki} \mid Y_{ki} \in C_k\}, \quad i = 1, \cdots, n_k;$

the total sample size for such ODS sample is $n = \sum_{i=0}^{K} n_i$.

Several ODS settings can be designed on the basis of the above general sampling scheme. For example, when $n_0 > 0$ and $n_k = 0$ for each k, the design reduces to an

SRS. The sampling augmented by a non-zero-observation SRS and several at-least-one-observation extra samples from strata, is another ODS format. Or, a random sample obtained can be strictly stratified, including supplemental samples with at least one subject for each but without an SRS. The data structure applied by Zhou et al (2002) is when $K = 3$ and $n_2 = 0$, $n_0 > 0$, $n_1 > 0$, $n_3 > 0$; that is, supplemental samples were observed from the tails of the distribution of Y.

1.3.3 The Multivariate-ODS Design Schemes

Let Y_{ij} be the j^{th} continuous outcome for the subject i, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$ ($p \geq 2$), and $\boldsymbol{X}_i$ be a vector of covariates, which can include both discrete and continuous components for the i^{th} subject; $\boldsymbol{Y}_i$ is a p-dimensional response vector ($p \geq 2$) for the ith subject. A *Multivariate-ODS* design includes two components: an overall *simple random sample* (SRS) from the base population and some *supplemental samples* randomly drawn from the domain of interest. Motivated by the CPP study and the Family Heart Study described in Section 1.2, we will discuss the following three selection criteria under a *Multivariate-ODS* scheme in this dissertation: (i) the maximum, (ii) the summation, and (iii) the general selection criteria. The likelihood functions and derivations for the corresponding estimators under each criterion will be presented in the following chapters.

The Multivariate-ODS with a Maximum Selection Criterion

The maximum selection criterion refers to the case where supplemental samples are chosen based on the maximum response out of each subject's outcome values. This

is particularly useful in the genetics studies. Suppose that the space of the maximum responses from the population, $\boldsymbol{Y}_{\max} = \{\max(Y_{i1}, \ldots, Y_{ip}), \forall i\}$, is the union of K mutually exclusive strata by the fixed constants $-\infty = a_0 < a_1 < \ldots < a_{K-1} < a_K = \infty$, and $C_k = (a_{k-1}, a_k]$, $k = 1, \ldots, K$, is the kth interval. The subject in the kth supplemental sample is observed if his/her maximum observation from the outcome values falls in the interval C_k. The data structure consisting of a *simple random sample* of size n_0 (≥ 0) and *supplemental samples* of size n_k (≥ 0) drawn from C_k is as follows:

(i) SRS Component: $\left\{\boldsymbol{Y}_i,\ \boldsymbol{X}_i\right\}, \quad i = 1, \ldots, n_0$;

(ii) Supplemental Component: $\left\{\boldsymbol{Y}_i,\ \boldsymbol{X}_i \mid \max\{Y_{i1}, \ldots, Y_{ip}\} \in C_k\right\}$, $i = 1, \ldots, n_k$ and $k = 1, \cdots, K$.

Let $n = \sum_{k=0}^{K} n_k$ be the total sample size of the *Multivariate-ODS* for which we observe complete data.

The Multivariate-ODS with a Summation Selection Criterion

This is the case when we observe supplemental random samples according to the sums of response measures. Assume that the domain of interest, the sums of responses, $\boldsymbol{Y}_\bullet = \left\{\sum_{j=1}^{p} Y_{ij},\ \forall i\right\}$, is partitioned into K mutually exclusive intervals by the known constants $-\infty = a_0 < a_1 < \ldots < a_{K-1} < a_K = \infty$, and the kth interval is denoted as $C_k = (a_{k-1}, a_k]$, $k = 1, \ldots, K$. The data structure of the *Multivariate-ODS* design under such selection criterion consists of two components: an overall *simple random sample* (SRS) of size n_0 (≥ 0) and a stratified *supplemental sample* of size n_k (≥ 0) randomly drawn from each interval C_k:

(i) SRS Component: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i\right\}$, $i = 1, \ldots, n_0$;

(ii) Supplemental Component: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid (\sum_{j=1}^{p} Y_{ij}) \in C_k\right\}$, $i = 1, \ldots, n_k$ and $k = 1, \ldots, K$.

The total sample size in the *Multivariate-ODS* is $n = \sum_{k=0}^{K} n_k$.

The Multivariate-ODS with a General Selection Criterion

The previous two selection criteria are discussed on the response domain of interest, the maximum and the summation, which is not the space of responses itself. In this case, the supplemental samples will be selected directly based on the response domain. Let $\boldsymbol{a} = \{a_j,\ j = 1, \ldots, p\}$ and $\boldsymbol{b} = \{b_j,\ j = 1, \ldots, p\}$, where a_j and b_j are known constants and $\{a_j > b_j, \forall j\}$, be the fixed cutpoints on the domain of $\boldsymbol{Y}_j = \{Y_{ij},\ \forall i\}$. The data structure of the *Multivariate-ODS* design under such selection criterion consists of three components: an overall *simple random sample* (SRS) of size n_0 (≥ 0), a *supplemental sample* of size n_1 (≥ 0) conditional on $\{Y_{i1} > a_1, Y_{i2} > a_2, \ldots, Y_{ip} > a_p\}$, and another *supplemental sample* of size n_2 conditional on $\{Y_{i1} < b_1, Y_{i2} < b_2, \ldots, Y_{ip} < b_p\}$:

(i) SRS Component: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i\right\}$, $i = 1, \cdots, n_0$;

(ii) Supplemental Component 1: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid \{Y_{i1} > a_1, Y_{i2} > a_2, \ldots, Y_{ip} > a_p\}\right\}$, $i = 1, \ldots, n_1$ and $j = 1, \ldots, p$;

(iii) Supplemental Component 2: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid \{Y_{i1} < b_1, Y_{i2} < b_2, \ldots, Y_{ip} < b_p\}\right\}$, $i = 1, \ldots, n_2$ and $j = 1, \ldots, p$;

the total sample size in the *Multivariate-ODS* is $n = \sum_{k=0}^{2} n_k$.

1.4 Literature Review

In this section, we will review related background and methods, some of which with some modifications could be applied to our *Multivariate-ODS* data structure, for making inferences about the parameters when data are obtained in an ODS scheme.

1.4.1 Methods for Data from a Case-Control Design

In epidemiological studies of correlating a disease with an exposure and other explanatory variables, the disease status is often dichotomous as having the disease or free of disease, and therefore epidemiologic cohort and case-control study designs are frequently used. A cohort study is a form of longitudinal and observational studies, based on data from a follow-up period of a group in which some have had, have or will have the exposure of interest, to determine the association between that exposure and the outcome. Studying infrequent events, such as death from cancer or a rare disease, using randomized clinical trials or other controlled prospective studies requires that relatively large populations be tracked for lengthy periods to observe disease development in order to yield reasonable results. These studies, however, can be prohibitively expensive because of the low likelihood that a certain disease will be developed.

An alternative is the case-control study design, which has several advantages, such as its efficiency, its applicability to rare as well as common diseases and its support of evaluating the cause-effect relationship (Breslow and Day, 1980). The basic and general tool allowing the scope of case-control study analysis is the linear logistic regression model. To be more specific, denote the case that the individual develops the disease as

$y = 1$ and the control for the disease-free individual as $y = 0$. The model that relates a single dichotomous outcome variable y to K regression variables $(x_1, \ldots, x_K)$ can be written as

$$\Pr(y = 1 \mid \boldsymbol{x}) = \frac{\exp(\alpha + \sum \beta_k x_k)}{1 + \exp(\alpha + \sum \beta_k x_k)} \tag{1.1}$$

or equivalently,

$$\text{logit } \Pr(y = 1 \mid \boldsymbol{x}) = \alpha + \sum \beta_k x_k$$

where α is the log odds of disease risk for a person with all the regression variables being zero and β_k is a parameter estimate for a multiplicative effect on the odds ratio.

However, the limitation of the case-control design is that it can only be applied to the dichotomous outcome variables under the logistic regression model. For a continuous outcome to fit the case-control design, dichotomizing the outcome may result in misclassification and tend to lose information. As a result, the method that takes the advantage of the case-control study and at the same time directly and fully utilize information in the continuous outcome has also been developed.

1.4.2 Other Extension of Case-Control Studies

White (1982) proposed a two-stage design especially for a rare disease and a rare exposure, whether it be cohort or case-control and used weighted least squares methods for estimating the relative risks. In the first stage, only data on the response and the exposure variables are collected from a large sample, which indeed is costly. During the second stage, random samples within the four groups: the case groups (the diseased and

exposed/unexposed) and the control groups (the non-diseased and exposed/unexposed) are chosen and information about other covariates is obtained. Under the scenario of a rare disease and exposure, one can expect very disparate sizes of the four groups and hence the advantage of the two-stage design is that additional observations from the smaller groups can also contribute to the estimation as well as those from the larger groups and together, such design can result in more efficient estimates of the parameters. Similar to the two-stage design, the case-cohort design also only consisting of the disease status and the exposure variable for all the subjects at the first stage but collecting covariate histories for all cases and only a random sample of the entire cohort at the second stage, is proposed to reduce redundant covariate information on disease free subjects (Prentice, 1986). Breslow and Cain (1988) proposed modified logistic regression for case-control data in the two-stage design and estimated parameters through 'conditional maximum likelihood' under the logistic model, which was developed for choice-based data by Manski and McFadden (1981) and Hsieh et al. (1985). In summary, Zhao and Lipsitz (1992) discussed a class of twelve possible designs within the framework of two-stage designs.

Breslow and Holubkov (1997) derived the full maximum likelihood (ML) estimator of logistic regression coefficients for data under a two-stage, ODS sampling design; data at the first stage are obtained as an ODS and at the second stage subjects are drawn using stratified random sampling from the first-stage subpopulations and explanatory variables are measured thereafter. Breslow and Holubkov's method demonstrated an advantage on efficiency of ML estimates for discrete data.

Hsieh et al. (1985) proposed an approach for estimation of response probabilities from choice-based data when retrospective data were augmented by auxiliary information

since case-control data alone cannot effectively identify response probabilities. Scott and Wild (1997) obtained maximum likelihood estimates of the parameters by fitting logistic regression models for stratified case-control and response-selective data. They showed the maximum likelihood estimates by simply iterating the pseudo-likelihood procedure by Wild (1991) with an "offset" parameter updated between iterations.

Wang and Zhou (2006) proposed a semiparametric empirical likelihood method for data in the two-stage ODS design, whose structure is

(i) SRS Component: $\{Y_i, \boldsymbol{X}_i, W_i\}, \quad i = 1, \ldots, m;$

(ii) Supplemental Component: $\bigcup_{j=1}^{J} \bigcup_{k=1}^{K} \left[\{\boldsymbol{X}_i | Y_i = j, W_i = k\}, \quad i = 1, \ldots, n_{jk} \right]$;

W is a categorical auxiliary variable for X where $\{W = k, k = 1, \ldots, K\}$. The key settings are that information on the parent cohort is unavailable and that the sampling probability is nonidentifiable. The empirical likelihood estimates for the marginal distribution of the covariates conditional on the auxiliary variable are estimated simultaneously. The proposed method can be applied to binary and multi-categorical outcome data.

1.4.3 Methods for Data from an ODS Design with Continuous Outcome

Recently, the continuous, univariate outcome in the ODS data has been considered along with the likelihood function derived.

Semiparametric Likelihood Method

Lawless et al. (1999) derived full semiparametric maximum likelihood estimates and developed several other semiparametric approaches for $\boldsymbol{\theta}$ for the response-selective data in the stratified ODS design, generated from the model $f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})g(\boldsymbol{x})$, where $\boldsymbol{y}$ is a response and $\boldsymbol{x}$ is a vector of covariates. They reviewed general semiparametric approaches for the stratified problems under the assumption that the strata totals for the sampling population are unknown. They presented theoretical asymptotic results for the estimators and handled the problems from the ODS, measurement error, and the missing data literature under a single framework.

Zhou et al. (2002) proposed a semiparametric empirical likelihood method for for data in an ODS design with a continuous outcome. Suppose $\boldsymbol{X}$ is a vector of covariates and the continuous outcome variable, Y, is partitioned into K mutually exclusive intervals by known constants satisfying that $-\infty = a_0 < a_1 < \ldots < a_{K-1} < a_K = \infty$ and let the kth interval be denoted as $C_k = (a_{k-1}, a_k], k = 1, \ldots, K$. Particularly, they considered the ODS sample consisting of an overall simple random sample of size n_0 and stratified supplemental random samples from the K intervals, each with size of n_k, $k = 1, 2, \ldots, K$. Let $L(\beta, G_{\boldsymbol{X}})$ denote the likelihood function for the ODS data

$$L(\beta, G_X) = \left[\prod_{i=1}^{n_0} f_\beta(y_{0i}|x_{0i}) g_X(x_{0i})\right] \times \left[\prod_{k=1}^{K}\prod_{j=1}^{n_k} f_\beta(y_{kj}, x_{kj}|y_{kj} \in C_k)\right], \qquad (1.2)$$

where β is the regression coefficient of interest. It can be further rewritten as

$$\begin{aligned} L(\beta, G_X) &= \left\{ \prod_{i=1}^{n_0} f_\beta(y_{0i}|x_{0i}) \prod_{k=1}^{K} \prod_{j=1}^{n_k} \frac{f_\beta(y_{kj}|x_{kj})}{F(a_k|x_{kj}) - F(a_{k-1}|x_{kj})} \right\} \\ &\quad \times \left\{ \prod_{i=1}^{n_0} g_X(x_{0i}) \prod_{k=1}^{K} \prod_{j=1}^{n_k} \frac{[F(a_k|x_{kj}) - F(a_{k-1}|x_{kj})] g_X(x_{kj})}{F(a_k) - F(a_{k-1})} \right\} \\ &= L_1(\beta) \times L_2(\beta, G_X), \end{aligned} \tag{1.3}$$

where $F(u) = \Pr(Y \leq u)$ and $F(u|x) = \Pr(Y \leq u|x)$. Zhou et al. obtained an estimate for β without specifying a form for G_X by profiling $L_2(\beta, G_X)$ by fixing β and then maximized the resulting profile likelihood function with respect to β. An empirical estimate of G_X, whose mass is located at each of the observed points $\boldsymbol{x}_i$, is obtained (Verdi, 1985, Owen, 1988, 1990, and Qin and Lawless, 1994). Denote $p_i = g_X(x_i)$ as discrete distributions with jumps at each point. $L_2(\beta, G_X)$ is proportional to

$$L_2(\beta, \{p_i\}) \propto \prod_{i=1}^{n} p_i \pi_1^{-n_1} \pi_3^{-n_3}, \tag{1.4}$$

where the case is taking when $K = 3$ and $n_2 = 0, n_1 > 0, n_3 > 0, n = n_0 + n_1 + n_3$; $\pi_1 = F(a_1)$ and $\pi_3 = 1 - F(a_2) = \bar{F}(a_2)$. Then L_2 is maximized over β and p_i subject to the following constraints:

$$\left\{ p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i \{F(a_1|w_i) - \pi_1\} = 0,\ \sum_{i=1}^{n} p_i \{\bar{F}(a_2|w_i) - \pi_3\} = 0 \right\}. \tag{1.5}$$

These constraints were implemented to uphold the properties of G_X as a discrete distribution function with support points at each observed point, x_i. Using the Lagrange

multiplier argument to maximize over p_i, one can write

$$H = \log L_2(\beta, \{p_i\}) + \rho(1 - \sum_{i=1}^{n} p_i) + n\lambda_1 \sum_{i=1}^{n} p_i\{F(a_1|w_i) - \pi_1\} + n\lambda_3 \sum_{i=1}^{n} p_i\{\bar{F}(a_2|w_i) - \pi_3\}\ , \tag{1.6}$$

where ρ, λ_1, and λ_3 are Lagrange multipliers. From the score equation of H with respect to p_i with the constraints in (1.5), one can show that $\rho = n$ and

$$\hat{p}_i = \frac{1}{n} \cdot \frac{1}{1 + \lambda_1\{F(a_1|w_i) - \pi_1\} + \lambda_3\{\bar{F}(a_2|w_i) - \pi_3\}},\ i = 1, \ldots, n. \tag{1.7}$$

Then an empirical profile log likelihood function can be obtained by plugging (1.7) into L_2 and the maximum semiparametric empirical likelihood estimator (MSELE) for the parameter vector can be derived. Zhou et al. showed efficient semiparametric estimation methods and likelihood ratio statistics that do not require specification of any distribution for the covariates.

Maximum Estimated Likelihood Estimator

Weaver and Zhou (2005) extended work above to the context of two-stage design, considering the population of whom Y is observed but $\boldsymbol{X}$ is unobserved, in additional to the ODS sample. Let n_V be the validation sample and $n_{\bar{V}}$ be the nonvalidation sample, referring to the complete observations and incomplete observations, respectively. Let N denote the total study population, including both complete and incomplete, N_k be the size of the kth stratum, $k = 1, \ldots, K$, and $n_{\bar{V},k} = N_k - n_{0,k} - n_k$ be the stratum size for

the nonvalidation sample. The full-information likelihood is

$$L_F(\boldsymbol{\beta}, G_X) = \left[\prod_{i\in V} f(Y_i|\boldsymbol{X}_i;\boldsymbol{\beta})\right]\left[\prod_{i\in V} g_X(\boldsymbol{X}_i)\right]\left[\prod_{j\in \bar{V}} f_Y(Y_j;\boldsymbol{\beta})\right]. \tag{1.8}$$

Unlike it in Zhou et al., a simple global empirical distribution function to estimate G_X is not valid since the data set on the covariate observations is not simple random sample. They proposed the estimator of G_X as

$$\widehat{G}_X(\boldsymbol{x}) = \sum_{k=1}^{K} \frac{N_k}{N}\widehat{G}_k(\boldsymbol{x}) \ , \tag{1.9}$$

where

$$\widehat{G}_k(\boldsymbol{x}) = \sum_{i\in V_k} \frac{I\{\boldsymbol{X}_i \le \boldsymbol{x}\}}{n_k + n_{0,k}} \tag{1.10}$$

is the empirical distribution function for the covariates in the stratum k. Then the last term in (1.8) is replaced with

$$\widehat{f}_Y(Y_j;\boldsymbol{\beta}) = \int f(Y_j|\boldsymbol{x};\boldsymbol{\beta}) d\widehat{G}_X(\boldsymbol{x}) = \sum_{k=1}^{K} \frac{N_k}{N(n_k + n_{0,k})} \sum_{i\in V_k} f(Y_j|\boldsymbol{X}_i;\boldsymbol{\beta}) \ , \tag{1.11}$$

After substituting the above equation into (1.8), the logarithm transformation of the likelihood function is

$$\widehat{l}_F(\boldsymbol{\beta}) = \left[\sum_{i\in V} \ln f(Y_i|\boldsymbol{X}_i;\boldsymbol{\beta})\right] + \left[\sum_{j\in \bar{V}} \ln\left\{\sum_{k=1}^{K} \frac{N_k}{N(n_k + n_{0,k})} \sum_{i\in V_k} f(Y_j|\boldsymbol{X}_i;\boldsymbol{\beta})\right\}\right]. \tag{1.12}$$

The maximum estimated likelihood estimators (MELEs), $\boldsymbol{\beta}$, can be obtained from the score equations of (1.12).

1.4.4 Methods for Modeling Multivariate Data under Non-ODS Setting

It is common in epidemiology that the response status from an individual is observed over time or repeatedly and therefore, data often comprise a binary or categorical time series. If there is only one single observation for each subject, the generalized linear models (GLMs) (McCullagh and Nelder, 1989), an extension of the linear modeling process, can be used to fit regression models on such univariate data, where response variables follow any probability distribution in the exponential family of distributions.

Longitudinal data for example consist an outcome variable, y_{it}, and a $p \times 1$ vector of covariates, x_{it}, at times $t = 1, \ldots, n_i$ for subjects $i = 1, \ldots, K$; statistical methods are already well-developed for modeling and analysis if data are approximately multivariate normally distributed. Laird and Ware (1982) proposed two-stage random-effects models for repeated measurements, where there is no requirement for balance in the data. The multiple measurements for each individual are assumed to follow the same probability distribution whereas the random-effects parameters of that distribution vary across subjects, which is so-called the second stage of the model. Ware (1985) presented and provided a detailed description of linear models for analyzing Gaussian longitudinal data.

For binary longitudinal data with time dependence within each individual's responses, logistic regression (Cox, 1958, 1970) for a single binary outcome for each subject is no longer valid because taking effect of dependence resulted from correlated data within

each subject into account is necessary. Zeger et al. (1985) analyzed binary longitudinal data with time-independent covariates with two proposed working models: one is that observations over time within each subject are assumed to be independent; the other one is that each series for each subject is a stationary Markov chain or order one, having a common first lag autocorrelation. They showed consistency property of both estimators under weak assumptions.

Extending GLMs to analyzing non-Gaussian longitudinal data, Liang and Zeger (1986) and Zeger and Liang (1986) further introduced a class of generalized estimating equations (GEEs) for regression parameters, accounting for the correlation among outcome observations for each subject, $\{\boldsymbol{y}_i\}$, in generalized linear models. The form for the joint distribution of the repeated measurements is not specified completely. Instead, the characteristic of using GEEs is that the marginal distribution of the dependent variable is considered rather than the conditional distribution given previous observations, and the marginal expectation (average response for observations sharing the same covariates) is modeled as a function of explanatory variables of interest. It makes it more difficult to obtain consistent estimators of the regression coefficients if the time dependence is not correctly specified; therefore, the GEEs for the estimates can guarantee consistency under minimal assumptions about the time dependence. Diggle, Liang, and Zeger (1994) provided a thorough review of marginal models and guideline to the choice of the correlation structures.

Zeger, Liang, and Albert (1988) introduced how a GEEs approach could be used in fitting both the subject-specific models, in which the heterogeneity is explicitly modelled, and the population-averaged models, where the regression coefficients are interpreted

for the population rather than for individuals. Liang, Zeger, and Qaqish (1992) also illustrated the use of the GEEs with multivariate categorical responses. Particularly, the method proposed allows to discuss marginal expectations of each response and pairwise associations.

1.4.5 Remarks

In this section, we will give a brief review and discuss the advantage and disadvantages of the methods described in this literature review and how the methods can relate to our proposed research.

In Sections 1.4.1 and 1.4.2, we presented several methods developed for discrete data, in a general ODS setting. Of these, some methods utilized data obtained in a two-stage ODS scheme (White, 1982; Prentice, 1986; Breslow and Cain, 1988; Zhao and Lipsitz, 1992; Breslow and Holubkov, 1997; Wang and Zhou, 2006) or choice-based study design (Manski and McFadden, 1981; Hsieh et al., 1985; Lawless, 1999). Breslow and Holubkov (1997) derived the full maximum likelihood estimator while Lawless (1999) developed full semiparametric maximum likehood estimates, which can be directly applicable for continuous outcome models in which ODS data are from stratified samples. Wang and Zhou (2006) further considered semiparametric empirical likelihood method for estimation, incorporating ODS data in two-stage along with additional information on an auxiliary variable.

In particular, the maximum semiparametric empirical likelihood estimators proposed by Zhou et al. (2002) and Weaver and Zhou (2005) were specifically developed for ODS

sampling scheme with continuous outcome variable, which were described in detail in Section 1.4.3. Therefore, our proposed methods will be an extension of those discussed by them for obtaining the estimates to multivariate continuous response variables.

Recently, a commonly applied approach is the GEEs (Liang and Zeger, 1986) if longitudinal data are non-Gaussian and comprised of repeated and correlated observations for an outcome variable. However, this method is available and applicable only when data are from simple random samples; in other words, for the *Multivariate-ODS* data we consider here, the assumptions for GEEs are invalid. Moreover, as we will discuss later, the estimator obtained, without knowing the marginal density of covariates, is not the most efficient.

It is clear from the discussion above that a method for estimating the parameters in a *Multivariate-ODS* regression model is needed for development. In this dissertation, we will propose and investigate such a method, as outlined in the next section.

1.5 Outline of the Remaining Dissertation

In Chapter 2, we will revisit the notation and the data structures under three *Multivariate-ODS* designs outlined in Section 1.3.3. We will develop the semiparametric likelihood approaches for each semiparametric empirical likelihood estimator for estimating the parameters in the model.

In Chapter 3, we will present asymptotic results for the proposed estimator using a maximum selection criterion. The consistency and asymptotic normality properties of the semiparametric maximum likelihood estimators will be shown and the asymptotic

variance structure will be derived.

In Chapter 4, we will study the small sample properties of the proposed estimator for the maximum selection criterion using simulated data with a bivariate normal regression model. The main goals are to see if (i) the asymptotic distribution derived in Chapter 3 is a reasonable approximation in the small samples and (ii) the proposed variance estimator is a good approximation to the actual variance in the small samples. Results obtained for the proposed estimator using the *Multivariate-ODS* design in this research will be compared to results obtained using other naive competing estimators. We will also study relative efficiencies by comparing our proposed estimator with the estimator from a simple random sample of the sample size as the *Multivariate-ODS* sample. In the end of this chapter, we will apply the proposed method to analyze the Collaborative Perinatal Project data described Section 1.2.

In Chapter 5, we will propose the estimator for the *Multivariate-ODS* design with a summation selection criterion to obtain the supplemental data. The proposed estimator will be shown to be consistent and asymptotically normally distributed. The asymptotic variance structure will be derived and a consistent variance estimator will be given. Then the small sample properties of the proposed estimator using a bivariate normal model will be studied. We will compare the proposed estimator to other competing estimators to determine what gains in efficiency, using simulated data generated from the conditional model specified to be a Normal density function. Then the proposed method will be applied to the CPP data. Asymptotic results for this estimator will be given in the end of this chapter. Chapter 5 is presented in a format of the manuscript.

In Chapter 6, we will study the estimator for the *Multivariate-ODS* design with a

general selection criterion to obtain the supplemental samples. We will show that the proposed estimator is consistent and asymptotically normally distributed, and derive the asymptotic variance structure. The small sample properties of the proposed estimator under the same model as in Chapters 4 and 5 will be discussed, where an extensive simulation study is carried out. The applications to the CPP data using the proposed method will be demonstrated. Again, this chapter is presented in a manuscript format and a sketched proof for the asymptotic results will be shown in the end of this chapter.

In Chapter 7, We will summarize this dissertation and suggest some possible extensions of the proposed methods in future research.

Advantages of the Proposed Estimators

Much research has been discussed for multivariate data, of which is a common and important form in epidemiological studies; nevertheless, the methods accounting for the *Multivariate-ODS* design are lacking. The proposed estimators by incorporating additional information into such *Multivariate-ODS* process can provide consistent and more efficient parameter estimates than those obtained by using a simple random sample of the same size. Our proposed estimators are semiparametric in nature that all the underlying distributions of covariates are modeled nonparametrically using the empirical likelihood methods. The *Multivariate-ODS* design, combined with an appropriate analysis, provides a cost-effective approach to conduct and analyze biomedical studies with multivariate responses for a given sample size.

CHAPTER 2

PROPOSED METHODS FOR THE MULTIVARIATE-ODS DESIGN

2.1 Introduction

In this chapter, we present three proposed semiparametric empirical likelihood methods for estimating the regression parameters for data obtained from an outcome-dependent sampling scheme with multivariate continuous outcomes according to the scenarios described in Section 1.3.3. In Section 2.2, we will develop the estimators from *Multivariate-ODS* data where the supplemental samples are obtained using the maximum selection criterion. In Section 2.3, we derive the estimators for the *Multivariate-ODS* with a summation selection criterion. The estimator for the *Multivariate-ODS* with a general selection criterion will be developed in Section 2.4.

2.2 The Multivariate-ODS with a Maximum Selection Criterion

2.2.1 Multivariate-ODS Likelihood for the Maximum Selection Criterion

Let Y_{ij} be the j^{th} continuous outcome for the subject i, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$ ($p \geq 2$), and $\boldsymbol{X}_i$ be a vector of covariates, which can include both discrete and continuous components for the i^{th} subject. The range of the random variable, $\boldsymbol{Y}_{\max} = \{\max(\boldsymbol{Y}_i),\ \forall i\}$ which consists of the maximum values of the responses, can be partitioned into K mutually exclusive intervals by the fixed constants $-\infty = a_0 < a_1 < \ldots < a_{K-1} < a_K = \infty$ and the kth interval is represented by $C_k = (a_{k-1}, a_k]$, where $k = 1, \ldots, K$. The data structure consists of two components: an overall *simple random sample* (SRS) of size n_0 and a stratified *supplemental sample* of size n_k randomly selected from the interval C_k:

(i) SRS Component: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i\right\}$, $i = 1, \ldots, n_0$;

(ii) Supplemental Component: for each k ($k = 1, \ldots, K$), $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid \max\{Y_{i1}, \ldots, Y_{ip}\} \in C_k\right\}$, $i = 1, \ldots, n_k$.

Note that $\boldsymbol{Y}_i$ is a p-dimensional response vector. The joint density of $(\boldsymbol{Y}_i, \boldsymbol{X}_i)$ can be written as $f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X}_i)$, where $f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})$ is the conditional density of $\boldsymbol{Y}_i$ given $\boldsymbol{X}_i$, $\boldsymbol{\theta}$ is the vector of regression coefficients of interest, and $g_{\boldsymbol{X}}(\boldsymbol{X}_i)$ is the marginal density of $\boldsymbol{X}_i$, which is independent of $\boldsymbol{\theta}$. The unknown distribution function of the covariates $\boldsymbol{X}_i$ is denoted as $G_{\boldsymbol{X}}(\boldsymbol{X}_i)$; $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ are assumed to be completely observed for all i. Without loss of generality, we assume that $p = 2$ and $K = 1$, meaning each

subject has two observations and the supplemental sample is randomly drawn from the upper tail of the distribution of $\max(\boldsymbol{Y})$, i.e. $C_1 = (a_1, \infty)$. That is, the ith subject in the supplemental sample is randomly selected if his/her maximum value of the responses is greater than a_1. For simplicity, we drop the subscription of a_1 and denote a. Thus, the likelihood function in correspondence to the *Multivariate-ODS* with a maximum selection criterion is

$$L_M(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})\ g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \times \left[\prod_{i=1}^{n_1} f(\boldsymbol{Y}_i, \boldsymbol{X}_i|\max(\boldsymbol{Y}_i) > a;\boldsymbol{\theta})\right] , \quad (2.1)$$

where the first bracket represents the quantity of the likelihood for the observations from the SRS of the *Multivariate-ODS* while the quantity in the second bracket is the likelihood contributed by the supplemental sample. Using Bayes' Law, the likelihood function can be further rewritten as

$$L_M(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} \frac{f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{1 - \Pr\{\max(\boldsymbol{Y}_i) < a\}}\right] . \quad (2.2)$$

To simplify notation, we define that

$$P_0(\boldsymbol{X};\boldsymbol{\theta}) = \Pr\{\max(\boldsymbol{Y}) < a|\boldsymbol{X}\} = \Pr\{Y_1 < a, Y_2 < a|\boldsymbol{X}\} = \int_{-\infty}^{a}\int_{-\infty}^{a} f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}) dY_1 dY_2 \quad (2.3)$$

and

$$\pi = \Pr\{\max(\boldsymbol{Y}) < a\} = \int_{\mathbb{X}} P_0(\boldsymbol{X};\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}) d\boldsymbol{X} \quad (2.4)$$

are the conditional and marginal probabilities that every element in $\boldsymbol{Y}$ is less than a, respectively. Rearranging the terms, we can then have

$$\begin{aligned} L_M(\boldsymbol{\theta}, G_{\boldsymbol{X}}) &= \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} f(\boldsymbol{Y}_1|\boldsymbol{X}_1;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i) \times \prod_{i=1}^{n_1} \frac{1}{1-\pi}\right] \\ &= \left[\prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})\right] \left[\left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \times (1-\pi)^{-n_1}\right] \\ &= L_{M1}(\boldsymbol{\theta}) \times L_{M2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) \ , \end{aligned} \tag{2.5}$$

where

$$L_{M1}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) \ , \tag{2.6}$$

$$L_{M2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \times (1-\pi)^{-n_1} \ . \tag{2.7}$$

There are several possible approaches that could be used to make inferences about $\boldsymbol{\theta}$. Without knowing $G_{\boldsymbol{X}}$, one of the naive approaches is to take the observations in the SRS portion of the *Multivariate-ODS* and derive a maximum likelihood estimator for $\boldsymbol{\theta}$. However, ignoring the information from the supplemental sample would lose accuracy and efficiency. Or, one could obtain $\boldsymbol{\theta}$ by maximizing the conditional likelihood based on the complete data in the *Multivariate-ODS*. Clearly, these two estimators are not the most efficient since the information regarding the supplemental sample is not fully accounted. If $G_{\boldsymbol{X}}(\boldsymbol{X})$ is parameterized to a parameter vector, say ξ, one could maximize the resulting $L_M(\boldsymbol{\theta}, \widehat{G}_{\boldsymbol{X}})$ subject to $(\boldsymbol{\theta}, \xi)$. However, misspecification of $G_{\boldsymbol{X}}$ could lead to erroneous conclusions so that such approach will be limited only if the form of $G_{\boldsymbol{X}}$ is correctly specified. As a result, a nonparametric modeling of $G_{\boldsymbol{X}}$ is desirable in this

case. Nevertheless, $G_{\boldsymbol{X}}$ cannot be easily factored out of $L_{M2}(\boldsymbol{\theta}, G_{\boldsymbol{X}})$ and is an infinite-dimensional nuisance parameter. Thus, to incorporate all the available information in the *Multivariate-ODS* data without specifying $G_{\boldsymbol{X}}$, one needs a new method that will be tractable both theoretically and computationally. We next describe a semiparametric empirical likelihood estimator, where $G_{\boldsymbol{X}}$ is left unspecified.

2.2.2 Semiparametric Empirical Likelihood Estimator for the Maximum Selection Criterion

Our plan is to obtain a profile log likelihood function for $\boldsymbol{\theta}$ by first fixing $\boldsymbol{\theta}$ and obtaining the empirical likelihood function of $G_{\boldsymbol{X}}$ in (2.5) (Vardi, 1985), which will be a function of $\boldsymbol{\theta}$ and π. Then we can obtain the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ by maximizing the resulting profile log likelihood function over $\boldsymbol{\theta}$.

First we maximize $L_M(\boldsymbol{\theta}, G_{\boldsymbol{X}})$, with $\boldsymbol{\theta}$ fixed, over all discrete distributions whose support includes the observed values by considering a discrete distribution function (i.e. a step function) which has all of its probability located at the observed data points (Vardi, 1985). Let $p_i = dG_{\boldsymbol{X}}(\boldsymbol{X}_i) = g_{\boldsymbol{X}}(\boldsymbol{X}_i)$, $i = 1, \ldots, n$, be the probability mass for the ith covariate vector. We search values for $\{\widehat{p}_i,\ \forall i\}$, which maximize the log likelihood function corresponding to (2.5)

$$l_M(\boldsymbol{\theta}, \{p_i\}) = \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i; \boldsymbol{\theta}) + \sum_{i=1}^{n} \ln p_i - n_1 \ln(1-\pi) \ , \tag{2.8}$$

subject to the following restrictions

$$\left\{ p_i \geq 0 \ \forall i, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i \Big(P_0(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi \Big) = 0 \right\} . \tag{2.9}$$

The above conditions reflect the fact that $G_{\boldsymbol{X}}$ is a discrete distribution function. For a fixed $\boldsymbol{\theta}$, there exists a unique maximum for $\{p_i\}$ in (2.8), subject to the constraints in (2.9) if 0 is inside the convex hull of the points $\{P_0(\boldsymbol{X}_i; \boldsymbol{\theta}), \forall i\}$ (Owen, 1988; Qin and Lawless, 1994). We consider the following Lagrange multiplier argument to maximize l_M over $\{p_i\}$,

$$\begin{aligned} H_M(\boldsymbol{\theta}, \ \{p_i\}, \ \eta, \ \lambda) \ &= \ \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i | \boldsymbol{X}_i; \boldsymbol{\theta}) + \sum_{i=1}^{n} \ln p_i - n_1 \ln (1 - \pi) \\ &\quad - \eta \left(\sum_{i=1}^{n} p_i - 1 \right) - n\lambda \sum_{i=1}^{n} p_i \Big(P_0(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi \Big), \end{aligned} \tag{2.10}$$

where η and λ are the Lagrange multipliers corresponding to the normalized restriction on the $\{\widehat{p}_i\}$. We take the derivative of H_M with respect to p_i and set it to equal 0 to obtain the score equation,

$$\frac{\partial H_M}{\partial p_i} = \frac{1}{p_i} - \eta - n\lambda \Big(P_0(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi \Big) = 0 \ , \tag{2.11}$$

which implies that

$$\widehat{p}_i = \frac{1}{\eta + n\lambda \Big(P_0(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi \Big)}. \tag{2.12}$$

Then, multiplying both sides of (2.12) by p_i, summing over i, and using the characteristics

of the restrictions, we then have

$$n - \eta - n\lambda \sum_{i=1}^{n} p_i \Big(P_0(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi \Big) = 0 \ , \tag{2.13}$$

which implies that $\widehat{\eta} = n$. Substituting $\widehat{\eta}$ back into (2.13), we have

$$\widehat{p}_i = \left\{ n \left[1 + \lambda \Big(P_0(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi \Big) \right] \right\}^{-1} \ . \tag{2.14}$$

Thus, we can obtain a function of $\boldsymbol{\theta}$, λ and π by replacing p_i in (2.8) with $\widehat{p}_i$. Let $\boldsymbol{\phi}_M^T = (\boldsymbol{\theta}^T, \lambda, \pi)$ represent the combined parameter vector and note that we are treating π as a parameter independent of $\boldsymbol{\theta}$ and so does λ. Thus, the resulting profile log likelihood function for $\boldsymbol{\phi}_M$ is

$$l_M(\boldsymbol{\phi}_M) = \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i | \boldsymbol{X}_i; \boldsymbol{\theta}) - \sum_{i=1}^{n} \ln \ n \left[1 + \lambda \Big(P_0(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi \Big) \right] - n_1 \ \ln(1 - \pi) \ , \tag{2.15}$$

which can be maximized over $\widehat{\boldsymbol{\phi}}_M$. We refer $\widehat{\boldsymbol{\phi}}_M$ as a *semiparametric empirical maximum likelihood estimator* (SEMLE). The Newton-Raphson algorithm is used to solve the score equations from (2.15) and find a root. In order to start the iterative procedure with consistent initial estimators, we will use the maximum likelihood estimators obtained from the likelihood function involving only the SRS portion of the *Multivariate-ODS* as our starting values for $\boldsymbol{\theta}$. These initial values are to be adequate for converging to the root corresponding to the SEMLE.

2.3 The Multivariate-ODS with a Summation Selection Criterion

2.3.1 Multivariate-ODS Likelihood for the Summation Selection Criterion

In this section, we present the estimator under a summation selection criterion introduced in Section 1.3.3. Recall that the domain of interest, the sums of responses $\boldsymbol{Y}_\bullet = \left\{\sum_{j=1}^{p} Y_{ij},\ \forall i\right\}$, can be partitioned into K mutually exclusive intervals by the known constants $-\infty = a_0 < a_1 < \ldots < a_{K-1} < a_K = \infty$, and the kth interval is denoted as $C_k = (a_{k-1}, a_k]$, $k = 1, \ldots, K$. The data structure of the *Multivariate-ODS* design under such selection criterion consists of two components: an overall *simple random sample* (SRS) of size n_0 (≥ 0) and a stratified *supplemental sample* of size n_k (≥ 0) randomly drawn from the interval, C_k:

(i) SRS Component: $\{\boldsymbol{Y}_i, \boldsymbol{X}_i\}$, $i = 1, \ldots, n_0$;

(ii) Supplemental Component: for each k ($k = 1, \ldots, K$), $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i | (\sum_{j=1}^{p} Y_{ij}) \in C_k\right\}$, $i = 1, \ldots, n_k$, $j = 1, \ldots, p$.

Without loss of generality, we assume that $p = 2$ and $K = 1$. That is, each individual has two observations and one only selects the supplemental sample in the upper tail of the distribution of $\{\sum_{j=1}^{p} Y_{ij}\ , \forall i\}$, i.e., $C_1 = (a_1, \infty)$. To simplify the notation, we denote a_1 as a. Let $n = n_0 + n_1$ be the total sample size of the *Multivariate-ODS* we observe. Let $L_S(\boldsymbol{\theta}, G_{\boldsymbol{X}})$ be the joint likelihood function for the observed data using the summation

selection criterion such that

$$L_S(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} f(\boldsymbol{Y}_j, \boldsymbol{X}_j|(Y_{i1}+Y_{i2}) > a;\boldsymbol{\theta})\right] \ , \qquad (2.16)$$

where the first bracket is the likelihood corresponding to the observations from the SRS portion of the *Multivariate-ODS* and the second quantity represents the likelihood contributions of the observations in the supplemental sample; $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X})$ is the joint density of $(\boldsymbol{Y}, \boldsymbol{X})$, where $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$ is the conditional density function of $\boldsymbol{Y}$ given $\boldsymbol{X}$, $\boldsymbol{\theta}$ is a vector of the regression coefficients of interest, and $g_{\boldsymbol{X}}(\boldsymbol{X})$ is the marginal density of $\boldsymbol{X}$, which is independent of $\boldsymbol{\theta}$. The corresponding unknown distribution function of $\boldsymbol{X}$ is denoted as $G_{\boldsymbol{X}}(\boldsymbol{X})$. Using Bayes' Law, we can rewrite (2.16) as

$$\begin{aligned}
L_S(\boldsymbol{\theta}, G_{\boldsymbol{X}}) &= \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} \frac{f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{1 - \Pr(Y_{i1}+Y_{i2} < a)}\right] \\
&= \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i) \times \prod_{i=1}^{n_1} \frac{1}{1 - \pi(\boldsymbol{\theta}, G_{\boldsymbol{X}})}\right] \\
&= \left[\prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})\right] \left[\left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \times (1-\pi)^{-n_1}\right] \\
&= L_1(\boldsymbol{\theta}) \times L_2(\boldsymbol{\theta}, G_{\boldsymbol{X}}) \ , \qquad (2.17)
\end{aligned}$$

where

$$L_{S1}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) \qquad (2.18)$$

and

$$L_{S2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \times (1-\pi)^{-n_1} \ . \qquad (2.19)$$

Note that for simplicity, we define that

$$P_0(\boldsymbol{X};\boldsymbol{\theta}) = \Pr\{Y_1 + Y_2 < a|\boldsymbol{X}\} = \int_{-\infty}^{\infty}\int_{-\infty}^{a-Y_2} f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})dY_1dY_2 \tag{2.20}$$

and

$$\pi = \pi(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \int_{\boldsymbol{X}} P_0(\boldsymbol{X};\boldsymbol{\theta})dG_{\boldsymbol{X}} \tag{2.21}$$

are the conditional and the marginal probabilities that the sum of the elements in $\boldsymbol{Y}$ is less than a, respectively.

2.3.2 Semiparametric Empirical Likelihood Estimator for the Summation Selection Criterion

The approach to obtain a profile log likelihood function for $\boldsymbol{\theta}$ is similar to the method presented in the previous section. We next give a brief derivation and will revisit this topic in details in Chapter 5. The log likelihood corresponding to (2.17) is

$$l_S(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) + \sum_{i=1}^{n} \ln g_{\boldsymbol{X}}(\boldsymbol{X}_i) - n_1 \ln(1-\pi)\ . \tag{2.22}$$

We use the similar argument for $G_{\boldsymbol{X}}(\boldsymbol{X})$ as discussed in the previous section for a maximum selection criterion. Let $p_i = dG_{\boldsymbol{X}}(\boldsymbol{X}_i) = g_{\boldsymbol{X}}(\boldsymbol{X}_i)$, $\forall i$, be the probability mass for the ith vector of covariates. We then search for values $\{\widehat{p}_i, \forall i\}$, maximizing the log

likelihood function

$$l_S(\boldsymbol{\theta}, \{p_i\}) = \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) + \sum_{i=1}^{n} \ln p_i - n_1 \ln(1-\pi) \tag{2.23}$$

subject to the following constraints:

$$\left\{ p_i \geq 0 \ \forall i,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i \Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi \Big) = 0 \right\}. \tag{2.24}$$

We then consider the following Lagrange function to maximize l_S over all $\{p_i, \forall i\}$,

$$\begin{aligned} H_S(\boldsymbol{\theta}, \{p_i\}, \mu, \lambda) \ &= \ \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) + \sum_{i=1}^{n} \ln p_i - n_1 \ln(1-\pi) \\ &\quad -\mu \left(\sum_{i=1}^{n} p_i - 1 \right) - n\lambda \sum_{i=1}^{n} p_i \Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi \Big) \ , \end{aligned} \tag{2.25}$$

where μ and λ are the Lagrange multipliers corresponding to the normalized restriction on the $\{\widehat{p}_i, \forall i\}$. With $\boldsymbol{\theta}$ fixed and taking the derivative of H_S with respect to p_i, the score equation is

$$\frac{\partial H_S}{\partial p_i} = \frac{1}{p_i} - \mu - n\lambda \Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi \Big) = 0 \ . \tag{2.26}$$

Together with the constraints in (2.24), it is straightforward to see that $\widehat{\mu} = n$ and

$$\widehat{p}_i = \left\{ n \left[1 + \lambda \Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi \Big) \right] \right\}^{-1} \ . \tag{2.27}$$

Substituting $\{\widehat{p}_i\}$ back into (2.23), we then have the resulting profile log likelihood function,

$$l_S(\boldsymbol{\phi}_{SM}) = \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) - \sum_{i=1}^{n} \ln n\Big[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\Big] - n_1 \ln(1-\pi)\ , \quad (2.28)$$

where $\boldsymbol{\phi}_{SM}^T = (\boldsymbol{\theta}^T, \lambda, \pi)$ is a combined parameter vector and λ and π are treated as the parameters independent of $\boldsymbol{\theta}$. The *semiparametric empirical maximum likelihood estimator* (SEMLE), $\widehat{\boldsymbol{\phi}}_{SM}$, is a maximizer for (2.28). The Newton-Raphson algorithm is used to solve the score equation from (2.28).

2.4 The Multivariate-ODS with a General Selection Criterion

2.4.1 Multivariate-ODS Likelihood for the General Selection Criterion

In this section, we present the proposed method with a more flexible and general selection criterion when considering the supplemental samples under the *Multivariate-ODS* design. To fix notation, let Y_{ij} be the j^{th} continuous outcome for the subject i, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$ ($p \geq 2$), and $\boldsymbol{X}_i$ be a vector of covariates for the i^{th} subject, which can include both discrete and continuous components. Recall the notation used in Section 1.3.3. We assume that $\boldsymbol{a} = \{a_j,\ j = 1, \ldots, p\}$ and $\boldsymbol{b} = \{b_j,\ j = 1, \ldots, p\}$, where a_j and b_j are known constants and satisfying $\{a_j > b_j, \forall j\}$, are the fixed cutpoints on the domain of $\boldsymbol{Y}_j = \{Y_{ij},\ \forall i\}$. Different from those in the previous *Multivariate-ODS* selection schemes, now the data structure under such selection criterion consists of three components: an overall *simple random sample* (SRS) of size n_0 (≥ 0), a *supplemental*

sample of size n_1 (≥ 0) conditional on $\{Y_{i1} > a_1, Y_{i2} > a_2, \ldots, Y_{ip} > a_p\}$, and another *supplemental sample* of size n_2 conditional on $\{Y_{i1} < b_1, Y_{i2} < b_2, \ldots, Y_{ip} < b_p\}$:

(i) SRS Component: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i\right\}$, $i = 1, \cdots, n_0$;

(ii) Supplemental Component 1: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid \{Y_{i1} > a_1, Y_{i2} > a_2, \ldots, Y_{ip} > a_p\}\right\}$, $i = 1, \ldots, n_1$ and $j = 1, \ldots, p$;

(iii) Supplemental Component 2: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid \{Y_{i1} < b_1, Y_{i2} < b_2, \ldots, Y_{ip} < b_p\}\right\}$, $i = 1, \ldots, n_2$ and $j = 1, \ldots, p$;

the total sample size in the *Multivariate-ODS* is $n = \sum_{k=0}^{2} n_k$.

Without loss of generality, we assume that $p = 2$, i.e., each individual has two responses, and therefore the cutpoints are set to be a_1, a_2, b_1 and b_2. The joint density of $(\boldsymbol{Y}, \boldsymbol{X})$ can be written as $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X})$, where $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$ is the conditional density function of $\boldsymbol{Y}$ given $\boldsymbol{X}$, $\boldsymbol{\theta}$ is a vector of the regression coefficients of interest, and $g_{\boldsymbol{X}}(\boldsymbol{X})$ is the marginal density of $\boldsymbol{X}$, which is independent of $\boldsymbol{\theta}$. The corresponding unknown distribution function of $\boldsymbol{X}$ can be denoted as $G_{\boldsymbol{X}}(\boldsymbol{X})$. We can then write the joint likelihood function, $L_{GL}(\boldsymbol{\theta}, G_{\boldsymbol{X}})$, for $(\boldsymbol{Y}, \boldsymbol{X})$ drawn into the *Multivariate-ODS* under the general selection criterion as

$$\begin{aligned} L_{GL}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) &= \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\theta})\right] \left[\prod_{i=1}^{n_1} f(Y_{i1}, Y_{i2}, \boldsymbol{X}_i; \boldsymbol{\theta} | Y_{i1} > a_1, Y_{i2} > a_2)\right] \\ &\quad \times \left[\prod_{i=1}^{n_2} f(Y_{i1}, Y_{i2}, \boldsymbol{X}_i; \boldsymbol{\theta} | Y_{i1} < b_1, Y_{i2} < b_2)\right] , \end{aligned} \tag{2.29}$$

where the first component is the likelihood from the SRS in the *Multivariate-ODS* while the last two parts are contributions from the two supplemental samples. By using Bayes'

Law, the above likelihood can be further rewritten as

$$\begin{aligned}
L_{GL}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) &= \left[\prod_{i=1}^{n_0} f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} \frac{f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{\Pr(Y_{i1} > a_1, Y_{i2} > a_2)}\right] \\
&\quad \times \left[\prod_{i=1}^{n_2} \frac{f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{\Pr(Y_{i1} < b_1, Y_{i2} < b_2)}\right] \\
&= \left[\prod_{i=1}^{n_0} f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} \frac{f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{\pi_1(\boldsymbol{\theta}, G_{\boldsymbol{X}})}\right] \\
&\quad \times \left[\prod_{i=1}^{n_2} \frac{f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{\pi_2(\boldsymbol{\theta}, G_{\boldsymbol{X}})}\right] \\
&= \left[\prod_{i=1}^{n} f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta})\right] \left[\left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \pi_1^{-n_1} \pi_2^{-n_2}\right] \\
&= L_{GL1}(\boldsymbol{\theta}) \times L_{GL2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) \ , \qquad (2.30)
\end{aligned}$$

where

$$L_{GL1}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) \qquad (2.31)$$

and

$$L_{GL2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \pi_1^{-n_1} \pi_2^{-n_2} \ ; \qquad (2.32)$$

and for simplicity, we define

$$P_1(\boldsymbol{X};\boldsymbol{\theta}) = \Pr\{Y_1 > a_1, Y_2 > a_2|\boldsymbol{X}\} = \int_{a_1}^{\infty} \int_{a_2}^{\infty} f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}) dY_1 dY_2 \qquad (2.33)$$

and

$$\pi_1 = \pi_1(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \int_{\mathbb{X}} P_1(\boldsymbol{x};\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}) d\boldsymbol{X} \qquad (2.34)$$

are the conditional and marginal probabilities that Y_1 and Y_2 satisfy $\{Y_1 > a_1, Y_2 > a_2\}$;

$$P_2(\boldsymbol{X};\boldsymbol{\theta}) = \Pr\{Y_1 < b_1, Y_2 < b_2|\boldsymbol{X}\} = \int_{-\infty}^{b_1}\int_{\infty}^{b_2} f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})dY_1dY_2 \tag{2.35}$$

and

$$\pi_2 = \pi_2(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \int_{\mathbb{X}} P_2(\boldsymbol{x};\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X})d\boldsymbol{X} \ , \tag{2.36}$$

are the conditional and marginal probabilities for $\{Y_1 < b_1, Y_2 < b_2\}$.

Using similar arguments for $G_{\boldsymbol{X}}(\boldsymbol{X})$, we avoid specifying a parametric form for $G_{\boldsymbol{X}}$ and consider a semiparametric empirical likelihood approach to maximizing $L_{GL}(\boldsymbol{\theta}, G_{\boldsymbol{X}})$ with respect to $(\boldsymbol{\theta}, G_{\boldsymbol{X}})$, which is desirable and tractable both theoretically and computationally.

2.4.2 Semiparametric Empirical Likelihood Estimator for the General Selection Criterion

We follow a similar approach to derive a profile log likelihood function for $\boldsymbol{\theta}$ in (2.30) as discussed in the previous sections. We will elaborate the proposed method for the general selection criterion again in Chapter 6.

We first maximize $L(\boldsymbol{\theta}, G_{\boldsymbol{X}})$, with $\boldsymbol{\theta}$ fixed, by considering a discrete distribution function (i.e. a step function) which has all of its probability located at the observed data points (Vardi, 1985) to over all discrete distributions whose support includes the observed values. Let $p_i = dG_{\boldsymbol{X}}(\boldsymbol{X}_i) = g_{\boldsymbol{X}}(\boldsymbol{X}_i)$, $i = 1, \ldots, n$, be the probability mass for the ith covariate vector. We want to search for values $\{\widehat{p}_i, \forall i\}$ which maximize the log

likelihood function with respect to (2.30)

$$l_{GL}(\boldsymbol{\theta}, \{p_i\}) = \sum_{i=1}^{n} \ln f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i; \boldsymbol{\theta}) + \sum_{i=1}^{n} \ln p_i - n_1 \ln \pi_1 - n_2 \ln \pi_2 \ , \tag{2.37}$$

subject to the following constraints:

$$\left\{ \{p_i\} \geq 0 \ \forall i, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i \Big(P_1(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi_1 \Big) = 0, \ \sum_{i=1}^{n} p_i \Big(P_2(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi_2 \Big) = 0 \right\} . \tag{2.38}$$

The above conditions reflect the fact that $G_{\boldsymbol{X}}$ is a discrete distribution function. For a fixed $\boldsymbol{\theta}$, there exists a unique maximum for $\{p_i\}$ in (2.37) subject to the constraints in (2.38) if 0 is inside the convex hull of the points $\{P_1(\boldsymbol{X}_i; \boldsymbol{\theta}), \ \forall \, i\}$ and $\{P_2(\boldsymbol{X}_i; \boldsymbol{\theta}), \ \forall \, i\}$ (Qin and Lawless, 1994). We use the Lagrange multiplier argument to maximize $l_{GL}(\boldsymbol{\theta}, \{p_i\})$ over all $\{p_i, \forall i\}$,

$$\begin{aligned} H_{GL}(\boldsymbol{\theta}, \{p_i\}, \delta, \lambda_1, \lambda_2) \ &= \ \sum_{i=1}^{n} \ln p_i - n_1 \ln \pi_1 - n_2 \ln \pi_2 - \delta \Big(\sum_{i=1}^{n} p_i - 1 \Big) \\ &\quad - n\lambda_1 \sum_{i=1}^{n} p_i \Big(P_1(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi_1 \Big) - n\lambda_2 \sum_{i=1}^{n} p_i \Big(P_2(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi_2 \Big) \ , \end{aligned}$$

where the restrictions that $\pi_1 = \sum_{i=1}^{n} p_i P_1(\boldsymbol{X}_i; \boldsymbol{\theta})$ and $\pi_2 = \sum_{i=1}^{n} p_i P_2(\boldsymbol{X}_i; \boldsymbol{\theta})$ are reflected; δ, λ_1 and λ_2 are the Lagrange multipliers corresponding to the normalized restriction on the $\{\widehat{p}_i, \forall i\}$. After taking the derivative of H_{GL} with respect to p_i and applying the constraints in (2.38), we obtain $\widehat{\delta} = n$ and

$$\widehat{p}_i = \left\{ n \ \left[1 + \lambda_1 \Big(P_1(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi_1 \Big) + \lambda_2 \Big(P_2(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi_2 \Big) \right] \right\}^{-1} , \tag{2.39}$$

where $i = 1, \ldots, n$. We can then substitute $\widehat{p}_i$ back to (2.30) to obtain a function of $\boldsymbol{\theta}$, π_1, π_2, λ_1 and λ_2. Define $\boldsymbol{\phi}_{GL}^T = (\boldsymbol{\theta}^T, \pi_1, \pi_2, \lambda_1, \lambda_2)$, representing the combined parameter vector and note that we are treating λ_1, λ_2, π_1 and π_2 as parameters independent of $\boldsymbol{\theta}$. Thus, the resulting profile log likelihood function for $\boldsymbol{\phi}_{GL}$ is

$$\begin{aligned} l_{GL}(\boldsymbol{\phi}_{GL}) &= \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) - \sum_{i=1}^{n} \ln\ n\Big[1 + \lambda_1\Big(P_1(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_1\Big) + \lambda_2\Big(P_2(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_2\Big)\Big] \\ &\quad -n_1 \ln \pi_1 - n_2 \ln \pi_2\ . \end{aligned} \tag{2.40}$$

We refer $\widehat{\boldsymbol{\phi}}_{GL}$ as the *semiparametric empirical maximum likelihood estimator* (SEMLE), which is a maximizer of (2.40). The Newton-Raphson algorithm will be used to solve the score equations with respect to (2.40) and the initial values to start the iterative procedure will be the maximum likelihood estimators obtained from the first term in the likelihood (2.29) which involves only the SRS part of the *Multivariate-ODS*.

CHAPTER 3

ASYMPTOTIC PROPERTIES OF THE SEMIPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATOR FOR THE MULTIVARIATE-ODS WITH THE MAXIMUM SELECTION CRITERION

3.1 Introduction

In this chapter, we will derive the asymptotic properties of the semiparametric empirical maximum likelihood estimator, $\boldsymbol{\phi}_M$, for the *Multivariate-ODS* design with a maximum selection criterion presented in Section 2.2. We will demonstrate the existence and consistency of these estimators and derive the asymptotic normal distribution for this estimator; furthermore, we will derive a consistent estimator for the asymptotic variance-covariance matrix. In Section 3.2, we introduce some additional and useful notations which will be used in the proofs later and present several assumptions required for the proofs along with notational conventions. In addition, we state some useful preliminary results which will be useful in the proofs. In Section 3.3, we demonstrate the main results for $\boldsymbol{\phi}_M$ regarding the consistency, asymptotic normality, and a consistent estimator for the asymptotic variance-covariance matrix as three theorems, respectively. Rigorous proofs of the main results will be provided in Sections 3.4 and 3.5.

3.2 Notation, Assumptions, and Useful Preliminary Results

3.2.1 Notation

Recall from Section 2.2.2 that the profile log-likelihood function for the *Multivariate-ODS* with a maximum selection criterion is

$$l_M(\boldsymbol{\phi}_M) = \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) - \sum_{i=1}^{n} \ln\ n\Big[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\Big] - n_1\ \ln(1-\pi)\ , \quad (3.1)$$

where $\boldsymbol{\phi}_M = (\boldsymbol{\theta}^T, \pi, \lambda)^T$ represents the combined parameter vector,

$$\begin{aligned} P_0(\boldsymbol{X};\boldsymbol{\theta}) &= \Pr\{\max(\boldsymbol{Y}) < a|\boldsymbol{X}\} = \Pr\{Y_1 < a, Y_2 < a|\boldsymbol{X}\} \\ &= \int_{-\infty}^{a}\int_{-\infty}^{a} f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})dY_1dY_2 \end{aligned} \quad (3.2)$$

and

$$\pi = \pi(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \Pr\{\max(\boldsymbol{Y}) < a\} = \int_{\mathbb{X}} P_0(\boldsymbol{X};\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}) d\boldsymbol{X} \quad (3.3)$$

are conditional and marginal probabilities, respectively. Here we assume that $\boldsymbol{\theta}$ is a p-dimensional parameter vector so that $\boldsymbol{\phi}_M$ is the combined parameter vector of dimension $(p+2)\times 1$.

We indicate $\boldsymbol{\phi}_M^0$ as the true parameter vector of interest containing $\boldsymbol{\theta}^0$, π^0 and λ^0, where π^0 is the true marginal probability that the maximum value of the observations from each individual is less than the cutpoint, a, and λ is the Lagrange multiplier. For any function $h(\boldsymbol{Y}, \boldsymbol{X})$, $\mathrm{E}\Big[h(\boldsymbol{Y}, \boldsymbol{X})\Big]$ denotes expectation conditional on $\{\max(\boldsymbol{Y}) < a\}$

so that

$$\mathrm{E}\Big[h(\boldsymbol{Y},\boldsymbol{X})\Big] = \int_{\mathbb{X}} \frac{1}{\pi^0} \int_{-\infty}^{a} \int_{-\infty}^{a} h(\boldsymbol{y},\boldsymbol{x}) f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}^0) d\boldsymbol{y} dG_{\boldsymbol{X}}(\boldsymbol{x}) \ .$$

3.2.2 Assumptions

We assume the following regularity conditions throughout this chapter:

A1. As $n \to \infty$, $\dfrac{n_1}{n} \to \gamma > 0$ and $\dfrac{n_0}{n} \to 1-\gamma > 0$, where γ represents the supplemental sampling fraction.

A2. The parameter space, $\boldsymbol{\Theta}$, is a compact subset of $\mathbb{R}^p$; $\boldsymbol{\theta}^0$ lies in the interior of $\boldsymbol{\Theta}$; the covariate space, $\mathbb{X}$, is a compact subset of $\mathbb{R}^q$, for some $q \geq 1$.

A3. $f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$ is continuous in both $\boldsymbol{y}$ and $\boldsymbol{\theta}$ and is strictly positive for all $\boldsymbol{y} \in \mathbb{Y}$, $\boldsymbol{x} \in \mathbb{X}$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Furthermore, the partial derivatives, $\partial f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})/\partial\theta_i$ and $\partial^2 f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})/\partial\theta_i\partial\theta_j$, for $i,\ j = 1,\ldots,p$, exist and are continuous for all $\boldsymbol{y} \in \mathbb{Y}$, $\boldsymbol{x} \in \mathbb{X}$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

A4. Interchanges of differentiation and integration of $f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$ are valid for the first and second partial derivatives with respect to $\boldsymbol{\theta}$.

A5. The expected value matrix,

$$\mathrm{E}\left[-\frac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right],$$

is finite and positive definite at $\boldsymbol{\theta}^0$.

A6. There exists a $\delta > 0$ such that for the set $A = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}^0| \leq \delta\}$,

$$\mathrm{E}\left[\sup_A \left| \frac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} \right|\right] < \infty,$$

for i, $j = 1, \ldots, p$.

A7. The derivatives, $\dfrac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\theta_j}$, $j = 1, \ldots, p$, are linearly independent. That is, suppose $\boldsymbol{t}$ is any $(p \times 1)$ such that

$$\sum_{j=1}^{p} t_j \frac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\theta_j} = 0$$

for almost all $\boldsymbol{x} \in \mathbb{X}$ if $\boldsymbol{t} = \mathbf{0}$.

Remarks Regarding the Assumptions:

(i) The compactness condition in A2, from Cosslett (1981b) which follows Jennrich (1969) and Amemiya (1973), is imposed to obtain uniform convergence properties, simplifying the complexity of the proofs.

(ii) We can extend the condition in A3 (first discussed in Weaver's Dissertation (2001)) from the conditional density $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$ directly to the marginal density $f(\boldsymbol{Y};\boldsymbol{X},\boldsymbol{\theta})$. A simple proof is as follows. Since $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, i.e. for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\left| f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}_1) - f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}_2) \right| < \epsilon$$

whenever $|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| < \delta$. Then

$$\begin{aligned} & \left| \int_{\mathbb{X}} f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}_1) dG_{\boldsymbol{X}}(\boldsymbol{X}) - \int_{\mathbb{X}} f(\boldsymbol{Y}|\boldsymbol{X}\boldsymbol{\theta}_2) dG_{\boldsymbol{X}}(\boldsymbol{X}) \right| \\ \leq & \int_{\mathbb{X}} \left| f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}_1) - f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}_2) \right| dG_{\boldsymbol{X}}(\boldsymbol{X}) \\ < & \int_{\mathbb{X}} \epsilon dG_{\boldsymbol{X}}(\boldsymbol{X}) \\ = & \epsilon \ , \end{aligned}$$

whenever $|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| < \delta$. This implies that $f(\boldsymbol{Y};\boldsymbol{X},\boldsymbol{\theta})$ is continuous of $\boldsymbol{\theta}$ as well. Similarly, this result can be applied to the first and second partial derivatives of $f(\boldsymbol{Y};\boldsymbol{X},\boldsymbol{\theta})$ is continuous of $\boldsymbol{\theta}$.

(iii) Assumptions A4 and A5 are standard assumptions.

(iv) Uniform convergence of the second derivative matrix of the log likelihood function to the information matrix can be obtained by using assumption A6. Note that this assumption can be directly extended to the marginal density of $f(\boldsymbol{Y};\boldsymbol{X},\boldsymbol{\theta})$ although this is stated in terms of the conditional density, $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$.

(v) A7, directly following Cosslett (1981b), is used to obtain the limiting form of the Hessian matrix of the profile likelihood function with respect to $\boldsymbol{\theta}$ being positive definite.

(vi) It can be shown that assumptions A2 through A6 provide sufficient conditions such that the usual consistency and asymptotic normality for maximum likelihood estimators hold for $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$ (see Foutz, 1977 and Sen and Singer, 1993). (The

proofs will be not be shown here.)

3.2.3 Preliminary Results

Before we prove three theorems in the next section, we will state some lemmas which are useful in our proofs. These results are well-known and frequently applied. The proofs of these lemmas can be found in the references provided.

The results of the following lemma are often used and its proof can be found in Lehmann (1999), page 50.

Lemma 3.1: *If X_n and Y_n are two sequences of random variables and a and b are two constants such that $X_n \xrightarrow{p} a$ and $Y_n \xrightarrow{p} b$, then*

$$X_n + Y_n \xrightarrow{p} a + b \ ,$$

$$X_n \times Y_n \xrightarrow{p} a \times b \ , and$$

$$X_n / Y_n \xrightarrow{p} a/b \ if \ b \neq 0 \ .$$

Lemma 3.2 below is taken directly from Weaver's (2001) Lemma 3.1, which is a restatement of Jennrich's (1969) Theorem 2. This lemma established the uniform convergence of a sample mean of functions bounded to its expected value in a sense of Law of Large Numbers. The proof of a similar result can be found in Jennrich (1969). Note that stronger results were established by Rao (1962).

Lemma 3.2: *Let $\boldsymbol{\Theta}$ be a compact subset of a Euclidean space and let Ψ be a Euclidean space. Let $g(\boldsymbol{\psi}, \boldsymbol{\theta})$ be a continuous function of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ for each $\boldsymbol{\psi} \in \Psi$, such that $|g(\boldsymbol{\psi}, \boldsymbol{\theta})|$ is bounded by some function $h(\boldsymbol{\psi})$ for all $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$, where $h(\boldsymbol{\psi})$ is integrable with respect to a probability distribution function F on Psi. If $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots$ is a random sample from F, then for almost every sequence $\{\boldsymbol{\psi}_i\}$,*

$$\frac{1}{N} \sum_{i=1} N g(\boldsymbol{\psi}_i; \boldsymbol{\theta}) \xrightarrow{p} \int_{\Psi} g(\boldsymbol{\psi}; \boldsymbol{\theta}) dF(\boldsymbol{\psi})$$

uniformly for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

A principal tool in the proof of consistency of our proposed estimators is the Inverse Function Theorem. The version of the theorem given below is taken from Foutz (1977, pp. 147).

The Inverse Function Theorem: *Suppose $\boldsymbol{f}$ is a mapping from an open set $\boldsymbol{\Theta}$ in Euclidean r space, E^r into E^r, the partial derivatives of $\boldsymbol{f}$ exist and are continuous on $\boldsymbol{\Theta}$, and the matrix of derivatives $\boldsymbol{f}'(\boldsymbol{\theta}^*)$ has inverse $\boldsymbol{f}'(\boldsymbol{\theta}^*)^{-1}$ for some $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. Write*

$$\lambda = 1/(4\|\boldsymbol{f}'(\boldsymbol{\theta}^*)^{-1}\|) \ .$$

Use the continuity of the elements of $\boldsymbol{f}'(\boldsymbol{\theta}^*)$ to fix a neighborhood $\boldsymbol{U}_\delta$ of $\boldsymbol{\theta}^*$ of sufficiently small radius $\delta > 0$ to insure that $\|\boldsymbol{f}'(\boldsymbol{\theta}) - \boldsymbol{f}'(\boldsymbol{\theta}^*)\| < 2\lambda$, whenever $\boldsymbol{\theta} \in \boldsymbol{U}_\delta$. Then (a) for every $\boldsymbol{\theta}_1, \ \boldsymbol{\theta}_2$ in $\boldsymbol{U}_\delta$,

$$|\boldsymbol{f}(\boldsymbol{\theta}_1) - \boldsymbol{f}(\boldsymbol{\theta}_2)| \geq 2\lambda |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| \ ,$$

and (b) the image set $\boldsymbol{f}(\boldsymbol{U}_\delta)$ *contains the open neighborhood with radius* $\lambda\delta$ *about* $\boldsymbol{f}(\boldsymbol{\theta}^*)$.

Conclusion (a) in the theorem above guarantees that $\boldsymbol{f}$ is one-to-one on $\boldsymbol{U}_\delta$ and that $\boldsymbol{f}^{-1}$ is well-defined on the image set $\boldsymbol{f}(\boldsymbol{U}_\delta)$. The theorem is proven in this form in Rudin and Walter (1964, pp. 193-194).

Lemma 3.3 below is more generally restated by Weaver (in his dissertation, 2001) from Foutz' (1977) result which established the existence of a unique consistent solution to the likelihood functions by using the Inverse Function Theorem. Similarly, we will weaken the requirement of the matrix of second derivatives of the log likelihood function being negative definite; in stead, we only require that the limiting second derivative matrices be invertible. This has been shown to be a sufficient condition for Foutz' result in Weaver's dissertation (2001, pp. 56 - 57).

Lemma 3.3: *Let* $\{\boldsymbol{f}_N(\boldsymbol{\theta})\}$ *be a sequence of continuous, random, vector-valued functions of* $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$. *Suppose that, for all N, the partial derivatives of* $\boldsymbol{f}_N(\boldsymbol{\theta})$ *with respect to* $\boldsymbol{\theta}$ *exist and are continuous on* $\boldsymbol{\Theta}$*; let* $\boldsymbol{f}'_N(\boldsymbol{\theta})$ *be the* $p \times p$ *dimensional matrix containing these partial derivatives. Let* $\boldsymbol{H}(\boldsymbol{\theta})$ *be a* $p \times p$ *dimensional matrix whose elements are continuous functions of* $\boldsymbol{\theta}$ *such that* $\boldsymbol{H}^{-1}(\boldsymbol{\theta}^*)$ *exists for some* $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. *Suppose that* $\boldsymbol{f}'_N(\boldsymbol{\theta}) \xrightarrow{p} \boldsymbol{H}(\boldsymbol{\theta})$ *as* $N \to \infty$ *uniformly for* $\boldsymbol{\theta}$ *in an open neighborhood around* $\boldsymbol{\theta}^*$. *Furthermore, assume that* $\boldsymbol{f}_N(\boldsymbol{\theta}^*) \xrightarrow{p} \boldsymbol{0}$. *Then, there exists a sequence* $\{\widehat{\boldsymbol{\phi}}_N\}$ *such that*

$$\boldsymbol{f}_N(\widehat{\boldsymbol{\theta}}_N) = \boldsymbol{0} \tag{3.4}$$

with probability going to one as $N \to \infty$, *and*

$$\widehat{\boldsymbol{\theta}}_N \xrightarrow{p} \boldsymbol{\theta}^*. \tag{3.5}$$

If another sequence $\overline{\boldsymbol{\theta}}_N$ *also satisfies (3.4) and (3.5), then* $\widehat{\boldsymbol{\theta}}_N = \overline{\boldsymbol{\theta}}_N$ *with probability going to one as* $N \to \infty$.

The lemma originally stated and proven by Amemiya (1973, Lemma 4) is slightly modified in the following to help us prove the asymptotic normality using the consistency result of an estimator.

Lemma 3.4: *Let* $\boldsymbol{f}_N(\boldsymbol{\theta})$, $N = 1, \ldots, \infty$, *be measurable functions on a measurable space* $\boldsymbol{\Omega}$ *and continuous functions for* $\boldsymbol{\theta}$ *in a compact set* $\boldsymbol{\Theta}$. *If* $\boldsymbol{f}_N(\boldsymbol{\theta})$ *converges to* $\boldsymbol{f}(\boldsymbol{\theta})$ *with probability approaching one uniformly for all* $\boldsymbol{\theta}$ *in* $\boldsymbol{\Theta}$ *as* $N \longrightarrow \infty$, *and if* $\widetilde{\boldsymbol{\theta}}_N$ *converges to* $\boldsymbol{\theta}^*$ *with probability approaching one, then* $\boldsymbol{f}_N(\widetilde{\boldsymbol{\theta}}_N)$ *converges to* $\boldsymbol{f}(\boldsymbol{\theta}^*)$ *with probability approaching one.*

3.3 Main Results for the SEMLE

We state the three main results for the SEMLE, $\widehat{\boldsymbol{\phi}}_M$, under the *Multivariate-ODS* with a maximum selection criterion in this section. Theorems 3.1 and 3.2 demonstrate the consistency and asymptotic normality, respectively; Theorem 3.3 establishes a consistent estimator for the asymptotic variance-covariance matrix derived in Theorem 3.2. Rigorous and detailed proofs of these theorems are provided in the following sections.

Theorem 3.1 (Consistency of the SEMLE $\widehat{\boldsymbol{\phi}}_M$): *With probability going to 1 as $n \to \infty$, there exists a sequence $\{\widehat{\boldsymbol{\phi}}_M\}$ of solutions to the score equations with respect to (3.1) such that $\widehat{\boldsymbol{\phi}}_M \xrightarrow{p} \boldsymbol{\phi}^0$, where $\boldsymbol{\phi}^0$ is the true parameter vector of interest. If another sequence $\{\overline{\boldsymbol{\phi}}_M\}$ of solutions to the score equations exists such that $\overline{\boldsymbol{\phi}}_M \xrightarrow{p} \boldsymbol{\phi}^0_M$, then $\overline{\boldsymbol{\phi}}_M = \widehat{\boldsymbol{\phi}}_M$ with probability going to 1 as $n \to \infty$.*

Theorem 3.2 (Asymptotic Normality of the SEMLE $\widehat{\boldsymbol{\phi}}_M$): *$\widehat{\boldsymbol{\phi}}_M$ has the following asymptotic normal distribution:*

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_M - \boldsymbol{\phi}^0_M) \xrightarrow{D} N_{(p+2)}(\mathbf{0},\ \boldsymbol{\Sigma}(\boldsymbol{\phi}^0_M)) \ ,$$

with the asymptotic variance-covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{V}\boldsymbol{J}^{-1} \ , \tag{3.6}$$

where

$$\boldsymbol{J} = -\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}^0_M)}{\partial \boldsymbol{\phi}_M \ \partial \boldsymbol{\phi}_M^T}$$

and

$$\boldsymbol{V} = Var\left[\frac{\partial l_M(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}^0_M)}{\partial \boldsymbol{\phi}_M}\right] \ ,$$

where $\widetilde{l}_M$ is the limiting form of l_M.

Theorem 3.3 (A Consistent Estimator for the Asymptotic Variance-Covariance Matrix): *A consistent estimator for the variance-covariance matrix shown in Equation*

(3.6) is

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\phi}}_M) = \widehat{\boldsymbol{J}}^{-1}(\widehat{\boldsymbol{\phi}}_M)\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_M)\widehat{\boldsymbol{J}}^{-1}(\widehat{\boldsymbol{\phi}}_M),$$

where

$$\widehat{\boldsymbol{J}}(\boldsymbol{\phi}_M) = -\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\phi}_M \, \partial \boldsymbol{\phi}_M^T}$$

and

$$\widehat{\boldsymbol{V}}(\boldsymbol{\phi}_M) = \frac{1}{n} \, \widehat{Var}_{\{i\}}\left[\frac{\partial l_M(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\phi}_M}\right].$$

3.4 Consistency of the SEMLE

Before we prove Theorem 3.1 in Section 3.4.4, we begin with the first and second derivatives of the log-likelihood function in Section 3.4.1, which will be useful for the derivation of consistency and asymptotic normality later. In order to apply the Inverse Function Theorem and Lemma 3.3 to the proof of the consistency, we first show that the log-likelihood function asymptotically has a root at the true parameter in Section 3.4.2, and then explore the nature of this root through its Hessian matrix in Section 3.4.3. Finally, we wrap up all of these results to prove Theorem 3.1.

3.4.1 First and Second Derivatives of the Log-likelihood Function

Recall the profile log-likelihood function in (3.1),

$$\begin{aligned} l_M(\boldsymbol{\phi}_M) &= \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) - \sum_{i=1}^{n} \ln\ n\Big[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\Big] - n_1 \ln(1-\pi) \\ &= \sum_{i=1}^{n} \ln h(\boldsymbol{Y}_i,\boldsymbol{X}_i;\boldsymbol{\phi}_M) - n\ln n - n_1\ln(1-\pi)\ , \end{aligned} \tag{3.7}$$

where

$$h(\boldsymbol{Y}_i,\boldsymbol{X}_i;\boldsymbol{\phi}_M) = \frac{f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})}{1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)}\ , \tag{3.8}$$

and $\boldsymbol{\phi}_M = (\boldsymbol{\theta}^T, \pi, \lambda)^T$.

The first and second derivatives with respect to each parameter in $\boldsymbol{\phi}_M$ are calculated in the following:

$$\begin{aligned} \frac{\partial l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^{n} \frac{\partial \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \sum_{i=1}^{n} \frac{\partial \ln\Big[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\Big]}{\partial \boldsymbol{\theta}} \\ &= \sum_{i=1}^{n} \frac{\partial \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \sum_{i=1}^{n} \frac{\lambda \dfrac{\partial P_0(\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)}\ ; \end{aligned} \tag{3.9}$$

$$\begin{aligned} \frac{\partial l_M(\boldsymbol{\phi}_M)}{\partial \pi} &= -\sum_{i=1}^{n} \frac{\partial \ln\Big[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\Big]}{\partial \pi} + \frac{n_1}{1-\pi} \\ &= \sum_{i=1}^{n} \frac{\lambda}{1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)} + \frac{n_1}{1-\pi}\ ; \end{aligned} \tag{3.10}$$

$$\begin{aligned}\frac{\partial l_M(\boldsymbol{\phi}_M)}{\partial \lambda} &= -\sum_{i=1}^{n} \frac{\partial \ln\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]}{\partial \lambda} \\ &= -\sum_{i=1}^{n} \frac{P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi}{1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)} \; ; \end{aligned} \tag{3.11}$$

$$\begin{aligned}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} &= \sum_{i=1}^{n} \frac{\partial^2 \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} - \sum_{i=1}^{n} \frac{\partial^2 \ \ln\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right])}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} \\ &= \sum_{i=1}^{n} \frac{\partial^2 \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} - \sum_{i=1}^{n} \frac{\lambda\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]\frac{\partial^2 P_0(\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}}{\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]^2} \\ &\quad + \sum_{i=1}^{n} \frac{\lambda^2 \frac{\partial P_0(\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial P_0(\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}}{\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]^2} \; ; \end{aligned} \tag{3.12}$$

$$\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}\partial \pi} = -\sum_{i=1}^{n} \frac{\lambda^2 \frac{\partial P_0(\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]^2} \; ; \tag{3.13}$$

$$\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}\partial \lambda} = -\sum_{i=1}^{n} \frac{\frac{\partial P_0(\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]^2} \; ; \tag{3.14}$$

$$\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \pi^2} = \sum_{i=1}^{n} \frac{\lambda^2}{\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]^2} + \frac{n_1}{(1-\pi)^2} \; ; \tag{3.15}$$

$$\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \pi \partial \lambda} = \sum_{i=1}^{n} \frac{1}{\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]^2} \; ; \tag{3.16}$$

$$\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \lambda^2} = \sum_{i=1}^{n} \frac{\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi\Big)^2}{\Big[1 + \lambda\Big[P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi\Big)\Big]^2} \ . \tag{3.17}$$

3.4.2 Limiting Form of the Profile Log-likelihood Function

From the restrictions described in (2.9), we can obtain the following identities:

$$\sum_{i=1}^{n} \frac{1}{n\Big[1 + \lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi\Big)\Big]} = 1 \tag{3.18}$$

and

$$\sum_{i=1}^{n} \frac{P_0(\boldsymbol{X}_i;\boldsymbol{\theta})}{n\Big[1 + \lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi\Big)\Big]} = \pi \ . \tag{3.19}$$

Moreover, we can obtain the score equation of π by setting $1/n$ times Equation (3.10) to equal zero, which becomes

$$\sum_{i=1}^{n} \frac{\lambda}{n\Big[1 + \lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi\Big)\Big]} = -\frac{n_1}{n}\frac{1}{1-\pi} \ . \tag{3.20}$$

Note that the left-hand side equals to λ from the identity (3.18). After rearranging both sides,

$$\lambda(1-\pi) = -\frac{n_1}{n} \ ,$$

and in conjunction with assumption A1, it is easy to see that $\lambda(1-\pi)$ converges to $-\gamma$ as n goes to ∞. As a result, we can further have the following identities,

$$\lambda \xrightarrow{p} \frac{-\gamma}{1-\pi^0} \tag{3.21}$$

and

$$\pi \xrightarrow{p} \frac{\gamma+\lambda^0}{\lambda^0} \ , \tag{3.22}$$

which are useful in the demonstration later. Multiplying Equation (3.9), the first derivative with respect to $\boldsymbol{\theta}$, by $1/n$ is

$$\frac{1}{n}\frac{\partial l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda\dfrac{\partial P_0(\boldsymbol{X}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)} \ . \tag{3.23}$$

Using assumption A1 and the Law of Large Numbers, we have

$$\frac{1}{n}\frac{\partial l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}} \xrightarrow{p} \frac{\partial \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}} \ , \tag{3.24}$$

where

$$\begin{aligned}
\frac{\partial \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}} &= \mathrm{E}\left[\frac{\partial \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\lambda \dfrac{\partial P_0(\boldsymbol{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{1+\lambda\Big(P_0(\boldsymbol{X};\boldsymbol{\theta})-\pi\Big)}\right] \\
&= \int_{\mathbb{X}}\int_{-\infty}^{a}\int_{-\infty}^{a} \frac{f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}^0)}{\pi^0}\frac{\partial \ln f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\, d\boldsymbol{y}\, dG_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\quad - \int_{\mathbb{X}}\int_{-\infty}^{a}\int_{-\infty}^{a} \frac{f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}^0)}{\pi^0}\frac{\lambda \dfrac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{1+\lambda\Big(P_0(\boldsymbol{x};\boldsymbol{\theta})-\pi\Big)}\, d\boldsymbol{y}\, dG_{\boldsymbol{X}}(\boldsymbol{x})\ .
\end{aligned} \tag{3.25}$$

At the true parameter values, the equation above becomes

$$\begin{aligned}
\left.\frac{\partial \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0 \\ \pi=\pi^0 \\ \lambda=\lambda^0}} &= \int_{\mathbb{X}} \frac{1}{\pi^0}\frac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\, dG_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\quad - \int_{\mathbb{X}} \frac{P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\pi^0}\frac{\lambda^0 \dfrac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}}{1+\lambda^0\Big(P_0(\boldsymbol{x};\boldsymbol{\theta}^0)-\pi^0\Big)}\, dG_{\boldsymbol{X}}(\boldsymbol{x}) \\
&= \frac{1}{\pi^0}\frac{\partial}{\partial \boldsymbol{\theta}}(\pi^0) - \frac{\lambda^0}{\pi^0}\left[\frac{\gamma+\lambda^0}{\lambda^0}\right]\int_{\mathbb{X}} \frac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} dG_{\boldsymbol{X}}(\boldsymbol{x}) \\
&= \mathbf{0}\ ,
\end{aligned} \tag{3.26}$$

since A4 is used and

$$\int_{\mathbb{X}}\int_{-\infty}^{a}\int_{-\infty}^{a} f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}^0) d\boldsymbol{y} dG_{\boldsymbol{X}}(\boldsymbol{x}) = \pi^0\ .$$

From Equation (3.10), we note that $1/n$ times the first derivative with respect to π is

$$\frac{1}{n}\frac{\partial l_M(\boldsymbol{\phi}_M)}{\partial \pi} = \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda}{1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)} + \frac{n_1}{n}\frac{1}{1-\pi}\ . \tag{3.27}$$

Applying the Law of Large Numbers, Equation (3.24) converges to

$$\frac{\partial \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \pi} = \frac{-\gamma}{1-\pi^0} + \frac{\gamma}{1-\pi}\ . \tag{3.28}$$

At the true parameter values, it is easy to see that

$$\left.\frac{\partial \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \pi}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0\\ \pi=\pi^0\\ \lambda=\lambda^0}} = 0\ . \tag{3.29}$$

For the last parameter in Equation (3.11), we multiply the first derivative with respect to λ by $1/n$ and obtain

$$\frac{1}{n}\frac{\partial l_M(\boldsymbol{\phi}_M)}{\partial \lambda} = -\sum_{i=1}^{n}\frac{P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi}{n\Big[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\Big]}\ . \tag{3.30}$$

Again using the identities, it is straightforward to see that

$$\left.\frac{\partial \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \lambda}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0\\ \pi=\pi^0\\ \lambda=\lambda^0}} = 0 \tag{3.31}$$

at the true parameter values.

Thus, we have shown that the profile log-likelihood function converges in probability

to a continuous, vector-valued function and a root of the likelihood equations exists; that is,

$$\frac{1}{n}\frac{\partial l_M(\boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\phi}_M} \xrightarrow{p} \mathbf{0} \ . \tag{3.32}$$

3.4.3 Limiting Form of the Hessian Matrix

Before taking the advantage of Foutz' results and Lemma 3.3, we need to show one more condition that the convergence in probability of the Hessian matrix to its limiting form is uniform for $\boldsymbol{\phi}_M$ in an open neighborhood about $\boldsymbol{\phi}_M^0$. To ensure the parameter estimators considered here lie in a compact neighborhood, we have to consider a neighborhood $\boldsymbol{U} = A \times A_\pi \times A_\lambda$ of $\boldsymbol{\phi}_M^0$ of sufficiently small radius, where $\pi \in A_\pi = [\pi^0 - \epsilon, \pi^0 + \epsilon]$ and $\lambda \in A_\lambda = [\lambda^0 - \delta, \lambda^0 + \delta]$, for some ϵ and δ that $0 < \epsilon < \pi^0\gamma$ and $0 < \delta < \lambda^0\gamma$.

Using the Law of Large Numbers, $1/n$ times Equation (3.12) can be shown that

$$\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} \ , \tag{3.33}$$

where

$$\begin{aligned} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} &= \mathrm{E}\left[\frac{\partial^2 \ln \widetilde{h}(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right] \\ &= \mathrm{E}\left[\frac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right] - \mathrm{E}\left[\frac{\partial^2 \ln\left[1 + \lambda\Big(P_0(\boldsymbol{X}; \boldsymbol{\theta}) - \pi\Big)\right]}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right] \ , \end{aligned} \tag{3.34}$$

where

$$\widetilde{h}(\boldsymbol{y}, \boldsymbol{x}; \boldsymbol{\phi}_M) = \frac{f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})}{1 + \lambda\Big(P_0(\boldsymbol{x}; \boldsymbol{\theta}) - \pi\Big)} \tag{3.35}$$

is the limiting form of (3.8). Note that the convergence of the first term in (3.12) to the first term in (3.34) is uniform for all $\boldsymbol{\theta} \in A$, by assumption A6 and Lemma 3.2. A is the neighborhood about $\boldsymbol{\theta}^0$ defined in A6. By assumptions A3 and A4, the existence of the first two derivatives of $P_0(\boldsymbol{x}; \boldsymbol{\theta})$ for all $\boldsymbol{x} \in \mathbb{X}$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is guaranteed. As a result, the first two derivatives with respect to $\boldsymbol{\theta}$ are uniformly bounded on $\boldsymbol{\Theta} \times \mathbb{X}$ since the derivatives only involve $\boldsymbol{x}$ and $\boldsymbol{\theta}$ and $\boldsymbol{\Theta} \times \mathbb{X}$ is compact by assumption A2. It is obvious to see that $P_0(\boldsymbol{x}; \boldsymbol{\theta})$ is uniformly bounded as well. Therefore, by Lemma 3.2, the second term in the first equation of (3.12) converges uniformly to the second term in (3.34).

Note that, at $(\boldsymbol{\phi}_M^0)$,

$$\begin{aligned}
& \mathrm{E}\left[\frac{\dfrac{\partial^2 \ln \widetilde{h}(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}}{\widetilde{h}(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}_M^0)}\right] \\
= & \int_{\mathbb{X}} \int_{-\infty}^{a} \int_{-\infty}^{a} \frac{f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}^0)}{\pi^0} \frac{\dfrac{\partial^2 \ln \widetilde{h}(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}}{\widetilde{h}(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}_M^0)} \, d\boldsymbol{y} \, dG_{\boldsymbol{X}}(\boldsymbol{x}) \\
= & \int_{\mathbb{X}} \left[1 + \lambda^0 \Big(P_0(\boldsymbol{X}; \boldsymbol{\theta}^0) - \pi^0\Big)\right] \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \left(\frac{\dfrac{P_0(\boldsymbol{X}; \boldsymbol{\theta}^0)}{\pi^0}}{1 + \lambda^0 \Big(P_0(\boldsymbol{X}; \boldsymbol{\theta}^0) - \pi^0\Big)}\right) dG_{\boldsymbol{X}}(\boldsymbol{x}) \\
= & \int_{\mathbb{X}} \left(1 + \lambda^0 \Big[P_0(\boldsymbol{X}; \boldsymbol{\theta}^0) - \pi^0\Big]\right) \left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} 1\right) dG_{\boldsymbol{X}}(\boldsymbol{x}) \\
= & \, \mathbf{0} \, ,
\end{aligned} \tag{3.36}$$

where the identity in (3.19) is used. Since

$$\frac{\partial^2 \ln \widetilde{h}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = -\frac{\partial \ln \widetilde{h}}{\partial \boldsymbol{\theta}} \frac{\partial \ln \widetilde{h}}{\partial \boldsymbol{\theta}^T} + \frac{1}{\widetilde{h}} \frac{\partial^2 \widetilde{h}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ ,$$

at the true parameter values, we can see that

$$\begin{aligned}
\left. \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0 \\ \pi=\pi^0 \\ \lambda=\lambda^0}} &= \mathrm{E}\left[\frac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] - \mathrm{E}\left[\frac{\partial^2 \ln \left[1 + \lambda^0 \Big(P_0(\boldsymbol{X};\boldsymbol{\theta}^0) - \pi^0 \Big) \right]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \\
&= -\mathrm{E}\left[\frac{\partial \ln \widetilde{h}(\boldsymbol{Y},\boldsymbol{X};\boldsymbol{\phi}^0)}{\partial \boldsymbol{\theta}} \frac{\partial \ln \widetilde{h}(\boldsymbol{Y},\boldsymbol{X};\boldsymbol{\phi}^0)}{\partial \boldsymbol{\theta}^T} \right] \\
&= -\boldsymbol{Q} \ . \qquad (3.37)
\end{aligned}$$

To check that $\boldsymbol{Q}$ is positive definite, we can consider an arbitrary quadratic form, $\boldsymbol{t}^T \boldsymbol{Q} \boldsymbol{t}$ for any $\boldsymbol{t}_{p\times 1} \neq \boldsymbol{0}$ and then apply Assumption A7. For $\boldsymbol{y} \in \mathbb{Y}$ and $\boldsymbol{x} \in \mathbb{X}$, it is straightforward to see the following equations:

$$\begin{aligned}
& \boldsymbol{t}^T \boldsymbol{Q} \boldsymbol{t} = \boldsymbol{0} \\
\iff \quad & \boldsymbol{t}^T \frac{\partial \ln \widetilde{h}(\boldsymbol{Y},\boldsymbol{X};\boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\theta}} = 0 \\
\iff \quad & \sum_{j=1}^{p} t_j \left[\frac{\partial \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} - \frac{\lambda^0 \frac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}}{1 + \lambda^0 \Big(P_0(\boldsymbol{X};\boldsymbol{\theta}^0) - \pi^0 \Big)} \right] = 0 \\
\iff \quad & \sum_{j=1}^{p} t_j \lambda^0 \frac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} = 0 \ .
\end{aligned}$$

From Assumption A7, we know that the last equation above will hold for almost all $\boldsymbol{x} \in \mathbb{X}$ only if $\boldsymbol{t} = \boldsymbol{0}$. Thus, $\boldsymbol{Q}$ is positive definite.

Next, by the Law of Large Numbers, we note that $1/n$ times Equation (3.12)

$$\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\theta}\partial\pi} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\theta}\partial\pi} \ , \tag{3.38}$$

where

$$\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\theta}\partial\pi} = -\mathrm{E}\left[\frac{\lambda^2 \dfrac{\partial P_0(\boldsymbol{X};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}}{\left[1+\lambda\Big(P_0(\boldsymbol{X};\boldsymbol{\theta})-\pi\Big)\right]^2}\right] \ . \tag{3.39}$$

Using similar arguments made previously, it is easy to see that the convergence above is uniform since the whole term is uniformly bounded. At the true parameter values, we then have

$$\begin{aligned}\left.\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\theta}\partial\pi}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0\\ \pi=\pi^0\\ \lambda=\lambda^0}} &= -\mathrm{E}\left[\frac{(\lambda^0)^2 \dfrac{\partial P_0(\boldsymbol{X};\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}}}{\left[1+\lambda^0\Big(P_0(\boldsymbol{X};\boldsymbol{\theta}^0)-\pi^0\Big)\right]^2}\right] \\ &= -s \ . \end{aligned} \tag{3.40}$$

Moving on to the rest of the second derivatives, by the Law of Large Numbers and assumption A1, we can see from Equations (3.14) to (3.17) that

$$\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\theta}\partial\lambda} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\theta}\partial\lambda} \ , \tag{3.41}$$

$$\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial\pi^2} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\pi^2} \ , \tag{3.42}$$

$$\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial\pi\partial\lambda} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\pi\partial\lambda} \ , \tag{3.43}$$

and

$$\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial\pi\partial\lambda} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\pi\partial\lambda} \tag{3.44}$$

uniformly for $\boldsymbol{\phi}_M$ on $\boldsymbol{U}$, where

$$\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\theta}\partial\lambda} = -\mathrm{E}\left[\frac{\dfrac{\partial P_0(\boldsymbol{X};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}}{\left[1+\lambda\Big(P_0(\boldsymbol{X};\boldsymbol{\theta})-\pi\Big)\right]^2}\right] , \tag{3.45}$$

$$\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\pi^2} = \mathrm{E}\left[\frac{\lambda^2}{\left[1+\lambda\Big(P_0(\boldsymbol{X};\boldsymbol{\theta})-\pi\Big)\right]^2}\right] + \frac{\gamma}{(1-\pi)^2} , \tag{3.46}$$

$$\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\pi\partial\lambda} = \mathrm{E}\left[\left[1+\lambda\Big(P_0(\boldsymbol{X};\boldsymbol{\theta})-\pi\Big)\right]^2\right] . \tag{3.47}$$

and

$$\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\lambda^2} = \mathrm{E}\left[\frac{\Big(P_0(\boldsymbol{X};\boldsymbol{\theta})-\pi\Big)^2}{\left[1+\lambda\Big(P_0(\boldsymbol{X};\boldsymbol{\theta})-\pi\Big)\right]^2}\right] . \tag{3.48}$$

Then at the true parameter values, the results are as follows:

$$\left.\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\theta}\partial\lambda}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0\\ \pi=\pi^0\\ \lambda=\lambda^0}} = -\mathrm{E}\left[\frac{\dfrac{\partial P_0(\boldsymbol{X};\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}}}{\left[1+\lambda^0\Big(P_0(\boldsymbol{X};\boldsymbol{\theta}^0)-\pi^0\Big)\right]^2}\right] = -r , \tag{3.49}$$

$$\left.\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\pi^2}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0\\ \pi=\pi^0\\ \lambda=\lambda^0}} = \mathrm{E}\left[\frac{(\lambda^0)^2}{\left[1+\lambda^0\Big(P_0(\boldsymbol{X};\boldsymbol{\theta}^0)-\pi^0\Big)\right]^2}\right] + \frac{\gamma}{(1-\pi^0)^2} = -t , \tag{3.50}$$

$$\left.\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\pi\partial\lambda}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0 \\ \pi=\pi^0 \\ \lambda=\lambda^0}} = \mathrm{E}\left[\left[1+\lambda^0\Big(P_0(\boldsymbol{X};\boldsymbol{\theta}^0)-\pi^0\Big)\right]^2\right] = -w \ . \tag{3.51}$$

and

$$\left.\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\lambda^2}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0 \\ \pi=\pi^0 \\ \lambda=\lambda^0}} = \mathrm{E}\left[\frac{\Big(P_0(\boldsymbol{X};\boldsymbol{\theta}^0)-\pi^0\Big)^2}{\left[1+\lambda^0\Big(P_0(\boldsymbol{X};\boldsymbol{\theta}^0)-\pi^0\Big)\right]^2}\right] = -z \ . \tag{3.52}$$

Combining all the results, we obtain that

$$-\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M^0)}{\partial\boldsymbol{\phi}_M\partial\boldsymbol{\phi}_M^T} \xrightarrow{p} \boldsymbol{J} = \begin{pmatrix} \boldsymbol{Q} & s & r \\ s & t & w \\ r & w & z \end{pmatrix} \ . \tag{3.53}$$

We have shown that $\boldsymbol{Q}$ is positive definite and hence nonsingular. It is clear to see that the other terms of $\boldsymbol{J}$ are integers. As a result, we note that the matrix $\boldsymbol{J}$ is nonsingular and therefore invertible.

3.4.4 Proof of Consistency

Using the results obtained in the previous sections, we are now ready to prove Theorem 3.1. The idea of our proof follow Theorem 2 in Foutz' (1977) closely, which showed the existence of a consistent solution to the likelihood equations and its uniqueness, and then we apply Lemma 3.3, which modifies Foutz' conditions and still leads to the same results.

Proof of Theorem 3.1. We have shown in Section 3.4.2 that

$$\frac{1}{n}\frac{\partial l_M(\boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\phi}_M} \xrightarrow{p} \mathbf{0} \ . \tag{3.54}$$

And in Section 3.4.3, we have demonstrated that the convergence in probability of

$$\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\phi}_M \partial \boldsymbol{\phi}_M^T} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\phi}_M \partial \boldsymbol{\phi}_M} \tag{3.55}$$

is uniform for $\boldsymbol{\phi}_M$ in an open neighborhood for $\boldsymbol{\phi}_M^0$, and at the true parameter values,

$$-\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\phi}_M \partial \boldsymbol{\phi}_M^T} = \boldsymbol{J} \ , \tag{3.56}$$

which has been shown to be invertible. To make use of Lemma 3.3, let

$$\boldsymbol{f}_N(\boldsymbol{\phi}_M) = \frac{1}{n}\frac{\partial l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\phi}_M} \ ,$$

$$\boldsymbol{f}'_N(\boldsymbol{\phi}_M) = \frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\phi}_M \partial \boldsymbol{\phi}_M^T}$$

and

$$\boldsymbol{H}(\boldsymbol{\theta}) = \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial \boldsymbol{\phi}_M \partial \boldsymbol{\phi}_M^T} \ .$$

Clearly, the conditions in Lemma 3.3 are satisfied so that we can apply Lemma 3.3 and conclude that $\widehat{\boldsymbol{\phi}}_M = \boldsymbol{f}^{-1}(\mathbf{0})$ exists with probability going to one as $n \to \infty$ and $\widehat{\boldsymbol{\phi}}_M \xrightarrow{p} \boldsymbol{\phi}_M^0$. Furthermore, by the one-to-oneness of $\boldsymbol{f}_N$, any other sequence $\{\overline{\boldsymbol{\phi}}_M\}$ being roots to $\boldsymbol{f}_N(\boldsymbol{\phi}_M) = \mathbf{0}$ must lie outside of $\boldsymbol{U}$ with probability approaching to one as

$n \to \infty$, which demonstrates its uniqueness. □

3.5 Asymptotic Normality of the SEMLE

3.5.1 Proof of Asymptotic Normality

We prove the asymptotic normality result in four steps in this section. We first start from a Taylor series expansion. The limiting form of the Hessian matrix is calculated thereafter; then, the asymptotic distribution of the estimated score function is derived. In the end of the proof, Slutsky's Theorem will be applied to obtain the desired result.

Proof of Theorem 3.2.

Step 1: The Taylor Series Expansion

In the previous section, we have established the consistency result for our proposed estimator, $\widehat{\boldsymbol{\phi}}_M$, i.e. this estimator is a consistent solution to the profile score equations. We consider a Taylor series expansion of the estimated score function around the true parameter $\boldsymbol{\phi}_M^0$ evaluated at $\widehat{\boldsymbol{\phi}}_M$,

$$\frac{\partial l_M(\widehat{\boldsymbol{\phi}}_M)}{\partial \boldsymbol{\phi}_M} = \frac{\partial l_M(\boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\phi}_M} + \frac{\partial^2 l_M(\widetilde{\boldsymbol{\phi}}_M)}{\partial \boldsymbol{\phi}_M \partial \boldsymbol{\phi}_M^T}(\widehat{\boldsymbol{\phi}}_M - \boldsymbol{\phi}_M^0) \ , \tag{3.57}$$

where $\widetilde{\boldsymbol{\phi}}_M = \kappa \boldsymbol{\phi}_M^0 + (1-\kappa)\widehat{\boldsymbol{\phi}}_M$ for some $\kappa \in [0,1]$, as in Cosslett (1981b). The left-hand side of (3.57) is equal to zero since our estimator $\widehat{\boldsymbol{\phi}}_M$ has been shown to be a consistent solution to $\partial l_M(\boldsymbol{\phi}_M)/\partial \boldsymbol{\phi}_M = \mathbf{0}$. Rearranging (3.57) gives

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_M - \boldsymbol{\phi}_M^0) = \left[-\frac{1}{n} \frac{\partial^2 l_M(\widetilde{\boldsymbol{\phi}}_M)}{\partial \boldsymbol{\phi}_M \partial \boldsymbol{\phi}_M^T} \right]^{-1} \left[\frac{1}{\sqrt{n}} \frac{\partial l_M(\boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\phi}_M} \right] . \tag{3.58}$$

To prove the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\phi}}_M - \boldsymbol{\phi}_M^0)$, it is sufficient to show that the first bracket of (3.58), $-(1/n)\partial^2 l_M(\widetilde{\boldsymbol{\phi}}_M)/\partial\boldsymbol{\phi}_M\partial\boldsymbol{\phi}_M^T$ converges to an invertible matrix in probability and $(1/\sqrt{n})\partial l_M(\boldsymbol{\phi}_M^0)/\partial\boldsymbol{\phi}_M$ has an asymptotic normal distribution.

Step 2: The Limiting Form of Hessian Matrix

From Theorem 3.1, we have known that $\widehat{\boldsymbol{\phi}}_M \xrightarrow{p} \boldsymbol{\phi}_M^0$, which implies that $\widetilde{\boldsymbol{\phi}}_M \xrightarrow{p} \boldsymbol{\phi}_M^0$. And we have shown in Section 3.4.3 that

$$\frac{1}{n}\frac{\partial^2 l_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\phi}_M\partial\boldsymbol{\phi}_M^T} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M)}{\partial\boldsymbol{\phi}_M\partial\boldsymbol{\phi}_M^T}$$

uniformly for $\boldsymbol{\phi}_M \in \boldsymbol{U}$. According to Lemma 3.4, we can see that

$$-\frac{1}{n}\frac{\partial^2 l_M(\widetilde{\boldsymbol{\phi}}_M)}{\partial\boldsymbol{\phi}_M\partial\boldsymbol{\phi}_M^T} \xrightarrow{p} -\frac{\partial^2 \widetilde{l}_M(\boldsymbol{\phi}_M^0)}{\partial\boldsymbol{\phi}_M\partial\boldsymbol{\phi}_M^T} = \boldsymbol{J} \ , \tag{3.59}$$

where $\boldsymbol{J}$ is given by (3.53). Since $\boldsymbol{J}$ is shown to be positive definite, it follows that its inverse exists.

Step 3: Derivation of the Asymptotic Distribution of the Estimated Score Function Next, we consider the asymptotic distribution of $n^{-1/2}\partial l_M(\boldsymbol{\phi}_M^0)/\partial\boldsymbol{\phi}_M$, the second bracket in (3.58). Note that

$$\mathrm{E}\left[\frac{\partial l_M(\boldsymbol{Y},\boldsymbol{X};\boldsymbol{\phi}_M^0)}{\partial\boldsymbol{\phi}_M}\right] = \mathbf{0} \ . \tag{3.60}$$

Then, by the Central Limit Theorem, we know that

$$\frac{1}{\sqrt{n}}\frac{\partial l_M(\boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\phi}_M} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{V}) \ ,$$

where

$$\boldsymbol{V} = \text{Var}\left[\frac{\partial l_M(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}_M^0)}{\partial \boldsymbol{\phi}_M}\right] \ . \tag{3.61}$$

Step 4: Application of Slutsky's Theorem Finally, combining the results obtained in Equations (3.59) and (3.61) and then applying Slutsky's Theorem (Sen and Singer, 1993) to Equation (3.58), we have that

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_M - \boldsymbol{\phi}_M^0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\phi}_M^0)) \ ,$$

where

$$\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{V}\boldsymbol{J},$$

which is the asymptotic covariance matrix of $\widehat{\boldsymbol{\phi}}_M$.

□

3.5.2 A Consistent Estimator for the Asymptotic Variance Matrix

Proof of Theorem 3.3. It is noted that the observations from any one component of the *Multivariate-ODS* design are *i.i.d.*; thus, the sample covariance matrix over the

observed values is consistent for $\boldsymbol{\Sigma}(\boldsymbol{\phi}_M)$. Then, it is straightforward to see that

$$\widehat{\boldsymbol{V}}(\boldsymbol{\phi}_M) = \frac{1}{n} \widehat{\text{Var}}_{\{i\}} \left[\frac{\partial l_M(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\phi}_M)}{\partial \boldsymbol{\phi}_M} \right] \xrightarrow{p} \boldsymbol{V}(\boldsymbol{\phi}_M) \ .$$

By Assumption 3, the components of $\boldsymbol{V}(\boldsymbol{\phi}_M)$ are continuous in $\boldsymbol{\phi}_M$. We can then use the triangle inequality to obtain that

$$\|\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_M) - \boldsymbol{V}(\boldsymbol{\phi}_M^0)\| \leq \|\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_M) - \boldsymbol{V}(\widehat{\boldsymbol{\phi}}_M)\| + \|\boldsymbol{V}(\widehat{\boldsymbol{\phi}}_M) - \boldsymbol{V}(\boldsymbol{\phi}_M^0)\| \xrightarrow{p} 0$$

as n goes to ∞. Furthermore, in the proof of Theorem 3.2, we have shown that

$$\widehat{\boldsymbol{J}}(\widehat{\boldsymbol{\phi}}_M) = -\frac{1}{n} \frac{\partial^2 l_M(\widehat{\boldsymbol{\phi}}_M)}{\partial \boldsymbol{\phi}_M \partial \boldsymbol{\phi}_M^T} \xrightarrow{p} \boldsymbol{J}(\boldsymbol{\phi}_M^0) \ ,$$

which was defined in (3.53). It then follows that $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\phi}}_M)$ is a consistent estimator of the asymptotic covariance matrix.

□

CHAPTER 4

NUMERICAL RESULTS FOR THE MULTIVARIATE-ODS WITH A MAXIMUM SELECTION CRITERION

4.1 Introduction

In Chapter 3, we established the asymptotic theory for the SEMLE, $\widehat{\boldsymbol{\phi}}_M$, under the *Multivariate-ODS* with a maximum selection criterion and derived the theoretical asymptotic properties. In this chapter, we study the performance of $\widehat{\boldsymbol{\phi}}_M$ in small samples, investigated by means of simulation studies. Furthermore, we compare our proposed estimator to several competing estimators using the simulated data. In the last section, we will apply our proposed estimator to the Collaborative Perinatal Project (CPP) study as described in Section 1.2.1.

We will examine the small sample properties of the proposed estimator under the model of continuous outcomes with a bivariate normal distribution by conducting simulation experiments with various settings of sampling design specifications. For each experiment, we compute the parameter estimates and the estimated standard errors for the proposed estimator and other competing estimators, and the nominal 95% confidence intervals will be calculated based on their asymptotic normal distributions.

The primary objectives of the simulation studies are:

1. To determine if the proposed estimator is an unbiased estimator in small samples. This is addressed by comparing the means of the parameter estimates to the true parameters used to generate the data.

2. To determine if the variance estimator of the proposed estimator is a good estimate of the true variance in small samples. The "true" variance is defined as the variance of the estimator calculated over the simulated data sets within each simulation. To satisfy this goal, we compare the means of the variance estimator with the simulation sample variance.

3. To determine if the asymptotic normality distribution of the proposed estimator is a reasonable approximation in small samples. We satisfy this goal by studying the actual distribution in small samples and comparing the coverage of nominal 95% confidence intervals.

4. To compare the proposed estimator to other competing estimators with respect to small sample relative efficiency. For this goal, we consider asymptotic relative efficiency (ARE) of the proposed estimator relative to the other estimator, defined as the ratio of the variances, $VAR(\widehat{\boldsymbol{\phi}}_{other})/VAR(\widehat{\boldsymbol{\phi}}_{proposed})$.

There are several factors of the sampling design that may affect the performance of the proposed estimator. These factors include the *Multivariate-ODS* sample size n, the sampling fraction $\gamma = n_1/n$ which is the allocation of the supplemental sample to other

components that make up the *Multivariate-ODS*, the location of the cutpoint a for partitioning the space of $\boldsymbol{Y}_{\max}$, and the correlation coefficient of the outcome responses ρ. We will investigate the performance of the proposed estimator under various configurations of these sampling specifications.

4.2 Data Generation

4.2.1 The Simulation Model

We consider the following bivariate normal model to generate the simulated data:

$$\boldsymbol{Y}|\boldsymbol{X} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} {\sigma_1}^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & {\sigma_2}^2 \end{pmatrix}\right],$$

where $\boldsymbol{Y} = \left(Y_1, Y_2\right)^T$, $\boldsymbol{X} = \left(X_1, X_2\right)^T$, $\mu_1 = \alpha_1 + \beta_1 X_1$ and $\mu_2 = \alpha_2 + \beta_2 X_2$; i.e., the conditional distributions of Y_1 given X_1 and Y_2 given X_2 are normally distributed with means $\alpha_1 + \beta_1 X_1$ and $\alpha_2 + \beta_2 X_1$, variances σ_1^2 and σ_2^2, respectively, and the correlation coefficient ρ. Our goal is to estimate the parameter vector $\boldsymbol{\theta}_P = (\alpha_1, \beta_1, \alpha_2, \beta_2, \sigma_1, \sigma_2, \rho)^T$. In particular, we will investigate the behavior of β_1 and β_2 by fixing $\alpha_1 = 0.5$, $\alpha_2 = -0.8$, $\sigma_1^2 = \sigma_2^2 = 1$ or $\sigma_1^2 = \sigma_2^2 = 1.5$, and allowing $\boldsymbol{\beta}$ to take different values for β_1 and β_2. Then the same models are applied to $\rho = 0.5$ and $\rho = 0.85$ to see how the magnitude of association between outcome variables affects the parameter estimates.

4.2.2 Sampling Design Specifications

The *Multivariate-ODS* sample sizes for investigation were $n = 200$ and $n = 800$. The *Multivariate-ODS* design for this study included an overall SRS and a supplemental sample from individuals whose maximum values of the outcomes were in the tail of the distribution of $\boldsymbol{Y}_{\max}$. For simulations, we chose the cutpoint to partition the space of $\boldsymbol{Y}_{\max}$, a, of the 80th or 90th percentile from the distribution of $\boldsymbol{Y}_{\max}$ under the study models. The supplemental sampling fraction, $\gamma = n_1/n$, was either 20% or 50%.

4.2.3 Algorithm of Data Generation

Since Y_1 and Y_2 are bivariate normally distributed, we then have

$$Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$$

and

$$Y_2|Y_1, \boldsymbol{X} \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}\left(Y_1 - \mu_1\right),\ \sigma_2^2\left(1-\rho^2\right)\right)\ ;$$

that is, we can write the linear regression models for Y_1 and Y_2 of the forms

$$Y_1 = \alpha_1 + \beta_1 X_1 + \sigma_1\epsilon_1 \tag{4.1}$$

and

$$Y_2 = \alpha_2 + \beta_2 X_2 + \rho(\sigma_2/\sigma_1)(Y_1 - \alpha_1 - \beta_1 X_1) + \sigma_2\sqrt{1-\rho^2}\epsilon_2\ , \tag{4.2}$$

where $\epsilon_1 \sim N(0,1)$ and $\epsilon_2 \sim N(0,1)$ are independent.

For all the investigations, we generated 1,000 realizations of data in accordance with the models specified above. The process was as follows:

1) We generated N independent error terms ϵ_1 and ϵ_2 for Y_1 and Y_2 respectively, from the standard normal distribution, where N represents the size of the underlying population. Note that N was set to be 20,000 in the simulation.

2) Next, independently from these error terms, we generated two independent covariate vectors X_1 and X_2, each of size N, from the standard normal distribution.

3) Then we obtained the response vector, Y_1, by plugging the generated errors ϵ_1 and the covariate vector X_1 into the model in (4.1) with the specified parameter values. With Y_1, X_1, X_2, ϵ_2 and the parameter values, Y_2 were then generated according to (4.2).

4) Selection of the SRS and the supplemental sample proceeded as follows: an SRS of size n_0 was randomly selected from the underlying population and a supplemental sample of size n_1 was drawn from the remaining realizations conditional on $\{\max(Y_1, Y_2) > a\}$ with the specified cutpoint a.

4.2.4 Competing Estimators

We compare our proposed estimator SEMLE, $\widehat{\boldsymbol{\theta}}_P$, to other competitive estimators under each setting in our simulation study: (i) the maximum likelihood estimator by maximizing the likelihood using only the SRS portion of the *Multivariate-ODS* data ($\widehat{\boldsymbol{\theta}}_R$), (ii) the maximum likelihood estimator by maximizing the conditional likelihood

based on the complete *Multivariate-ODS* data ($\widehat{\boldsymbol{\theta}}_C$), and (iii) the maximum likelihood estimator obtained from a random sample of the same size as the *Multivariate-ODS* sample ($\widehat{\boldsymbol{\theta}}_S$). Comparing $\widehat{\boldsymbol{\theta}}_P$ with $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$ will give us an insight of the impact on ignoring the information from the supplemental sample. The comparison between $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_S$ will demonstrate the efficiency gain of the *Multivariate-ODS* design over the simple random sample of the same size.

4.3 Summary of Results

The simulation results are presented in Tables 4.1 through 4.20. The results in the tables are presented for three different combinations of $\boldsymbol{\beta}$, the correlation coefficient ρ, various cutpoints a, the sampling fractions γ, and sample sizes n, with three methods. Within each table, the sampling specifications and the covariate distribution are fixed. Tables 4.1 - 4.16 include the small sample properties of the proposed estimator $\widehat{\boldsymbol{\theta}}_P$ and the competing estimators, $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$. Tables 4.17 - 4.20 present the efficiencies of $\widehat{\boldsymbol{\theta}}_S$ versus $\widehat{\boldsymbol{\theta}}_P$ based on the models for Tables 4.1 - 4.16.

4.3.1 The Unbiasedness, the Normality and the Variance Estimator

Tables 4.1 through 4.4 contain simulation results for the cases in which $\beta_1 = \beta_2 = 0$: $n = 200$ in Tables 4.1 and 4.2 with the correlation coefficients of $\rho = 0.5$ and $\rho = 0.85$, respectively; the same models were considered in Tables 4.3 and 4.4 but with $n = 800$. We make the following observations concerning the results presented in Tables 4.1 - 4.4.

1. The proposed method $\widehat{\boldsymbol{\theta}}_P$ along with $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$ produced unbiased estimates compared

with the "true" parameter values under four settings. As the sample size n increased, the bias was even hardly observed.

2. The proposed method $\widehat{\boldsymbol{\theta}}_P$ produced the smallest standard errors for estimating the model parameters whereas $\widehat{\boldsymbol{\theta}}_R$ always provided the least efficient estimators. The standard errors were smaller as the sample size n increased.

3. The proposed estimator $\widehat{\boldsymbol{\theta}}_P$ provided a very good estimate of the true variability; for $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$, the means of the standard error estimates were close to the simulation standard errors as well.

4. The confidence intervals based on the proposed estimator $\widehat{\boldsymbol{\theta}}_P$ provided good coverage close to the nominal 95% level. The same findings were seen for both $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$.

5. In Table 4.1, for the same sampling fraction, the standard errors of $\widehat{\boldsymbol{\theta}}_P$ decreased as the percentile of the cutpoint a increased, indicating that our proposed method was more efficient and favored when the supplemental sample included more extreme observations. Similar results were obtained in Tables 4.2 - 4.4.

6. Above observations were true for both $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

4.3.2 Additional Results for the Unbiasedness, the Normality and the Variance Estimator

Tables 4.5 through 4.8 presented the results for $\beta_1 = 0$ and $\beta_2 = 0.5$ with the same sampling specifications as those in Tables 4.1 - 4.4 respectively. We observed similar

tendencies exhibited in Tables 4.1 - 4.4. The proposed estimator $\widehat{\boldsymbol{\theta}}_P$ continued to outperform the competing estimators. We note that the variance estimator for $\widehat{\boldsymbol{\theta}}_P$ appeared to decrease as β_2 increased, which for example, could be seen from Tables 4.1 and 4.5.

For Tables 4.9 through 4.12, we changed $\boldsymbol{\beta}$ to be non-zero that $\beta_1 = -0.5$ and $\beta_2 = \ln(2)$ and kept the relative sampling specifications the same as before. The results observed in these tables were comparable to those in Tables 4.1 - 4.4. The proposed estimator $\widehat{\boldsymbol{\theta}}_P$ provided consistency and good variance estimates.

Tables 4.13 through 4.16 presented the results using the same models as Table 4.9 - 4.12 except that σ_1 and σ_2 increased to $\sigma_1 = \sigma_2 = 1.5$. We observed similar results as those in Tables 4.9 - 4.12. As the variances increased, the standard errors were larger, which was expected.

4.3.3 The Performance of $\widehat{ARE}$ ($= Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$)

We further investigated the amount of information gained by the use of the *Multivariate-ODS* design over a simple random sample of the same size, and the results of the relative efficiencies (ratios of variances, $Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$) were summarized in Tables 4.17 through 4.20 with different model settings. Throughout the four tables, $\widehat{AREs}$ were greater than one, except for some cases in Table 4.17 which were indeed closer to one. We make the following observations concerning the results in Tables 4.17 through 4.20.

1. Except for some cases where $\beta_1 = 0$ and $\beta_2 = 0$, the estimates of $\boldsymbol{\beta}$ from the proposed method $\widehat{\boldsymbol{\theta}}_P$ were more efficient than $\widehat{\boldsymbol{\theta}}_S$, indicating that the supplemental sample contained substantial information and the proposed method led to more efficiency

gains; $\widehat{\boldsymbol{\theta}}_P$ was more efficient than $\widehat{\boldsymbol{\theta}}_S$ with gains as large as 51% for estimating β_1 and 56% for β_2, which can be found in Table 4.20.

2. With the correlation coefficient and the sampling fraction fixed, the efficiency gains of $\widehat{\boldsymbol{\theta}}_P$ over $\widehat{\boldsymbol{\theta}}_S$ increased as the cutpoint was located further in the tail of the distribution.

3. With the cutpoint and the sampling fraction fixed, there was an increase in the relative efficiencies as the data were more correlated for most cases.

4. Observing the effect of the sample size n, with a higher correlation coefficient, the efficiency gained by using $\widehat{\boldsymbol{\theta}}_P$ over the $\widehat{\boldsymbol{\theta}}_S$ tended to increase as the samples size increased.

5. Comparing the results in Tables 4.19 and 4.20, we note that there was an increase overall as the variances of β_1 and β_2 increased.

From above results, we see that the observed efficiency gains for using $\widehat{\boldsymbol{\theta}}_P$ were noticeably larger than $\widehat{\boldsymbol{\theta}}_S$.

4.3.4 The Effect of Changing Supplemental Sampling Fractions on $\widehat{ARE}$

To investigate the effect of changing the supplemental sampling fractions on the improvement of the *Multivariate-ODS* design over other simple random sample designs, we conducted several simulation experiments using the same simulation models used in Tables 4.9 and 4.11 but with the cutpoint $a = 80\%$. Figures 4.1 and 4.2 presented the relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ over $\widehat{\boldsymbol{\theta}}_R$. Clearly, the efficiency gains of the *Multivariate-ODS* design over the simple random sample design increased with the supplemental sampling

fractions, agreed by both sample size considerations, and $\widehat{\boldsymbol{\theta}}_P$ was consistently more efficient than $\widehat{\boldsymbol{\theta}}_R$ regardless of the sampling fractions. Although the efficiency gains increased as the supplemental sample size increased, it was not practical in reality since it may not be easy to have enough individuals in the extreme tails. We suggested the possible remedy for an appropriate proportion of the supplemental sample to be in the region from 0.3 to 0.6. Figures 4.3 through 4.6 illustrated the standard errors of $\widehat{\boldsymbol{\theta}}_P$ and the relative efficiencies of the *Multivariate-ODS* design to a simple random sample of the same sample size across various supplemental sampling fractions γ. The increase in the relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ was not monotone over the fractions although $\widehat{\boldsymbol{\theta}}_P$ was substantially more efficient than $\widehat{\boldsymbol{\theta}}_S$ in most sampling fractions. We observed the most efficiency gain at $\gamma = 0.3$ for both β_1 and β_2 among all the figures. As γ was more than 60%, there was a drop in the relative efficiency. These results suggested that a great efficiency gain can be achieved when γ was between 0.3 and 0.6.

4.3.5 Conclusions

In this section, we demonstrated the asymptotic properties of our proposed method derived in Chapter 3 using the simulation studies and showed that the small-sample properties approximated well for samples with even a relative small sample size. The simulation results showed that our proposed method produced unbiased estimators, the variance estimators were good estimates for the true variances, and the proposed estimator was asymptotically normally distributed.

In terms of small sample relative efficiency studies, the proposed method provided

a more efficient parameter estimate than it obtained using a simple random sample of the same sample size. More efficiency gains were observed in the samples with a higher correlation. We suggested to achieve the greatest gain in efficiency, the supplemental sampling fraction was around 30%. We also illustrated that the efficiency gain of the proposed method over the estimator obtained only using the SRS was more substantial as the portion of the supplemental sample increased and the cutpoint moved further out in the tail of the $\boldsymbol{Y}_{\max}$ distribution. For practice, the suggested region for considering the proportion of the supplemental sample in the *Multivariate-ODS* was between 0.3 and 0.6.

In the next section, we will demonstrate the utility of our proposed method by applying it to the real data.

Table 4.1: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.357$ (80^{th} percentile) and 1.791 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.002	0.079	0.080	0.952	−0.003	0.082	0.080	0.946
		$\boldsymbol{\theta}_C$	0.002	0.077	0.077	0.953	−0.001	0.076	0.073	0.941
		$\boldsymbol{\theta}_P$	0.002	0.064	0.066	0.960	−0.001	0.070	0.069	0.948
	50%	$\boldsymbol{\theta}_R$	0.004	0.105	0.101	0.948	0.001	0.104	0.100	0.939
		$\boldsymbol{\theta}_C$	0.005	0.091	0.090	0.954	0.001	0.077	0.076	0.955
		$\boldsymbol{\theta}_P$	0.005	0.068	0.068	0.941	0.001	0.072	0.069	0.945
90%	20%	$\boldsymbol{\theta}_R$	−0.004	0.079	0.080	0.955	0.003	0.081	0.080	0.942
		$\boldsymbol{\theta}_C$	−0.005	0.077	0.078	0.957	0.001	0.076	0.073	0.930
		$\boldsymbol{\theta}_P$	−0.002	0.060	0.062	0.953	0.002	0.070	0.068	0.931
	50%	$\boldsymbol{\theta}_R$	0.001	0.103	0.101	0.938	−0.001	0.101	0.101	0.954
		$\boldsymbol{\theta}_C$	−0.001	0.096	0.092	0.943	−0.001	0.075	0.077	0.954
		$\boldsymbol{\theta}_P$	0.001	0.064	0.061	0.948	0.000	0.065	0.067	0.963

Table 4.2: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.340$ (80^{th} percentile) and 1.784 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.001	0.076	0.079	0.956	0.000	0.078	0.079	0.954
		$\boldsymbol{\theta}_C$	−0.001	0.075	0.077	0.953	−0.001	0.075	0.076	0.955
		$\boldsymbol{\theta}_P$	−0.001	0.064	0.066	0.961	−0.001	0.066	0.067	0.962
	50%	$\boldsymbol{\theta}_R$	0.002	0.102	0.102	0.949	0.003	0.102	0.102	0.953
		$\boldsymbol{\theta}_C$	0.001	0.091	0.092	0.950	0.003	0.088	0.086	0.944
		$\boldsymbol{\theta}_P$	0.001	0.066	0.068	0.951	0.002	0.069	0.069	0.945
90%	20%	$\boldsymbol{\theta}_R$	0.001	0.079	0.080	0.953	0.002	0.080	0.080	0.943
		$\boldsymbol{\theta}_C$	0.001	0.077	0.078	0.951	0.002	0.077	0.076	0.952
		$\boldsymbol{\theta}_P$	0.003	0.061	0.062	0.951	0.004	0.065	0.065	0.943
	50%	$\boldsymbol{\theta}_R$	0.001	0.102	0.100	0.942	0.001	0.099	0.101	0.959
		$\boldsymbol{\theta}_C$	0.001	0.095	0.093	0.942	0.001	0.089	0.088	0.946
		$\boldsymbol{\theta}_P$	0.001	0.062	0.061	0.947	0.001	0.065	0.064	0.944

Table 4.3: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.357$ (80^{th} percentile) and 1.791 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.002	0.040	0.040	0.950	0.002	0.041	0.040	0.939
		$\boldsymbol{\theta}_C$	0.002	0.039	0.038	0.948	0.002	0.037	0.036	0.940
		$\boldsymbol{\theta}_P$	0.002	0.034	0.033	0.949	0.002	0.035	0.034	0.950
	50%	$\boldsymbol{\theta}_R$	−0.001	0.050	0.050	0.953	−0.002	0.051	0.050	0.945
		$\boldsymbol{\theta}_C$	0.000	0.043	0.045	0.959	−0.001	0.039	0.038	0.942
		$\boldsymbol{\theta}_P$	0.001	0.034	0.034	0.947	0.000	0.035	0.034	0.940
90%	20%	$\boldsymbol{\theta}_R$	0.000	0.038	0.040	0.953	0.001	0.039	0.040	0.947
		$\boldsymbol{\theta}_C$	0.000	0.038	0.039	0.948	0.000	0.036	0.036	0.953
		$\boldsymbol{\theta}_P$	0.001	0.031	0.031	0.955	0.000	0.034	0.034	0.956
	50%	$\boldsymbol{\theta}_R$	0.000	0.049	0.050	0.962	0.000	0.049	0.050	0.953
		$\boldsymbol{\theta}_C$	0.000	0.045	0.046	0.956	−0.001	0.037	0.038	0.954
		$\boldsymbol{\theta}_P$	−0.001	0.030	0.030	0.948	−0.001	0.033	0.033	0.954

Table 4.4: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.340$ (80^{th} percentile) and 1.784 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.001	0.039	0.040	0.951	0.001	0.038	0.040	0.967
		$\boldsymbol{\theta}_C$	0.001	0.038	0.039	0.951	0.001	0.037	0.038	0.963
		$\boldsymbol{\theta}_P$	0.001	0.032	0.033	0.957	0.000	0.033	0.034	0.958
	50%	$\boldsymbol{\theta}_R$	0.001	0.050	0.050	0.952	0.002	0.049	0.050	0.952
		$\boldsymbol{\theta}_C$	0.001	0.044	0.045	0.951	0.001	0.042	0.043	0.950
		$\boldsymbol{\theta}_P$	0.000	0.033	0.034	0.958	0.000	0.033	0.034	0.955
90%	20%	$\boldsymbol{\theta}_R$	0.001	0.040	0.040	0.946	0.001	0.040	0.040	0.952
		$\boldsymbol{\theta}_C$	0.001	0.039	0.039	0.943	0.001	0.038	0.038	0.946
		$\boldsymbol{\theta}_P$	0.001	0.033	0.031	0.934	0.001	0.033	0.032	0.950
	50%	$\boldsymbol{\theta}_R$	−0.001	0.051	0.050	0.953	−0.002	0.049	0.050	0.952
		$\boldsymbol{\theta}_C$	−0.001	0.047	0.046	0.950	−0.001	0.043	0.044	0.962
		$\boldsymbol{\theta}_P$	0.001	0.031	0.030	0.952	0.001	0.032	0.032	0.956

Table 4.5: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.379$ (80^{th} percentile) and 1.814 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.001	0.079	0.079	0.948	0.504	0.079	0.080	0.947
		$\boldsymbol{\theta}_C$	0.001	0.076	0.076	0.951	0.504	0.070	0.073	0.962
		$\boldsymbol{\theta}_P$	0.002	0.067	0.066	0.948	0.504	0.066	0.068	0.966
	50%	$\boldsymbol{\theta}_R$	0.002	0.099	0.102	0.958	0.502	0.099	0.102	0.961
		$\boldsymbol{\theta}_C$	0.002	0.082	0.087	0.968	0.501	0.071	0.077	0.972
		$\boldsymbol{\theta}_P$	−0.001	0.065	0.067	0.963	0.499	0.065	0.068	0.959
90%	20%	$\boldsymbol{\theta}_R$	−0.007	0.078	0.080	0.953	0.497	0.079	0.080	0.954
		$\boldsymbol{\theta}_C$	−0.005	0.074	0.076	0.952	0.497	0.073	0.072	0.953
		$\boldsymbol{\theta}_P$	−0.004	0.060	0.062	0.963	0.498	0.068	0.067	0.949
	50%	$\boldsymbol{\theta}_R$	0.002	0.102	0.101	0.951	0.503	0.099	0.101	0.963
		$\boldsymbol{\theta}_C$	0.005	0.089	0.088	0.944	0.504	0.077	0.077	0.961
		$\boldsymbol{\theta}_P$	0.001	0.062	0.061	0.938	0.502	0.067	0.065	0.947

Table 4.6: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.346$ (80^{th} percentile) and 1.786 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.002	0.076	0.080	0.968	0.496	0.076	0.079	0.956
		$\boldsymbol{\theta}_C$	−0.002	0.073	0.077	0.966	0.496	0.073	0.076	0.953
		$\boldsymbol{\theta}_P$	0.000	0.064	0.066	0.964	0.498	0.065	0.067	0.963
	50%	$\boldsymbol{\theta}_R$	−0.001	0.102	0.101	0.962	0.501	0.099	0.101	0.958
		$\boldsymbol{\theta}_C$	−0.001	0.088	0.090	0.959	0.499	0.084	0.086	0.959
		$\boldsymbol{\theta}_P$	−0.001	0.067	0.068	0.958	0.498	0.067	0.069	0.954
90%	20%	$\boldsymbol{\theta}_R$	0.003	0.078	0.080	0.961	0.502	0.078	0.080	0.951
		$\boldsymbol{\theta}_C$	0.003	0.076	0.078	0.959	0.502	0.075	0.076	0.950
		$\boldsymbol{\theta}_P$	0.001	0.062	0.062	0.957	0.501	0.063	0.064	0.951
	50%	$\boldsymbol{\theta}_R$	0.003	0.102	0.101	0.947	0.505	0.103	0.101	0.949
		$\boldsymbol{\theta}_C$	0.003	0.093	0.092	0.947	0.503	0.089	0.088	0.947
		$\boldsymbol{\theta}_P$	0.001	0.062	0.061	0.955	0.501	0.065	0.063	0.943

Table 4.7: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.379$ (80^{th} percentile) and 1.814 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.001	0.040	0.040	0.950	0.498	0.040	0.040	0.946
		$\boldsymbol{\theta}_C$	−0.001	0.039	0.038	0.950	0.499	0.035	0.036	0.951
		$\boldsymbol{\theta}_P$	−0.001	0.033	0.033	0.948	0.499	0.034	0.034	0.953
	50%	$\boldsymbol{\theta}_R$	0.002	0.049	0.050	0.946	0.502	0.047	0.050	0.962
		$\boldsymbol{\theta}_C$	0.001	0.042	0.043	0.957	0.502	0.037	0.038	0.953
		$\boldsymbol{\theta}_P$	0.001	0.034	0.033	0.948	0.502	0.034	0.034	0.952
90%	20%	$\boldsymbol{\theta}_R$	0.000	0.039	0.040	0.956	0.501	0.039	0.040	0.953
		$\boldsymbol{\theta}_C$	0.000	0.037	0.038	0.961	0.500	0.035	0.036	0.957
		$\boldsymbol{\theta}_P$	0.000	0.030	0.031	0.952	0.501	0.032	0.033	0.958
	50%	$\boldsymbol{\theta}_R$	−0.003	0.051	0.050	0.953	0.501	0.051	0.050	0.955
		$\boldsymbol{\theta}_C$	−0.001	0.044	0.044	0.952	0.500	0.039	0.038	0.937
		$\boldsymbol{\theta}_P$	0.000	0.030	0.030	0.958	0.501	0.033	0.032	0.945

Table 4.8: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.346$ (80^{th} percentile) and 1.786 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.000	0.039	0.040	0.949	0.500	0.039	0.040	0.955
		$\boldsymbol{\theta}_C$	0.000	0.038	0.038	0.950	0.500	0.038	0.038	0.952
		$\boldsymbol{\theta}_P$	0.000	0.032	0.033	0.959	0.500	0.032	0.033	0.959
	50%	$\boldsymbol{\theta}_R$	−0.001	0.047	0.050	0.959	0.500	0.049	0.050	0.952
		$\boldsymbol{\theta}_C$	0.000	0.043	0.045	0.963	0.500	0.042	0.043	0.948
		$\boldsymbol{\theta}_P$	0.000	0.033	0.034	0.948	0.500	0.034	0.034	0.948
90%	20%	$\boldsymbol{\theta}_R$	0.001	0.039	0.040	0.955	0.502	0.039	0.040	0.956
		$\boldsymbol{\theta}_C$	0.001	0.038	0.039	0.952	0.502	0.037	0.038	0.962
		$\boldsymbol{\theta}_P$	0.000	0.031	0.031	0.951	0.501	0.031	0.032	0.954
	50%	$\boldsymbol{\theta}_R$	0.001	0.051	0.050	0.948	0.499	0.051	0.050	0.945
		$\boldsymbol{\theta}_C$	0.001	0.046	0.046	0.946	0.499	0.044	0.044	0.953
		$\boldsymbol{\theta}_P$	−0.001	0.031	0.031	0.944	0.498	0.031	0.032	0.951

Table 4.9: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.529$ (80^{th} percentile) and 1.991 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.499	0.080	0.080	0.943	0.693	0.079	0.080	0.952
		$\boldsymbol{\theta}_C$	−0.499	0.075	0.074	0.949	0.694	0.072	0.071	0.943
		$\boldsymbol{\theta}_P$	−0.500	0.068	0.067	0.945	0.694	0.066	0.066	0.946
	50%	$\boldsymbol{\theta}_R$	−0.503	0.105	0.101	0.946	0.692	0.102	0.101	0.945
		$\boldsymbol{\theta}_C$	−0.498	0.082	0.079	0.942	0.693	0.073	0.074	0.947
		$\boldsymbol{\theta}_P$	−0.500	0.071	0.067	0.953	0.692	0.063	0.064	0.953
90%	20%	$\boldsymbol{\theta}_R$	−0.501	0.080	0.080	0.944	0.693	0.081	0.080	0.954
		$\boldsymbol{\theta}_C$	−0.499	0.074	0.074	0.952	0.694	0.072	0.071	0.938
		$\boldsymbol{\theta}_P$	−0.502	0.063	0.065	0.952	0.691	0.064	0.064	0.953
	50%	$\boldsymbol{\theta}_R$	−0.494	0.103	0.101	0.950	0.695	0.101	0.101	0.959
		$\boldsymbol{\theta}_C$	−0.494	0.081	0.079	0.944	0.697	0.076	0.073	0.950
		$\boldsymbol{\theta}_P$	−0.499	0.065	0.064	0.945	0.694	0.061	0.060	0.944

Table 4.10: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.506$ (80^{th} percentile) and 1.979 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.501	0.081	0.080	0.945	0.692	0.080	0.080	0.950
		$\boldsymbol{\theta}_C$	−0.499	0.075	0.075	0.945	0.693	0.073	0.074	0.955
		$\boldsymbol{\theta}_P$	−0.500	0.067	0.067	0.953	0.692	0.066	0.066	0.957
	50%	$\boldsymbol{\theta}_R$	−0.493	0.105	0.101	0.943	0.699	0.104	0.101	0.941
		$\boldsymbol{\theta}_C$	−0.495	0.084	0.082	0.949	0.699	0.082	0.081	0.951
		$\boldsymbol{\theta}_P$	−0.499	0.068	0.067	0.953	0.694	0.066	0.066	0.958
90%	20%	$\boldsymbol{\theta}_R$	−0.500	0.081	0.080	0.952	0.694	0.079	0.080	0.959
		$\boldsymbol{\theta}_C$	−0.501	0.077	0.075	0.944	0.694	0.075	0.074	0.950
		$\boldsymbol{\theta}_P$	−0.504	0.066	0.064	0.949	0.691	0.066	0.063	0.952
	50%	$\boldsymbol{\theta}_R$	−0.502	0.101	0.101	0.945	0.691	0.103	0.101	0.941
		$\boldsymbol{\theta}_C$	−0.500	0.084	0.083	0.942	0.693	0.082	0.081	0.936
		$\boldsymbol{\theta}_P$	−0.500	0.065	0.063	0.938	0.693	0.063	0.061	0.938

Table 4.11: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.991$ (80^{th} percentile) and 1.529 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.501	0.039	0.040	0.945	0.692	0.040	0.040	0.948
		$\boldsymbol{\theta}_C$	−0.502	0.036	0.037	0.950	0.693	0.035	0.036	0.952
		$\boldsymbol{\theta}_P$	−0.502	0.033	0.035	0.958	0.693	0.033	0.033	0.954
	50%	$\boldsymbol{\theta}_R$	−0.504	0.051	0.050	0.939	0.688	0.050	0.050	0.945
		$\boldsymbol{\theta}_C$	−0.504	0.038	0.039	0.953	0.690	0.037	0.037	0.948
		$\boldsymbol{\theta}_P$	−0.502	0.033	0.034	0.959	0.691	0.032	0.032	0.958
90%	20%	$\boldsymbol{\theta}_R$	−0.503	0.039	0.040	0.959	0.693	0.041	0.040	0.932
		$\boldsymbol{\theta}_C$	−0.503	0.036	0.037	0.961	0.693	0.036	0.035	0.947
		$\boldsymbol{\theta}_P$	−0.502	0.031	0.032	0.955	0.694	0.032	0.032	0.949
	50%	$\boldsymbol{\theta}_R$	−0.499	0.048	0.050	0.958	0.694	0.049	0.050	0.950
		$\boldsymbol{\theta}_C$	−0.500	0.039	0.039	0.960	0.693	0.036	0.036	0.951
		$\boldsymbol{\theta}_P$	−0.499	0.031	0.030	0.957	0.693	0.032	0.030	0.949

Table 4.12: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.506$ (80^{th} percentile) and 1.979 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.500	0.039	0.040	0.954	0.694	0.038	0.040	0.953
		$\boldsymbol{\theta}_C$	−0.500	0.036	0.037	0.957	0.694	0.035	0.037	0.958
		$\boldsymbol{\theta}_P$	−0.500	0.033	0.033	0.954	0.693	0.032	0.033	0.956
	50%	$\boldsymbol{\theta}_R$	−0.499	0.049	0.050	0.953	0.693	0.049	0.050	0.954
		$\boldsymbol{\theta}_C$	−0.500	0.040	0.041	0.960	0.693	0.039	0.040	0.959
		$\boldsymbol{\theta}_P$	−0.500	0.033	0.034	0.955	0.692	0.032	0.033	0.961
90%	20%	$\boldsymbol{\theta}_R$	−0.498	0.041	0.040	0.945	0.695	0.039	0.040	0.945
		$\boldsymbol{\theta}_C$	−0.498	0.038	0.037	0.950	0.695	0.037	0.037	0.939
		$\boldsymbol{\theta}_P$	−0.499	0.032	0.032	0.953	0.694	0.031	0.031	0.950
	50%	$\boldsymbol{\theta}_R$	−0.497	0.052	0.050	0.944	0.695	0.051	0.050	0.946
		$\boldsymbol{\theta}_C$	−0.498	0.042	0.041	0.950	0.695	0.040	0.040	0.955
		$\boldsymbol{\theta}_P$	−0.499	0.031	0.031	0.947	0.695	0.030	0.030	0.956

Table 4.13: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.5$, $a = 1.999$ (80^{th} percentile) and 2.649 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.497	0.120	0.119	0.942	0.694	0.117	0.119	0.956
		$\boldsymbol{\theta}_C$	−0.496	0.115	0.112	0.943	0.692	0.104	0.108	0.959
		$\boldsymbol{\theta}_P$	−0.499	0.105	0.100	0.935	0.691	0.097	0.100	0.960
	50%	$\boldsymbol{\theta}_R$	−0.494	0.151	0.152	0.951	0.697	0.148	0.152	0.958
		$\boldsymbol{\theta}_C$	−0.497	0.117	0.122	0.956	0.700	0.110	0.114	0.950
		$\boldsymbol{\theta}_P$	−0.500	0.097	0.101	0.968	0.699	0.095	0.099	0.952
90%	20%	$\boldsymbol{\theta}_R$	−0.499	0.120	0.119	0.952	0.694	0.115	0.119	0.963
		$\boldsymbol{\theta}_C$	−0.499	0.111	0.112	0.953	0.694	0.104	0.108	0.953
		$\boldsymbol{\theta}_P$	−0.501	0.097	0.095	0.951	0.693	0.094	0.096	0.950
	50%	$\boldsymbol{\theta}_R$	−0.504	0.157	0.152	0.940	0.680	0.153	0.152	0.953
		$\boldsymbol{\theta}_C$	−0.502	0.124	0.123	0.950	0.688	0.110	0.113	0.949
		$\boldsymbol{\theta}_P$	−0.503	0.096	0.094	0.943	0.688	0.091	0.092	0.949

Table 4.14: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.85$, $a = 1.923$ (80^{th} percentile) and 2.593 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.502	0.119	0.120	0.952	0.692	0.119	0.120	0.954
		$\boldsymbol{\theta}_C$	−0.502	0.113	0.114	0.961	0.692	0.111	0.113	0.955
		$\boldsymbol{\theta}_P$	−0.503	0.099	0.100	0.959	0.692	0.099	0.100	0.954
	50%	$\boldsymbol{\theta}_R$	−0.491	0.150	0.152	0.951	0.698	0.148	0.152	0.956
		$\boldsymbol{\theta}_C$	−0.492	0.124	0.128	0.951	0.699	0.120	0.125	0.958
		$\boldsymbol{\theta}_P$	−0.496	0.099	0.102	0.955	0.695	0.097	0.100	0.960
90%	20%	$\boldsymbol{\theta}_R$	−0.503	0.118	0.119	0.954	0.690	0.117	0.119	0.947
		$\boldsymbol{\theta}_C$	−0.503	0.113	0.114	0.956	0.688	0.112	0.112	0.958
		$\boldsymbol{\theta}_P$	−0.503	0.094	0.094	0.954	0.688	0.094	0.094	0.950
	50%	$\boldsymbol{\theta}_R$	−0.501	0.152	0.151	0.948	0.691	0.156	0.151	0.944
		$\boldsymbol{\theta}_C$	−0.503	0.130	0.129	0.952	0.690	0.128	0.126	0.952
		$\boldsymbol{\theta}_P$	−0.500	0.092	0.093	0.952	0.693	0.093	0.092	0.949

Table 4.15: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.5$, $a = 1.999$ (80^{th} percentile) and 2.649 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.502	0.062	0.059	0.937	0.690	0.059	0.059	0.962
		$\boldsymbol{\theta}_C$	−0.501	0.057	0.055	0.946	0.692	0.053	0.054	0.948
		$\boldsymbol{\theta}_P$	−0.501	0.052	0.050	0.936	0.692	0.048	0.050	0.949
	50%	$\boldsymbol{\theta}_R$	−0.501	0.077	0.075	0.947	0.693	0.072	0.075	0.964
		$\boldsymbol{\theta}_C$	−0.502	0.061	0.060	0.953	0.694	0.055	0.056	0.954
		$\boldsymbol{\theta}_P$	−0.502	0.051	0.050	0.945	0.694	0.048	0.049	0.956
90%	20%	$\boldsymbol{\theta}_R$	−0.500	0.063	0.059	0.941	0.695	0.060	0.059	0.943
		$\boldsymbol{\theta}_C$	−0.501	0.059	0.056	0.936	0.693	0.053	0.054	0.954
		$\boldsymbol{\theta}_P$	−0.501	0.050	0.047	0.933	0.693	0.047	0.048	0.953
	50%	$\boldsymbol{\theta}_R$	−0.497	0.076	0.075	0.950	0.692	0.076	0.075	0.952
		$\boldsymbol{\theta}_C$	−0.499	0.060	0.061	0.946	0.692	0.056	0.056	0.954
		$\boldsymbol{\theta}_P$	−0.500	0.046	0.047	0.953	0.691	0.046	0.046	0.951

Table 4.16: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.85$, $a = 1.923$ (80^{th} percentile) and 2.593 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.499	0.061	0.059	0.943	0.693	0.061	0.059	0.940
		$\boldsymbol{\theta}_C$	−0.500	0.058	0.057	0.935	0.693	0.058	0.056	0.935
		$\boldsymbol{\theta}_P$	−0.500	0.051	0.050	0.949	0.693	0.050	0.050	0.955
	50%	$\boldsymbol{\theta}_R$	−0.495	0.079	0.075	0.937	0.697	0.078	0.075	0.946
		$\boldsymbol{\theta}_C$	−0.496	0.065	0.064	0.955	0.697	0.065	0.062	0.949
		$\boldsymbol{\theta}_P$	−0.496	0.051	0.051	0.949	0.697	0.051	0.050	0.947
90%	20%	$\boldsymbol{\theta}_R$	−0.502	0.060	0.059	0.950	0.691	0.061	0.059	0.952
		$\boldsymbol{\theta}_C$	−0.502	0.058	0.057	0.948	0.691	0.057	0.056	0.943
		$\boldsymbol{\theta}_P$	−0.501	0.049	0.047	0.940	0.692	0.049	0.047	0.938
	50%	$\boldsymbol{\theta}_R$	−0.499	0.076	0.075	0.937	0.694	0.075	0.075	0.950
		$\boldsymbol{\theta}_C$	−0.500	0.064	0.065	0.953	0.693	0.063	0.063	0.956
		$\boldsymbol{\theta}_P$	−0.500	0.045	0.046	0.959	0.694	0.045	0.046	0.964

Table 4.17: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $X_1 = X_2 \sim N(0,1)$.

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	80%	20%	1.29	0.96	1.03	1.04
		50%	1.05	0.98	1.10	0.98
	90%	20%	1.41	1.06	1.34	1.14
		50%	1.21	1.24	1.29	1.11
0.85	80%	20%	1.21	1.22	1.18	1.15
		50%	1.14	1.10	1.12	1.14
	90%	20%	1.45	1.25	1.26	1.19
		50%	1.34	1.30	1.44	1.40

Table 4.18: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$ and $X_1 = X_2 \sim N(0,1)$.

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	80%	20%	1.07	1.11	1.12	1.00
		50%	1.24	1.33	1.07	1.07
	90%	20%	1.40	1.15	1.35	1.21
		50%	1.41	1.13	1.50	1.15
0.85	80%	20%	1.32	1.18	1.30	1.28
		50%	1.22	1.08	1.23	1.12
	90%	20%	1.37	1.28	1.34	1.22
		50%	1.30	1.18	1.31	1.27

Table 4.19: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$ and $X_1 = X_2 \sim N(0,1)$.

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	80%	20%	1.04	1.28	1.02	1.05
		50%	1.13	1.19	1.15	1.27
	90%	20%	1.22	1.23	1.15	1.27
		50%	1.18	1.39	1.36	1.39
0.85	80%	20%	1.20	1.23	1.04	1.02
		50%	1.08	1.12	1.12	1.07
	90%	20%	1.22	1.17	1.28	1.32
		50%	1.28	1.33	1.47	1.52

Table 4.20: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$ and $X_1 = X_2 \sim N(0,1)$.

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	80%	20%	1.04	1.27	1.02	1.05
		50%	1.13	1.16	1.14	1.26
	90%	20%	1.22	1.23	1.14	1.26
		50%	1.18	1.33	1.34	1.37
0.85	80%	20%	1.20	1.23	1.04	1.02
		50%	1.10	1.13	1.12	1.07
	90%	20%	1.23	1.18	1.29	1.31
		50%	1.32	1.37	1.51	1.56

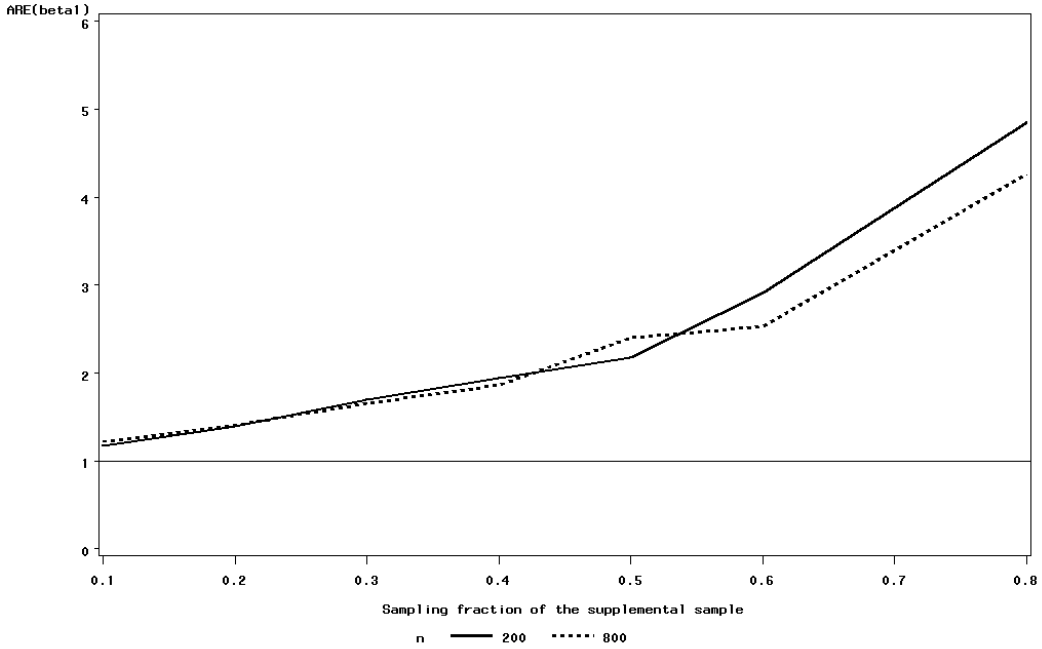

Figure 4.1: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_R$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample, under the models in Tables 4.9 and 4.11 with $a = 80\%$.

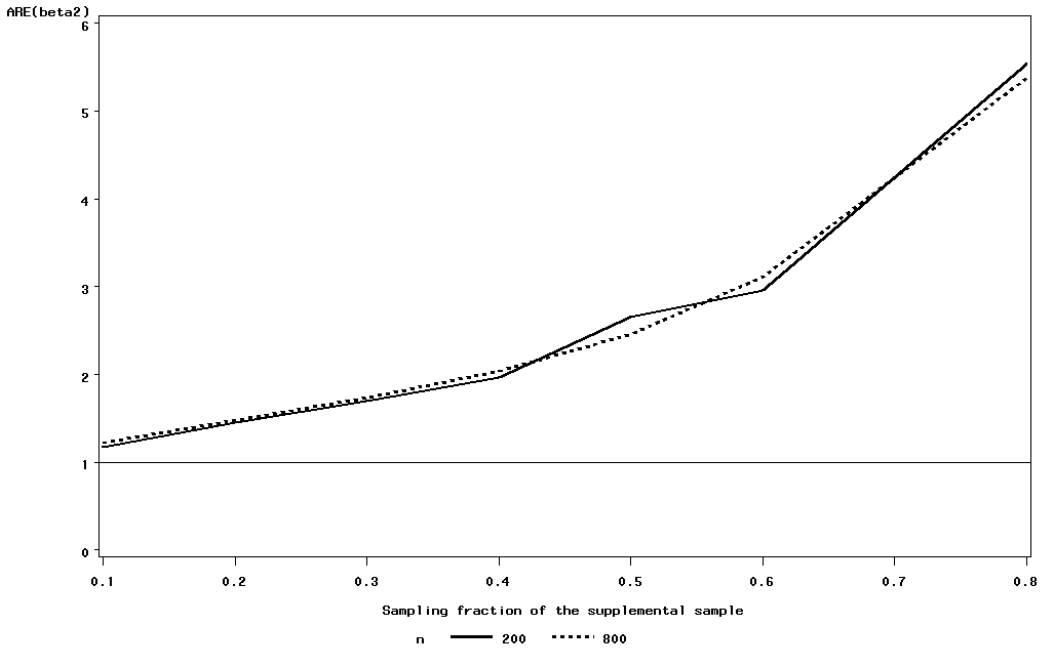

Figure 4.2: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_R$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample, under the models in Tables 4.9 and 4.11 with $a = 80\%$.

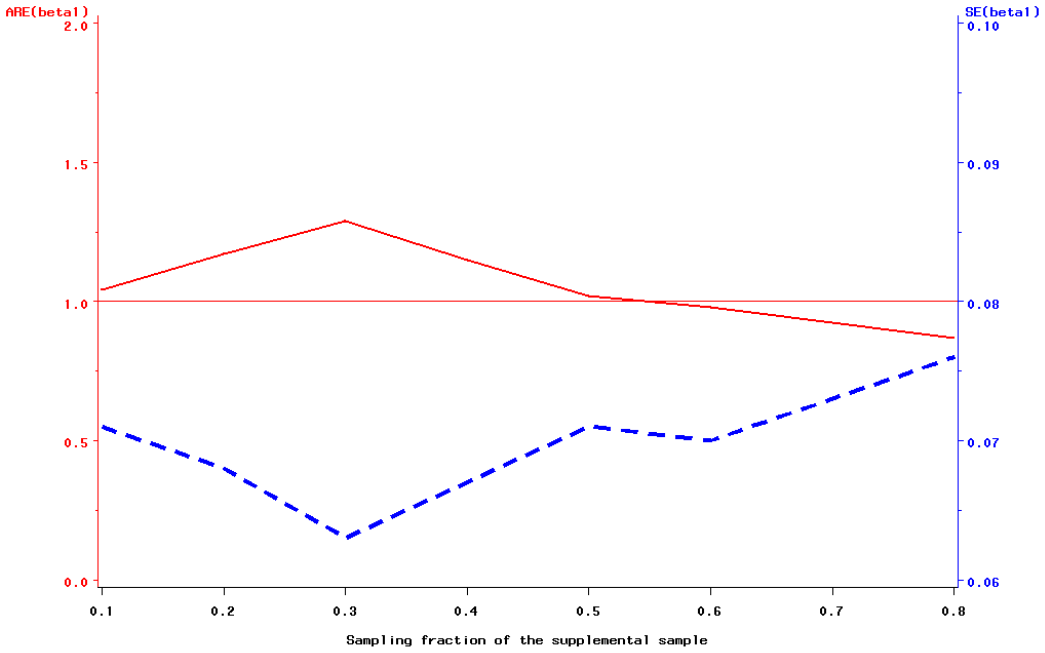

Figure 4.3: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample, under the model in Table 4.9 with $a = 80\%$.

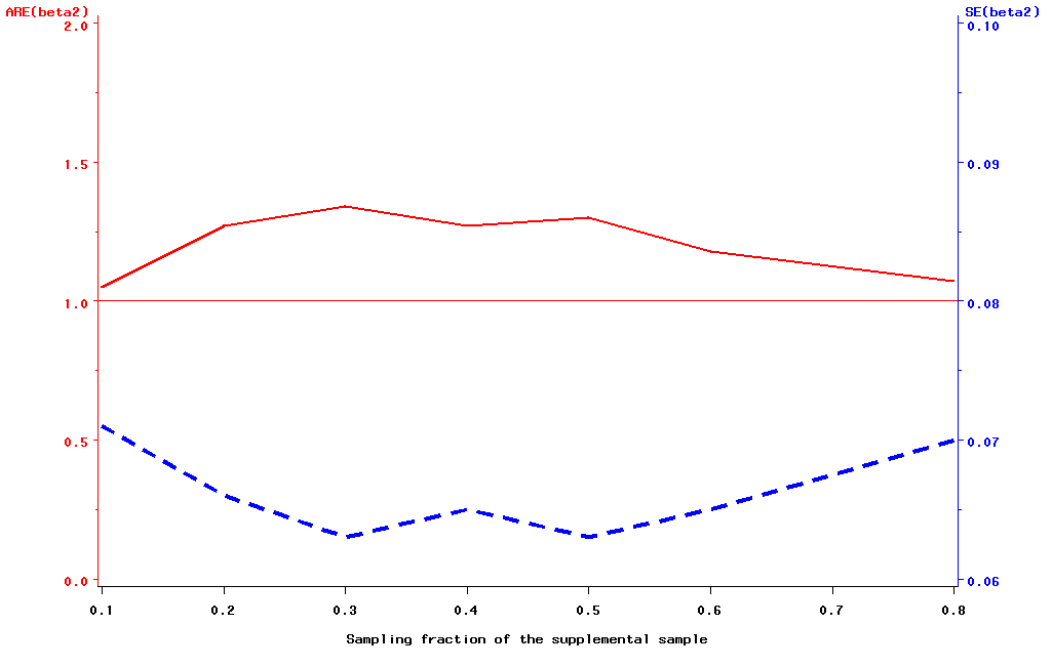

Figure 4.4: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample, under the model in Table 4.9 with $a = 80\%$.

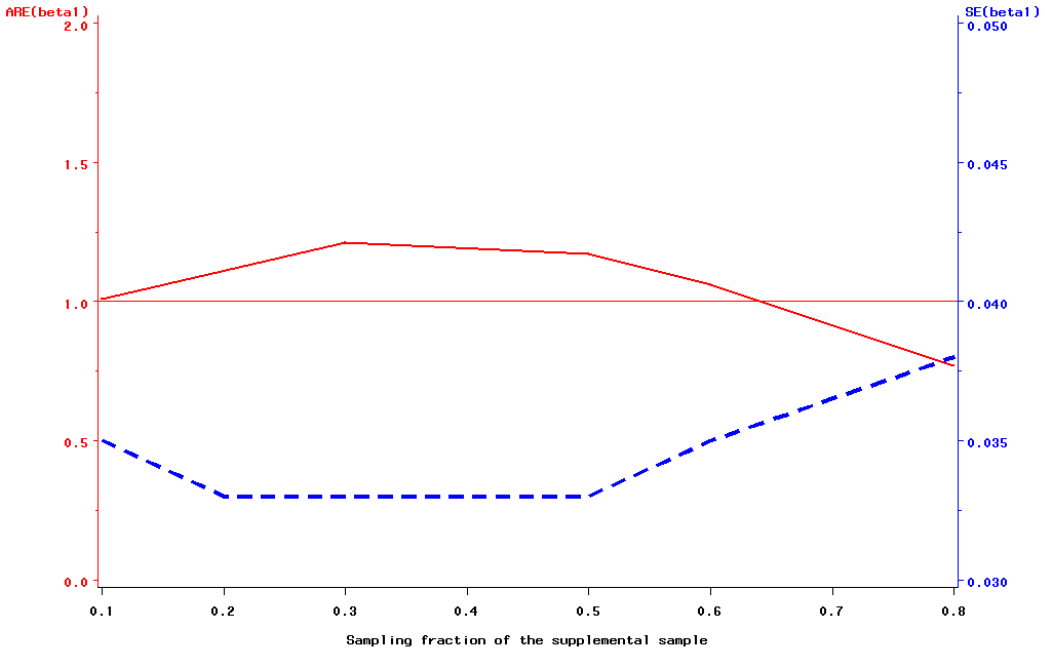

Figure 4.5: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample under the model in Table 4.11 with $a = 80\%$.

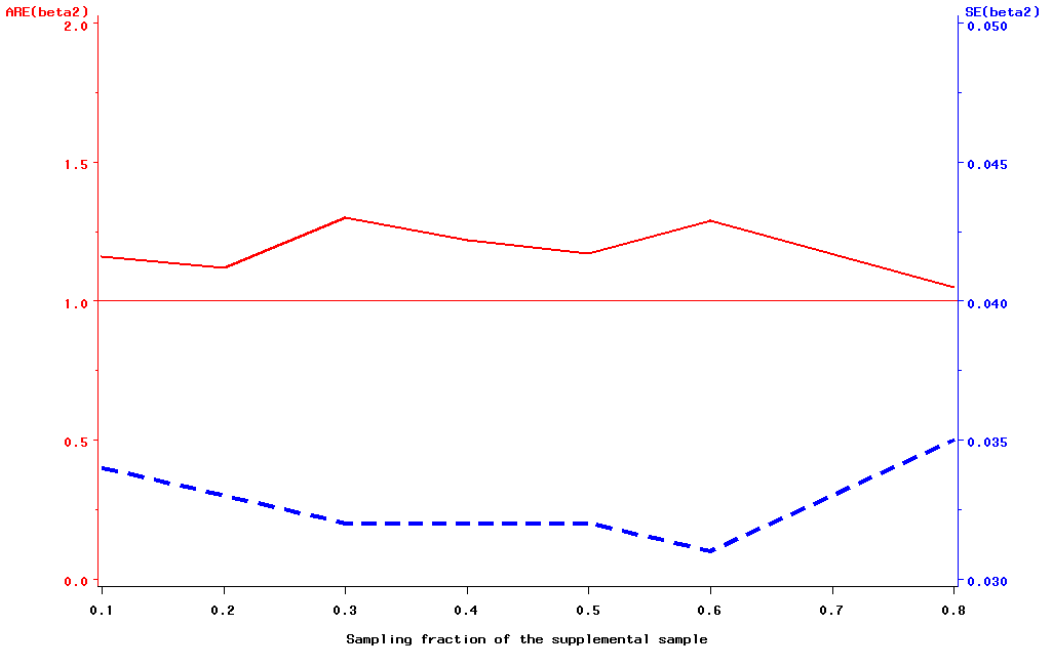

Figure 4.6: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample under the model in Table 4.11 with $a = 80\%$.

4.4 Application to the Collaborative Perinatal Project Data

4.4.1 The CPP Data

We applied the proposed method to analyze the Collaborative Perinatal Project (CPP) data to study the effect of the third trimester maternal pregnancy serum level of polychlorinated biphenyls (PCBs) on hearing loss children. The CPP was a prospective study designed to identify determinants of neurodevelopmental deficits in children. Details were described in Section 1.2.1. Nearly 56,000 pregnant women were recruited into the CPP study from 1959 through 1966 through 12 study centers across the United States. Women were enrolled, usually at their first prenatal visit; it resulted in 55,908 pregnancies. Data were collected on the mothers at each prenatal visit and at delivery and when the children were 24 hours, 4 and 8 months, and 1, 3, 4, 7, and 8 years.

In a recent environmental epidemiologic study (Longnecker et al., 2001 and 2004), the researchers were interested in studying the relationship between the audiometric evaluation, which was done when the children were approximately 8 years old, and *in utero* exposure to polychlorinated biphenyls (PCBs) measured as the third trimester maternal serum PCB level. The study subjects were children born into the CPP. There were 44,075 eligible children who met the following criteria: (1) live born singleton, and (2) a 3-ml third trimester maternal serum specimen was available. The investigators obtained exposure measurements for an outcome-dependent subsample from the population. In particular, the planned sampling design included an SRS of 1,200 subjects from eligible children, of whom 726 had an 8-year audiometric evaluation and a supplemental sample of 200 children whose audiometric evaluation showed sensorineural hearing loss (SNHL),

defined defined by a hearing threshold $\geq$ 13.3 dB according to the average across both ears at 1000, 2000, and 4000 Hz, without any evidence of conductive hearing loss. Evidence of conductive hearing loss exists when the air-bone difference in hearing threshold is $\geq$ 10 dB again based on the average across both ears. It was anticipated that a sampling design where children with SNHL were oversampled was to enhance the study efficiency relative to an SRS design of the same size.

In our analysis, we took the average measurements at frequencies 1000, 2000, and 4000 Hz for each ear separately to be the continuous outcome variables. The exposure variable of interest was the third trimester maternal serum PCB level (PCB) measured in $\mu g/L$. Additional factors considered potentially confounding included, for the mother, the socioeconomic index (SEI) score and the highest education level attained when giving birth (EDUC), and the race (RACE) and the gender (SEX) of the child. The covariate of RACE was coded to have two levels: 1 = "White", 0 = "Black and Others". The covariate SEX was coded 1 for males and 0 for females.

We considered the subjects who did not have missing observations for the variables selected into the model fitting and we assumed that missing data were missing completely at random. Of the 44,075 eligible children, 1,256 subjects were selected at random, of which 729 had complete data for the variables mentioned above and will then represent the study population in our data analysis. In order to adjust for our selection criterion described in the previous section, we considered the cutpoint of 13.3 dB. As a result, 156 out of 729 subjects represented the SNHL sample, whose maximum hearing levels were above the cutpoint. To illustrate our proposed method with the application of real data, we considered the following design with the total sample size $n = 200$ under the

Multivariate-ODS design with a maximum selection criterion: an overall simple random sample of size $n_0 = 150$ from 729 supplemented with an additional supplemental sample of $n_1 = 50$ drawn from the remaining subjects in the SNHL sample.

4.4.2 The Conditional Model

After examining the distributions of the hearing levels across three frequencies for each ear, we transformed the outcome variables on the natural log scale in order to exploit the normal properties. We therefore fitted the following linear model to the CPP *Multivariate-ODS* data,

$$\ln(Hearing_{ij}) = \beta_{0j} + \beta_{1j}PCB_i + \beta_{2j}SEX_{ij} + \beta_{3j}RACE_{ij} + \beta_{4j}EDUC_{ij} + \beta_{5j}SEI_{ij} + \epsilon_j \ , \tag{4.3}$$

where $i = 1, \ldots, 828$; $\epsilon_j \sim N(0, \sigma_j^2)$, where $j = 1$ representing the hearing level across three frequencies from the left ear and $j = 2$ from the right ear; $\rho = Corr(\epsilon_1, \epsilon_2)$. We assumed that $f(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\theta})$ is bivariate normal, where $\boldsymbol{\theta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \sigma_1^2, \sigma_2^2)$, and $\boldsymbol{\beta}_j^T = (\beta_{0j}, \ldots, \beta_{5j})$ and $j = 1, 2$. We estimated the parameters using the methods considered in the simulation studies: $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_R$.

4.4.3 Results

Table 4.21 presented the results of the parameter estimates, the estimated standard errors and the approximate 95% confidence intervals calculated based on the asymptotic normal distributions for each method. Both analyses showed that the corresponding 95%

confidence intervals for the PCB effect included 0. Thus, we would conclude that *in utero* PCB exposure did not have a significant effect on hearing levels for both ears. Observing the confidence intervals for other confounding parameters for the left ear, the covariate RACE showed a significant effect at the nominal level of 0.05, agreed by both methods and the significance was concluded for both ears. It indicated that white children had negative impact on hearing loss; in other words, white children were more likely to have better hearing ability than black and other children.

Although PCB was not significant, we could still see some efficiency gains from the results; the observed 95% confidence intervals for PCB provided by the proposed estimator $\widehat{\boldsymbol{\theta}}_P$ were narrower for both ears, compared with the CI obtained by $\widehat{\boldsymbol{\theta}}_R$; take the CI from the left ear as an example: $(-0.02, 0.10)$ for $\widehat{\boldsymbol{\theta}}_P$ versus $(-0.04, 0.12)$ for $\widehat{\boldsymbol{\theta}}_R$. Furthermore, it is clear to see that the proposed method resulted in substantially smaller standard errors for both ears than the competing method and there were gains in efficiency of the proposed method. We believe that our proposed method performed well when considering data in the *Multivariate-ODS* design.

Table 4.21: Results of modeling fitting for the CPP data with $n_0 = 150$, $n_1 = 50$, and $n = 200$.

		$\boldsymbol{\theta}_R$ ($n_0 = 150$)			$\boldsymbol{\theta}_P$ ($n = 200$)		
		$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI	$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI
Left Ear	Int	1.92	0.39	$(1.16, 2.67)$	2.06	0.28	$(1.50, 2.61)$
	PCB	0.04	0.04	$(-0.04, 0.12)$	0.04	0.03	$(-0.02, 0.10)$
	SEX	0.11	0.14	$(-0.17, 0.39)$	0.14	0.11	$(-0.07, 0.35)$
	RACE	-0.79	0.17	$(-1.12, -0.45)$	-0.74	0.13	$(-0.99, -0.49)$
	EDUC	-0.02	0.04	$(-0.09, 0.06)$	-0.03	0.03	$(-0.09, 0.02)$
	SEI	0.04	0.05	$(-0.06, 0.13)$	0.03	0.04	$(-0.04, 0.10)$
Right Ear	Int	2.35	0.36	$(1.65, 3.06)$	2.28	0.27	$(1.76, 2.79)$
	PCB	0.01	0.04	$(-0.06, 0.08)$	-0.02	0.03	$(-0.07, 0.04)$
	SEX	-0.10	0.13	$(-0.36, 0.16)$	-0.07	0.10	$(-0.26, 0.13)$
	RACE	-0.66	0.16	$(-0.96, -0.36)$	-0.54	0.12	$(-0.78, -0.31)$
	EDUC	-0.04	0.04	$(-0.11, 0.03)$	-0.02	0.03	$(-0.07, 0.03)$
	SEI	0.03	0.05	$(-0.05, 0.12)$	0.00	0.03	$(-0.06, 0.06)$

CHAPTER 5

STATISTICAL INFERENCES FOR MULTIVARIATE-ODS WITH A SUMMATION CRITERION

5.1 Introduction

To investigate the relationships between a disease outcome and an exposure given other characteristics, epidemiology and other biomedical studies often rely on the observational study designs. Cohort and case-control studies are most commonly used designs. The cohort study is to observe several individual exposures and the individual disease occurrence on the basis of a follow-up period and could take a long time to obtain the results. It could cost a lot to conduct a study especially when the disease is rare. Case-control design, on the other hand, is retrospective and studying the patients already having a disease to yield more information on risk factors of this group of people that differ from those who are free of disease (Cornfield, 1951). The case-control study in epidemiology or the choice-based sampling in econometrics are examples of a general scheme, *outcome-dependent sampling* (ODS) design, where the individuals are selected with probabilities depending on their observed outcome variables. The ODS design is appealing in practice because it allows the researchers to concentrate resources

on observations with the greatest amount of information of primary interest (Anderson, 1972).

Much work for studying dichotomous outcomes under an ODS setting has been continuously developed (e.g., White, 1982; Prentice, 1986; Brewlow and Cain, 1988; Lawless et al., 1999; Zhao and Lipsitz, 1992; Schill et al., 1993; Wacholder and Weinberg, 1994; Breslow and Holubkov, 1997; Wang and Zhou, 2006, 2008). The approach to dichotomize or categorize the outcome variable is commonly applied when the outcome is continuous and then one can conduct available statistical methods on the categorical outcomes. However, a selection bias often occurs since such a simplification for the outcome would induce a loss of efficiency and information and increase the risk for misclassification (Suissa, 1991; Zhou et al., 2002; Weaver and Zhou, 2005), especially when the results are sensitive to the choice of the cutpoints.

To directly apply the continuous scale of the outcome variable without losing information on dichotomization, Zhou et al. (2002) considered a general ODS scheme where (i) an overall simple random sample was drawn from the base population (the prospective component); and (ii) additional supplement samples were randomly selected from segments of the outcome space of particular interest (the retrospective component). They proposed a maximum semiparametric empirical likelihood inference procedure without specifying the underlying distribution for the covariates. Weaver and Zhou (2005) further developed a maximum estimated likelihood estimator (MELE) for the continuous outcome under a two-stage ODS scheme. These methods, however, were developed for the case with the univariate continuous outcome.

In practice, multivariate data arise in many contexts, for example, in epidemiological

cohort studies where the outcomes are recorded for members within families, in animal experiments in which treatments are applied to samples of littermates, or in most clinical trials where study subjects are experiencing multiple events. Among these studies, the correlation between the responses cannot be neglected. An increasing number of studies are indeed performed using the *Multivariate-ODS* design, a further generalization of the biased sampling, which is built on the idea of the ODS design with an aggregate of the responses in the multivariate form and at the same preserves the advantages of the ODS. An example of the ongoing study will be given to illustrate this idea in the next paragraph. The usual statistical method for analyzing the multivariate data if accounting for the *Multivariate-ODS* design is no longer appropriate. A statistical inference procedure is needed to take advantage of the *Multivariate-ODS* setting.

We are motivated by the Collaborative Perinatal Project (CPP), a prospective cohort study designed to identify determinants of neurodevelopmental deficits in children (Niswander and Gordon, 1972; Gray et al., 2000). Longnecker et al. (2004) studied the association in humans between maternal third trimester serum polychlorinated biphenyls (PCBs) levels and audiometry results in offsprings at approximately 8 years old. The sample selected by the investigators was according to an ODS scheme: 726 having an 8-year audiometric evaluation of 1200 subjects were selected at random from the underlying population and a supplemental sample of 200 eligible children was randomly selected from the 440 children whose 8-year audiometric evaluation showed sensorineural hearing loss (SNHL). It was anticipated that a sampling design where children with SNHL were oversampled was to enhance the study efficiency relative to an SRS design of the same size. The outcome variable discussed in the paper was whether the child had hearing loss,

defined from each individual's mean hearing level across both ears and then dichotomized by a threshold. Our goal is to develop a proper inference procedure by considering the continuous hearing measures from both ears simultaneously under the *Multivariate-ODS* design to achieve greater efficiency than only considering a simple random sample or alternatively simply dichotomizing the continuous outcome.

In this chapter we consider statistical inferences on regression models under a *Multivariate-ODS* design. Specifically, we model the underlying distributions of covariates nonparametrically using the empirical likelihood methods. A novelty of the proposed method is that one can make inferences on the regression parameters without postulating any of the distributions for the covariates by combining a nonparametric component with a parametric regression model. We show that the proposed estimator with the outcome-dependent nature accounted for is more efficient and statistically powerful than other alternative methods. We also investigate that the sampling strategies under the *Multivariate-ODS* framework can be used to design a cost-effective study. The remainder of this chapter is as follows. Section 5.2 presents the notation and the data structure under the *Multivariate-ODS* design with multivariate continuous outcomes. We then demonstrate the likelihood approaches and derive the asymptotic properties. Section 5.3 describes the simulation studies and the small sample properties of our proposed estimator and compares with other methods. We thereafter apply the proposed method to analyze the data in Collaborative Perinatal Project study in Section 5.4 and Section 5.5 gives a brief discussion and suggests some possible extensions of the proposed method in future research. Additional simulation results are given in Section 5.6.

5.2 The Multivariate-ODS Design and Inference

5.2.1 The Multivariate-ODS Data Structure and Likelihood

Let Y_{ij} be the jth continuous outcome for the subject i, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$ $(p \geq 2)$, and $\boldsymbol{X}_i$ be a vector of covariates for the ith subject, which can include both discrete and continuous components. Motivated by the Collaborative Perinatal Project study described in Section 1, we will consider the supplemental sample selected based on the summation criterion under the *Multivariate-ODS* mechanism through this paper. Assume that the domain of interest, the sums of responses $\left\{\sum_{j=1}^{p} Y_{ij} \ , \forall i\right\}$, is partitioned into K mutually exclusive intervals by the known constants $-\infty = a_0 < a_1 < \ldots < a_{K-1} < a_K = \infty$, and the kth interval is denoted as $C_k = (a_{k-1}, a_k]$, $k = 1, \ldots, K$. The data structure of the *Multivariate-ODS* design consists of two components: an overall *simple random sample* (SRS) of size n_0 (≥ 0) and a stratified *supplemental sample* of size n_k (≥ 0) randomly drawn from the interval, C_k:

(i) SRS Component: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i\right\}$, $i = 1, \ldots, n_0$;

(ii) Supplemental Component: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid \left(\sum_{j=1}^{p} Y_{ij}\right) \in C_k\right\}$, $i = 1, \ldots, n_k$, $j = 1, \ldots, p$ and $k = 1, \ldots, K$.

Without loss of generality, we assume that $p = 2$ and $K = 1$; in other words, each individual has two observations and one only selects the supplemental sample in the upper tail of the distribution of $\left\{\sum_{j=1}^{p} Y_{ij} \ , \forall i\right\}$, i.e., $C_1 = (a_1, \infty)$. To simplify the notation, we drop the subscript of the cutpoint a_1 and denote as a. Let $n = n_0 + n_1$ be the total sample size of the *Multivariate-ODS* we observe. The joint density of

$(\boldsymbol{Y}, \boldsymbol{X})$ can be written as $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X})$, where $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$ is the conditional density function of $\boldsymbol{Y}$ given $\boldsymbol{X}$, $\boldsymbol{\theta}$ is a vector of the regression coefficients of interest, and $g_{\boldsymbol{X}}(\boldsymbol{X})$ is the marginal density of $\boldsymbol{X}$, which is independent of $\boldsymbol{\theta}$. The corresponding unknown distribution function of $\boldsymbol{X}$ is denoted as $G_{\boldsymbol{X}}(\boldsymbol{X})$. The joint likelihood function for the observed data obtained through the *Multivariate-ODS* design is

$$L_S(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right]\left[\prod_{i=1}^{n_1} f(\boldsymbol{Y}_i, \boldsymbol{X}_i|(Y_{i1}+Y_{i2}) > a;\boldsymbol{\theta})\right] , \tag{5.1}$$

where the first bracket is the likelihood corresponding to the observations from the SRS portion of the *Multivariate-ODS* and the second quantity represents the likelihood contributions of the observations in the supplemental sample. Using Bayes' Law, the likelihood function can be further rewritten as

$$\begin{aligned} L_S(\boldsymbol{\theta}, G_{\boldsymbol{X}}) &= \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right]\left[\prod_{i=1}^{n_1} \frac{f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{1-\Pr(Y_{i1}+Y_{i2}<a)}\right] \\ &= \left[\prod_{i=1}^{n_0} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right]\left[\prod_{i=1}^{n_1} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X}_i) \times \prod_{i=1}^{n_1} \frac{1}{1-\pi(\boldsymbol{\theta}, G_{\boldsymbol{X}})}\right] \\ &= \left[\prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})\right]\left[\left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \times (1-\pi)^{-n_1}\right] \\ &= L_{S1}(\boldsymbol{\theta}) \times L_{S2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) , \end{aligned} \tag{5.2}$$

where

$$L_{S1}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) \tag{5.3}$$

and

$$L_{S2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \times (1-\pi)^{-n_1} \ . \tag{5.4}$$

Note that for simplicity, we define that

$$P_0(\boldsymbol{X};\boldsymbol{\theta}) = \Pr\{Y_1 + Y_2 < a|\boldsymbol{X}\} = \int_{-\infty}^{\infty}\int_{-\infty}^{a-Y_2} f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})dY_1dY_2 \tag{5.5}$$

and

$$\pi = \pi(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \int_{\boldsymbol{X}} P_0(\boldsymbol{X};\boldsymbol{\theta})dG_{\boldsymbol{X}} \tag{5.6}$$

are the conditional and the marginal probabilities that the sum of the elements in $\boldsymbol{Y}$ is less than a, respectively.

There are several possible approaches that could be used to make inferences about $\boldsymbol{\theta}$. Without knowing $G_{\boldsymbol{X}}$, one of the naive approaches is to take the observations in the SRS portion of the *Multivariate-ODS* and derive a maximum likelihood estimator for $\boldsymbol{\theta}$. However, ignoring the information from the supplemental sample would lose accuracy and efficiency. Or, one could obtain $\boldsymbol{\theta}$ by maximizing the conditional likelihood based on the complete data in the *Multivariate-ODS*. Clearly, these two estimators are not the most efficient since the information regarding the supplemental sample is not fully accounted. If $G_{\boldsymbol{X}}(\boldsymbol{X})$ is parameterized to a parameter vector, say ξ, one could maximize the resulting $L_S(\boldsymbol{\theta}, \widehat{G}_{\boldsymbol{X}})$ subject to $(\boldsymbol{\theta}, \xi)$. However, misspecification of $G_{\boldsymbol{X}}$ could lead to erroneous conclusions so that such approach will be limited only if the form of $G_{\boldsymbol{X}}$ is correctly specified. As a result, a nonparametric modeling of $G_{\boldsymbol{X}}$ is desirable in this case. Nevertheless, $G_{\boldsymbol{X}}$ is an infinite-dimensional nuisance parameter and cannot be easily factored out of $L_{S2}(\boldsymbol{\theta}, G_{\boldsymbol{X}})$. Thus, to incorporate all the available information in the *Multivariate-ODS* data without specifying $G_{\boldsymbol{X}}$, one needs a new method that will

be tractable both theoretically and computationally. We next describe a semiparametric empirical likelihood estimator, where $G_{\boldsymbol{X}}$ is left unspecified.

5.2.2 A Semiparametric Likelihood Approach for the Multivariate-ODS

To outline our approach for estimating $\boldsymbol{\theta}$, we develop a profile likelihood function for $\boldsymbol{\theta}$ by first maximizing $L_{S2}(\boldsymbol{\theta}, G_{\boldsymbol{X}})$ with $\boldsymbol{\theta}$ fixed and $G_{\boldsymbol{X}}$ treated as a nonparametric maximum likelihood estimate (NPMLE) (Vardi, 1985), a function of $\boldsymbol{\theta}$ and π, and obtaining an empirical estimator $\widehat{\boldsymbol{\theta}}$ by maximizing the resulting profile log likelihood function over $\boldsymbol{\theta}$. The procedure is detailed in the following.

We first maximize $L_S(\boldsymbol{\theta}, G_{\boldsymbol{X}})$, with $\boldsymbol{\theta}$ fixed, over all discrete distributions whose support includes the observed values by considering a discrete distribution function (i.e. a step function) which has all of its probability located at the observed data points (Vardi, 1985). Let $p_i = dG_{\boldsymbol{X}}(\boldsymbol{X}_i) = g_{\boldsymbol{X}}(\boldsymbol{X}_i)$, $i = 1, \ldots, n$, be the probability mass for the ith covariate vector. We want to find values $\{\widehat{p}_i, \forall i\}$, which maximize the log likelihood function corresponding to (5.2)

$$l_S(\boldsymbol{\theta}, \{p_i\}) = \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i | \boldsymbol{X}_i; \boldsymbol{\theta}) + \sum_{i=1}^{n} \ln p_i - n_1 \ln(1-\pi) \tag{5.7}$$

under the following constraints:

$$\left\{ p_i \geq 0 \ \forall i, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i \Big(P_0(\boldsymbol{X}_i; \boldsymbol{\theta}) - \pi \Big) = 0 \right\} . \tag{5.8}$$

The above conditions reflect the fact that $G_{\boldsymbol{X}}$ is a discrete distribution function. For a

fixed $\boldsymbol{\theta}$, there exists a unique maximum for $\{p_i\}$ in (5.7) subject to the constraints in (5.8) if 0 is inside the convex hull of the points $\{P_0(\boldsymbol{X}_i;\boldsymbol{\theta}),\ \forall\ i\}$ (Qin and Lawless, 1994). We use the Lagrange multiplier argument to obtain $\widehat{p}_i$ through maximizing H_S,

$$\begin{aligned} H_S(\boldsymbol{\theta},\{p_i\},\mu,\lambda) &= \sum_{i=1}^{n}\ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})+\sum_{i=1}^{n}\ln p_i - n_1\ln(1-\pi) \\ &\quad -\mu\left(\sum_{i=1}^{n}p_i-1\right)-n\lambda\sum_{i=1}^{n}p_i\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\ , \end{aligned} \tag{5.9}$$

where μ and λ are the Lagrange multipliers corresponding to the normalized restriction on the $\{\widehat{p}_i,\forall i\}$. With $\boldsymbol{\theta}$ fixed, taking the derivative of H with respect to p_i, solving the score equation and applying the constraints in (5.8), we obtain $\widehat{\mu}=n$ and

$$\widehat{p}_i=\left\{n\ \left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]\right\}^{-1}\ . \tag{5.10}$$

Substituting $\{\widehat{p}_i\}$ back into (5.7), we then have the resulting profile log likelihood function,

$$l_S(\boldsymbol{\phi}_{SM})=\sum_{i=1}^{n}\ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta})-\sum_{i=1}^{n}\ln n\left[1+\lambda\Big(P_0(\boldsymbol{X}_i;\boldsymbol{\theta})-\pi\Big)\right]-n_1\ln(1-\pi)\ , \tag{5.11}$$

where $\boldsymbol{\phi}_{SM}^T=(\boldsymbol{\theta}^T,\lambda,\pi)$ is a combined parameter vector; λ and π are treated as the parameters independent of $\boldsymbol{\theta}$. We refer $\widehat{\boldsymbol{\phi}}_{SM}$, a maximizer for (5.11), as the *semiparametric empirical maximum likelihood estimator* (SEMLE). The Newton-Raphson iterative algorithm is used to solve the score equation from (5.11).

5.2.3 Asymptotic Properties of the SEMLE

The main results for $\boldsymbol{\phi}_{SM}$ regarding the existence and consistency, asymptotic normality, and a consistent estimator for the asymptotic variance-covariance matrix are demonstrated as three theorems, respectively. Outlines of the proofs of the main results are provided in the Appendix.

We indicate $\boldsymbol{\phi}^0_{SM}$ as the true parameter vector of interest containing $\boldsymbol{\theta}^0$, π^0 and λ^0, where π^0 is the true marginal probability that the sum of one's observations is less than the cutpoint, a, and λ^0 is the true Lagrange multiplier.

Theorem 5.1 (Consistency of the SEMLE): *With probability going to 1 as $N \to \infty$, there exists a sequence $\{\widehat{\boldsymbol{\phi}}_{SM}\}$ of solutions to the score equations from (5.11) such that $\widehat{\boldsymbol{\phi}}_{SM} \xrightarrow{p} \boldsymbol{\phi}^0_{SM}$, where $\boldsymbol{\phi}^0_{SM}$ is the true parameter vector of interest. If another sequence $\{\bar{\boldsymbol{\phi}}_{SM}\}$ of solutions to the score equations exists such that $\bar{\boldsymbol{\phi}}_{SM} \xrightarrow{p} \boldsymbol{\phi}^0_{SM}$, then $\bar{\boldsymbol{\phi}}_{SM} = \widehat{\boldsymbol{\phi}}_{SM}$ with probability going to 1 as $n \to \infty$.*

Theorem 5.2 (Asymptotic Normality of the SEMLE): *The SEMLE has the following asymptotic normal distribution:*

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_{SM} - \boldsymbol{\phi}^0_{SM}) \xrightarrow{D} N_{(p+2)}(\mathbf{0},\ \boldsymbol{\Sigma}(\boldsymbol{\phi}^0_{SM}))\ ,$$

with the asymptotic variance-covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{V}\boldsymbol{J}^{-1}\ , \tag{5.12}$$

where

$$\boldsymbol{J} = -\frac{\partial^2 \widetilde{l}_S(\boldsymbol{\phi}^0_{SM})}{\partial \boldsymbol{\phi}_{SM} \partial \boldsymbol{\phi}^T_{SM}}$$

and

$$\boldsymbol{V} = \mathrm{Var}\left[\frac{\partial l_S(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}^0_{SM})}{\partial \boldsymbol{\phi}_{SM}}\right],$$

where $\widetilde{l}_S$ is the limiting form of l_S.

Theorem 5.3 (A Consistent Estimator for the Asymptotic Variance-Covariance Matrix): *A consistent estimator for the variance-covariance matrix shown in Equation (5.12) is*

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\phi}}_{SM}) = \widehat{\boldsymbol{J}}^{-1}(\widehat{\boldsymbol{\phi}}_{SM}) \widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_{SM}) \widehat{\boldsymbol{J}}^{-1}(\widehat{\boldsymbol{\phi}}_{SM}),$$

where

$$\widehat{\boldsymbol{J}}(\boldsymbol{\phi}_{SM}) = -\frac{1}{n} \frac{\partial^2 l_S(\boldsymbol{\phi}_{SM})}{\partial \boldsymbol{\phi}_{SM} \partial \boldsymbol{\phi}^T_{SM}}$$

and

$$\widehat{\boldsymbol{V}}(\boldsymbol{\phi}_{SM}) = \frac{1}{n} \widehat{\mathrm{Var}}_{\{i\}}\left[\frac{\partial l_S(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\phi}^0_{SM})}{\partial \boldsymbol{\phi}_{SM}}\right].$$

5.3 Simulation Studies

We evaluate the performance of the proposed estimator in the small sample settings using the simulated data, generated according to the bivariate normal model:

$$\boldsymbol{Y}|\boldsymbol{X} \sim N\left[\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \\ \mu_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} {\sigma_1}^2 & \rho\sigma_1\sigma_2 \\ \\ \rho\sigma_1\sigma_2 & {\sigma_2}^2 \end{pmatrix}\right],$$

where $\boldsymbol{Y} = \left(Y_1, Y_2\right)^T$, $\boldsymbol{X} = \left(X_1, X_2\right)^T$, $\mu_1 = \alpha_1 + \beta_1 X_1$ and $\mu_2 = \alpha_2 + \beta_2 X_2$; i.e., the conditional distributions of Y_1 given X_1 and Y_2 given X_2 are normally distributed with means $\alpha_1 + \beta_1 X$ and $\alpha_2 + \beta_2 X$, variances σ_1^2 and σ_2^2, respectively, and the correlation coefficient ρ. Our goal is to estimate the parameter vector $\boldsymbol{\theta}_P = (\alpha_1, \beta_1, \alpha_2, \beta_2, \sigma_1, \sigma_2, \rho)^T$. In particular, we will investigate the behavior of β_1 and β_2 by fixing $\alpha_1 = 0.5$, $\alpha_2 = -0.8$, and $\sigma_1^2 = \sigma_2^2 = 1$. Then the same models are applied to $\rho = 0.5$ and $\rho = 0.85$ to see how the magnitude of association between outcome variables affects the parameter estimates.

The study *Multivariate-ODS* sample size was set to be $n = 200$. The *Multivariate-ODS* design we considered included an SRS and a supplemental sample from individuals with their summation of outcome values in the tail of the distribution of $\sum(\boldsymbol{Y})$, where $\sum(\boldsymbol{Y}) = \left\{\sum_{j=1}^{p} Y_{1j}, \sum_{j=1}^{p} Y_{2j}, \ldots, \sum_{j=1}^{p} Y_{nj}\right\}$. We considered the cutpoint to partition the space of $\sum(\boldsymbol{Y})$, a, of the 80th or 90th percentile from the distribution of $\sum(\boldsymbol{Y})$ under the study models. And the supplemental sampling fraction, $\gamma = n_1/n$, was either 20% or 50%. The parameter estimates and the corresponding standard errors for each setting were obtained from independent 1,000 data sets generated.

We compare our proposed estimator, $\widehat{\boldsymbol{\theta}}_P$, to the following competitive estimators under each setting in our simulation study: (i) the maximum likelihood estimator by maximizing the likelihood from the SRS portion of the *Multivariate-ODS* data ($\widehat{\boldsymbol{\theta}}_R$), (ii) the maximum likelihood estimator by maximizing the conditional likelihood based on the complete *Multivariate-ODS* data ($\widehat{\boldsymbol{\theta}}_C$), and (iii) the maximum likelihood estimator obtained from a random sample of the same size as the *Multivariate-ODS* sample ($\widehat{\boldsymbol{\theta}}_S$). Comparing $\widehat{\boldsymbol{\theta}}_P$ with $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$ will give us an insight of the impact on ignoring the part

of the information from the *Multivariate-ODS* sample. The comparison between $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_S$ will demonstrate the efficiency gain of the *Multivariate-ODS* design over the simple random sample of the same size.

The simulation results were presented in Tables 5.1 through 5.3. Within each table, the sampling specifications and the covariate distribution were fixed. The *Multivariate-ODS* sample size in Tables 5.1 and 5.2 was set to be $n = 200$ and the correlation coefficient was $\rho = 0.5$ in Table 5.1 and $\rho = 0.85$ in Table 5.2. The results in Tables 5.1 and 5.2 included the small sample properties of the proposed estimator and the competing estimators. Table 5.3 presented the relative efficiencies (ratios of variances) to evaluate the amount of information gained by implementing the *Multivariate-ODS* design.

In Table 5.1, we observed that three methods yielded unbiased means of the estimates compared with the "true" parameter values for all four settings. The proposed estimator $\widehat{\boldsymbol{\theta}}_P$ produced the smallest standard errors for estimating the model parameters, compared with $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$. On the other hand, $\widehat{\boldsymbol{\theta}}_R$ was the least efficient, which was expected since $\widehat{\boldsymbol{\theta}}_R$ was obtained using the least information. For $\widehat{\boldsymbol{\theta}}_P$, the means of the standard errors were relatively close to the "true" simulated standard errors. The confidence intervals based on the proposed estimator provided good coverage close to the nominal 95% level. The same findings were observed for both $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$. Above observations were true for both $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

Table 5.2 presented the same model designs as Table 5.1 but with a higher correlation coefficient, $\rho = 0.85$. The observations from Table 5.1 held in Table 5.2. Note that the standard errors for $\widehat{\boldsymbol{\theta}}_P$ were smaller as the correlation coefficient increased. For example, the standard errors were 0.068, 0.068, 0.065, and 0.065 for $\widehat{\beta}_1$ from $\widehat{\boldsymbol{\theta}}_P$ in Table 5.1 under

four settings whereas the corresponding standard errors in Table 5.2 were 0.064, 0.065, 0.064, and 0.061, respectively. This suggested that the proposed estimator be favored and even more efficient when the outcomes were more correlated. The same trend was observed for $\widehat{\beta}_2$. The pattern above agreed well with larger sample sizes (the results were not shown here).

For Table 5.3, we presented results from a relative efficiency study by comparing the *Multivariate-ODS* design to the design of a simple random sample of the same sample size under the same settings studied in Tables 5.1 and 5.2. We calculated the asymptotic relative efficiencies (ARE) of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$, $Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$, under two sample size considerations, $n = 200$ and $n = 800$. We observed that all the $AREs$ were greater than 1, suggesting that $\widehat{\boldsymbol{\theta}}_P$ was more efficient than $\widehat{\boldsymbol{\theta}}_S$ under all the circumstances. A higher degree of efficiency gains was observed when the two outcomes were more correlated; for example, as the correlation coefficient $\rho = 0.85$, the cutpoint $a = 80\%$ and the sampling fraction $\gamma = 50\%$, the efficiency gain for $\widehat{\boldsymbol{\theta}}_P$ over $\widehat{\boldsymbol{\theta}}_S$ was 33% whereas the efficiency gain was 17% as $\rho = 0.5$. From the efficiency study, we see that $\widehat{\boldsymbol{\theta}}_P$ led to more efficiency gains over $\widehat{\boldsymbol{\theta}}_S$ as the proportion of the supplemental data in the *Multivariate-ODS* increased, in which the outcomes were more correlated.

Table 5.1: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.175$ (80^{th} percentile) and 1.958 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.501	0.081	0.080	0.948	0.692	0.081	0.080	0.952
		$\boldsymbol{\theta}_C$	−0.499	0.079	0.076	0.945	0.693	0.077	0.076	0.952
		$\boldsymbol{\theta}_P$	−0.499	0.068	0.068	0.953	0.693	0.067	0.068	0.951
	50%	$\boldsymbol{\theta}_R$	−0.500	0.105	0.101	0.941	0.690	0.099	0.101	0.959
		$\boldsymbol{\theta}_C$	−0.500	0.087	0.086	0.949	0.689	0.086	0.086	0.955
		$\boldsymbol{\theta}_P$	−0.500	0.068	0.069	0.951	0.690	0.069	0.069	0.942
90%	20%	$\boldsymbol{\theta}_R$	−0.495	0.080	0.079	0.955	0.698	0.082	0.079	0.937
		$\boldsymbol{\theta}_C$	−0.494	0.077	0.076	0.948	0.696	0.079	0.076	0.942
		$\boldsymbol{\theta}_P$	−0.497	0.065	0.064	0.940	0.693	0.067	0.064	0.944
	50%	$\boldsymbol{\theta}_R$	−0.500	0.102	0.102	0.954	0.691	0.102	0.102	0.949
		$\boldsymbol{\theta}_C$	−0.500	0.090	0.088	0.946	0.694	0.088	0.088	0.948
		$\boldsymbol{\theta}_P$	−0.500	0.065	0.064	0.941	0.694	0.063	0.064	0.951

Table 5.2: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.335$ (80^{th} percentile) and 2.192 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.499	0.075	0.080	0.961	0.693	0.078	0.080	0.958
		$\boldsymbol{\theta}_C$	−0.499	0.073	0.077	0.955	0.693	0.075	0.077	0.958
		$\boldsymbol{\theta}_P$	−0.500	0.064	0.067	0.957	0.692	0.066	0.067	0.962
	50%	$\boldsymbol{\theta}_R$	−0.497	0.099	0.101	0.956	0.694	0.101	0.101	0.944
		$\boldsymbol{\theta}_C$	−0.499	0.088	0.090	0.948	0.694	0.089	0.090	0.955
		$\boldsymbol{\theta}_P$	−0.499	0.065	0.068	0.968	0.694	0.065	0.068	0.960
90%	20%	$\boldsymbol{\theta}_R$	−0.498	0.080	0.080	0.954	0.696	0.081	0.080	0.946
		$\boldsymbol{\theta}_C$	−0.499	0.078	0.077	0.950	0.695	0.080	0.077	0.943
		$\boldsymbol{\theta}_P$	−0.501	0.064	0.063	0.940	0.693	0.065	0.063	0.941
	50%	$\boldsymbol{\theta}_R$	−0.502	0.100	0.101	0.952	0.682	0.102	0.101	0.943
		$\boldsymbol{\theta}_C$	−0.500	0.090	0.092	0.954	0.683	0.090	0.092	0.962
		$\boldsymbol{\theta}_P$	−0.496	0.061	0.062	0.953	0.697	0.062	0.062	0.950

Table 5.3: Simulation Results: Relative efficiencies ($Var_{\hat{\boldsymbol{\theta}}_S}/Var_{\hat{\boldsymbol{\theta}}_P}$) from the models presented in Tables 5.1 and 5.2 under different sample sizes, n.

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\hat{\beta}_1}$	$\widehat{ARE}_{\hat{\beta}_2}$	$\widehat{ARE}_{\hat{\beta}_1}$	$\widehat{ARE}_{\hat{\beta}_2}$
0.5	80%	20%	1.08	1.08	1.02	1.07
		50%	1.17	1.10	1.25	1.15
	90%	20%	1.24	1.11	1.32	1.29
		50%	1.27	1.33	1.22	1.29
0.85	80%	20%	1.30	1.22	1.12	1.08
		50%	1.33	1.25	1.12	1.16
	90%	20%	1.32	1.25	1.44	1.35
		50%	1.38	1.29	1.43	1.36

5.4 Analysis of the Collaborative Perinatal Project Data

We applied the proposed method to analyze the Collaborative Perinatal Project (CPP) data to study the effect of the third trimester maternal pregnancy serum level of polychlorinated biphenyls (PCBs) on hearing loss children. Nearly 56,000 pregnant women were recruited into the CPP study from 1959 through 1966 through 12 study centers across the United States. Women were enrolled, usually at their first prenatal visit; it resulted in 55,908 pregnancies (9,161 women contributed multiple pregnancies to the study). Data were collected on the mothers at each prenatal visit and at delivery and when the children were 24 hours, 4 and 8 months, and 1, 3, 4, 7, and 8 years. Among all the measures, we were interested in audiometric evaluation, which was done when the children were approximately 8 years old. In our selection of the subjects, we closely follow the selection criteria and the sampling scheme used in Longnecker et. al. (2004). There were 44,075 eligible children who met the following criteria: (1) live born singleton, and (2) a 3-ml third trimester maternal serum specimen was available. The audiometric evaluations showed sensorineural hearing loss (SNHL) was defined by a hearing threshold ≥ 13.3 dB according to the average across both ears at 1000, 2000, and 4000 Hz, without any evidence of conductive hearing loss. Evidence of conductive hearing loss exists when the air-bone difference in hearing threshold is ≥ 10 dB again based on the average across both ears.

We took the average measurements at frequencies 1000, 2000, and 4000 Hz for each ear separately to be the continuous outcome variables in our analysis of the CPP data. The exposure variable of interest was PCB measured in $\mu g/L$. Additional factors considered

potentially confounding included, for the mother, the socioeconomic index (SEI) score and the highest education level attained when giving birth (EDUC), and the race (RACE) and the gender (SEX) for children.

For the data analysis, we considered the subjects who did not have missing observations for any variable in the model fitting and we assumed that missing data were missing completely at random. With exclusion of the subjects having incomplete data, we had a total sample size of 828 in the *Multivariate-ODS* sample composed of 640 in the simple random sample and 188 in the SNHL sample. After examining the distributions of the hearing levels across three frequencies for each ear, we transformed the outcome variables on the natural log scale in order to exploit the normal properties. We therefore fitted the following linear model to the CPP *Multivariate-ODS* data,

$$\ln(Hearing_{ij}) = \beta_{0j} + \beta_{1j} PCB_i + \beta_{2j} SEX_{ij} + \beta_{3j} RACE_{ij} + \beta_{4j} EDUC_{ij} + \beta_{5j} SEI_{ij} + \epsilon_j \ , \tag{5.13}$$

where $\epsilon_j \sim N(0, \sigma_j^2)$, $i = 1, \ldots, 828$ and $j = 1$ representing the hearing level across three frequencies from the left ear and $j = 2$ from the right ear; $\rho = Corr(\epsilon_1, \epsilon_2)$. We assumed that $f(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\theta})$ is bivariate normal, where $\boldsymbol{\theta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \sigma_1^2, \sigma_2^2)$ and $\boldsymbol{\beta}_j = (\beta_{0j}, \ldots, \beta_{5j})$ and $j = 1, 2$. We estimated the parameters using the methods considered in the simulation studies: our proposed method, $\boldsymbol{\theta}_P$, and the competing method of $\boldsymbol{\theta}_R$.

Table 5.4 presented the results of the parameter estimates, the estimated standard errors and the approximate 95% confidence intervals calculated based on the asymptotic normal distributions for each method. Both analyses showed that the corresponding 95% confidence intervals for the PCB effect included 0. Thus, we would conclude that *in utero*

PCB exposure did not have a significant effect on hearing levels for both ears. Observing the confidence intervals for other confounding parameters for the left ear, the covariate RACE showed a significant effect at the nominal level of 0.05, agreed by three methods and the significance was concluded for both ears. It indicated that white children had negative impact on hearing loss; in other words, white children were more likely to have better hearing ability than black and other children. For another confounding variable SEX, $\widehat{\boldsymbol{\theta}}_P$ exhibited significance on the borderline for the left ear, suggesting that the hearing level for girls be lower than it for boys.

Although PCB was not significant, we could still see some efficiency gains from the results; the observed 95% confidence intervals for PCB provided by the proposed estimator $\widehat{\boldsymbol{\theta}}_P$ were narrower for both ears, compared with the CI obtained by $\widehat{\boldsymbol{\theta}}_R$; for example, on the left ear, $(-0.03, 0.03)$ for $\widehat{\boldsymbol{\theta}}_P$ versus $(-0.01, 0.06)$ for $\widehat{\boldsymbol{\theta}}_R$. Furthermore, it is clear to see that $\widehat{\boldsymbol{\theta}}_P$ resulted in substantially smaller standard errors for both ears than $\widehat{\boldsymbol{\theta}}_R$ and there were gains in efficiency of the proposed method.

In Table 5.5, we fitted the same model but only considering the univariate case by taking the grand mean over both ears across three frequencies. The point estimates for $\widehat{\boldsymbol{\theta}}_R$ in Table 5.4 were similar to those in Table 5.5. Note that the standard errors from $\boldsymbol{\theta}_P$ in Table 5.4 were relatively smaller compared with those from $\widehat{\boldsymbol{\theta}}_R$ in Table 5.5, which was agreed for both ears. From above analyses, we can clearly see that there are observable benefits by incorporating the supplemental data under the *Multivariate-ODS* design when estimating the model parameters.

Table 5.4: Results of modeling fitting for the CPP data using the *Multivariate-ODS* design.

		$\boldsymbol{\theta}_R$			$\boldsymbol{\theta}_P$		
		$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI	$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI
Left Ear	Int	1.83	0.19	$(1.47, 2.20)$	1.87	0.15	$(1.57, 2.17)$
	PCB	0.03	0.02	$(-0.01, 0.06)$	-0.00	0.02	$(-0.03, 0.03)$
	SEX	-0.11	0.07	$(-0.24, 0.02)$	-0.11	0.05	$(-0.21, -0.00)$
	RACE	-0.69	0.07	$(-0.83, -0.54)$	-0.31	0.06	$(-0.43, -0.19)$
	EDUC	0.00	0.02	$(-0.03, 0.04)$	0.01	0.01	$(-0.02, 0.03)$
	SEI	0.03	0.02	$(-0.01, 0.07)$	0.01	0.02	$(-0.03, 0.04)$
Right Ear	Int	1.65	0.18	$(1.30, 1.99)$	1.74	0.15	$(1.45, 2.04)$
	PCB	0.02	0.02	$(-0.02, 0.05)$	-0.01	0.01	$(-0.04, 0.02)$
	SEX	-0.00	0.06	$(-0.13, 0.12)$	-0.03	0.05	$(-0.13, 0.08)$
	RACE	-0.69	0.07	$(-0.82, -0.55)$	-0.34	0.06	$(-0.45, -0.22)$
	EDUC	0.01	0.02	$(-0.02, 0.05)$	0.01	0.01	$(-0.02, 0.04)$
	SEI	0.02	0.02	$(-0.02, 0.05)$	-0.00	0.02	$(-0.04, 0.03)$

Table 5.5: Results of modeling fitting for the CPP data using the univariate ODS design.

	$\boldsymbol{\theta}_R$			$\boldsymbol{\theta}_C$		
	$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI	$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI
Int	1.784	0.162	$(1.466, 2.102)$	2.201	0.174	$(1.860, 2.542)$
PCB	0.024	0.015	$(-0.005, 0.053)$	-0.003	0.017	$(-0.036, 0.030)$
SEX	-0.055	0.057	$(-0.167, 0.057)$	-0.087	0.062	$(-0.209, 0.035)$
RACE	-0.641	0.063	$(-0.764, -0.518)$	-0.394	0.069	$(-0.529, -0.259)$
EDUC	0.006	0.015	$(-0.023, 0.035)$	0.007	0.016	$(-0.024, 0.038)$
SEI	0.021	0.018	$(-0.014, 0.056)$	0.002	0.019	$(-0.035, 0.039)$

5.5 Discussion

Much research has been discussed for multivariate continuous data, of which is a common and important form; nevertheless, the methods accounting for the *Multivariate-ODS* design are lacking. Throughout previous sections, we have demonstrated the need for developing the statistical inferences on the *Multivariate-ODS* and proposed a semiparametric empirical likelihood method for multivariate continuous outcomes. The proposed estimator is semiparametric in nature that the underlying distributions of the covariates are modeled nonparametrically using the empirical likelihood methods. We have shown that the proposed estimator is consistent and asymptotically normally distributed and a consistent estimator for the asymptotic variance-covariance exists, by incorporating additional information into such *Multivariate-ODS* design process. We used simulated data generated from a standard linear regression model with Normal errors to examine the performance and the small-sample properties of our proposed estimator. Our limited simulation results indicated that the proposed estimator, $\boldsymbol{\theta}_P$, holds well for all the properties and is more efficient than $\boldsymbol{\theta}_R$, which only takes the simple random sample into consideration, and $\boldsymbol{\theta}_C$, the conditional estimator, using the complete *Multivariate-ODS* data but ignoring additional information in the supplemental sample. For the relative efficiency studies, we observed that $\boldsymbol{\theta}_P$ exhibits more efficiency gains than $\boldsymbol{\theta}_S$, using a simple random sample of the same size as the *Multivariate-ODS* from the underlying population, in terms of different correlation coefficients between the outcomes, the allocations of the cutpoints and the the supplemental fractions. We conclude that the

Multivariate-ODS design, combined with an appropriate analysis, can provide a cost-effective approach to further improve study efficiency, for a given sample size. Finally, we apply the proposed method to the Collaborative Perinatal Project data, where the researchers are interested in studying the association between a child's hearing loss and *in utero* exposure to PCBs as well as other covariates. The estimator obtained by $\boldsymbol{\theta}_P$ clearly gained more efficiency and as more precise than the other competing estimator, $\boldsymbol{\theta}_R$, although PCBs could not be concluded as a significant effect.

Our simulated studies also suggest that the greatest gain of efficiency takes place when the supplemental sampling fraction is in the region from 0.2 to 0.6, similar to the guidance suggested by Zhou et al. (2002) in using the ODS design concerning these issues under one continuous outcome variable. Further investigation for the sample size determination, the optimal sample allocations, the optimal correlation coefficient between the outcomes and power analyses aimed at multivariate outcomes under the *Multivariate-ODS* is required. We considered two-dimensional multivariate data in this dissertation; the future work may include the flexibility of incorporating the covariance structures for higher-dimensional data. Our proposed method can also be applied to the quantitative genetics studies, in which the quantitative trait is modeled as a continuous variable; more and more studies in order to limit the expenses on the DNA analysis are actually adopting the form of the ODS design. We believe that the proposed method can be a useful tool toward such studies.

5.6 Additional Simulation Results

Complete simulation studies were presented in this section. The simulation results were presented in Tables 5.6 through 5.25. The results in the tables were presented for three different combinations of $\boldsymbol{\beta}$, the correlation coefficients ρ, various cutpoints a, the sampling fractions γ, and sample sizes n, with three methods. Within each table, the sampling specifications and the covariate distribution are fixed. Tables 5.6 through 5.21 included the small sample properties of the proposed estimator and the competing estimators and Tables 5.22 - 5.25 presented the efficiencies of $\widehat{\boldsymbol{\theta}}_S$ versus $\widehat{\boldsymbol{\theta}}_P$ based on the models for Tables 5.6 - 5.21. The results were comparable to those discussed in the previous section.

To investigate the effect of changing the supplemental sampling fractions on the improvement of the *Multivariate-ODS* design over other simple random sample designs, we conducted several simulation experiments using the same simulation models used in Tables 5.14 and 5.16 but with the cutpoint $a = 80\%$. Figures 5.1 and 5.2 presented the relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ over $\widehat{\boldsymbol{\theta}}_R$. Clearly, the efficiency gains of the *Multivariate-ODS* design over the simple random sample design increased with the supplemental sampling fractions, agreed by both sample size considerations, and $\widehat{\boldsymbol{\theta}}_P$ was consistently more efficient than $\widehat{\boldsymbol{\theta}}_R$ regardless of the sampling fractions. Although the efficiency gains increased as the supplemental sample size increased, it was not practical in reality since it may not be easy to have enough individuals in the extreme tails. We suggested the possible remedy for an appropriate proportion of the supplemental sample to be in the region from 0.3 to 0.6. Figures 5.3 through 5.6 illustrated the standard errors for $\widehat{\boldsymbol{\theta}}_P$, and the

relative efficiency of the *Multivariate-ODS* design to a simple random sample of the same sample size across various supplemental sampling fractions γ. The increase in the relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ was not monotone over the fractions although $\widehat{\boldsymbol{\theta}}_P$ was more efficient than $\widehat{\boldsymbol{\theta}}_S$ with most of the sampling fractions. For both sample sizes, we observed the most efficiency gain at $\gamma = 0.3$, which β_1 and β_2 both agreed. As γ was more than 60%, there was a decrease in the relative efficiency. The results suggested that a great efficiency gain can be achieved when γ was between 0.3 and 0.6.

Table 5.6: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.165$ (80^{th} percentile) and 1.936 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.002	0.080	0.079	0.947	0.005	0.079	0.079	0.945
		$\boldsymbol{\theta}_C$	0.003	0.076	0.076	0.944	0.003	0.075	0.075	0.958
		$\boldsymbol{\theta}_P$	0.002	0.067	0.067	0.949	0.002	0.066	0.067	0.956
	50%	$\boldsymbol{\theta}_R$	0.000	0.104	0.100	0.943	−0.002	0.102	0.100	0.946
		$\boldsymbol{\theta}_C$	−0.002	0.087	0.087	0.942	−0.002	0.087	0.086	0.943
		$\boldsymbol{\theta}_P$	0.001	0.070	0.068	0.942	0.000	0.069	0.068	0.955
90%	20%	$\boldsymbol{\theta}_R$	−0.001	0.080	0.079	0.938	0.000	0.081	0.079	0.949
		$\boldsymbol{\theta}_C$	−0.001	0.076	0.076	0.945	−0.001	0.077	0.076	0.955
		$\boldsymbol{\theta}_P$	−0.001	0.064	0.064	0.953	0.003	0.064	0.064	0.952
	50%	$\boldsymbol{\theta}_R$	−0.001	0.105	0.101	0.938	−0.002	0.103	0.101	0.949
		$\boldsymbol{\theta}_C$	−0.003	0.091	0.088	0.942	−0.004	0.088	0.088	0.952
		$\boldsymbol{\theta}_P$	−0.002	0.065	0.063	0.939	−0.004	0.065	0.064	0.945

Table 5.7: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.318$ (80^{th} percentile) and 2.183 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.001	0.078	0.079	0.946	0.000	0.080	0.079	0.951
		$\boldsymbol{\theta}_C$	−0.001	0.074	0.077	0.951	0.001	0.077	0.077	0.950
		$\boldsymbol{\theta}_P$	0.000	0.066	0.066	0.946	0.001	0.068	0.066	0.938
	50%	$\boldsymbol{\theta}_R$	−0.001	0.096	0.100	0.956	−0.002	0.097	0.100	0.948
		$\boldsymbol{\theta}_C$	−0.001	0.090	0.089	0.942	−0.002	0.091	0.089	0.936
		$\boldsymbol{\theta}_P$	0.001	0.066	0.068	0.951	−0.001	0.066	0.067	0.945
90%	20%	$\boldsymbol{\theta}_R$	0.000	0.079	0.079	0.949	−0.001	0.080	0.079	0.946
		$\boldsymbol{\theta}_C$	0.000	0.076	0.077	0.956	−0.001	0.077	0.077	0.954
		$\boldsymbol{\theta}_P$	0.002	0.061	0.062	0.952	0.001	0.062	0.062	0.944
	50%	$\boldsymbol{\theta}_R$	−0.002	0.101	0.100	0.956	0.000	0.098	0.100	0.947
		$\boldsymbol{\theta}_C$	0.001	0.091	0.092	0.959	0.002	0.088	0.092	0.958
		$\boldsymbol{\theta}_P$	−0.002	0.061	0.061	0.951	−0.001	0.060	0.061	0.955

Table 5.8: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.165$ (80^{th} percentile) and 1.936 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.000	0.039	0.039	0.950	−0.001	0.040	0.040	0.942
		$\boldsymbol{\theta}_C$	0.000	0.037	0.038	0.962	0.000	0.039	0.038	0.945
		$\boldsymbol{\theta}_P$	0.001	0.034	0.033	0.951	0.000	0.035	0.033	0.944
	50%	$\boldsymbol{\theta}_R$	−0.004	0.050	0.050	0.952	−0.003	0.051	0.050	0.941
		$\boldsymbol{\theta}_C$	−0.003	0.042	0.043	0.654	−0.003	0.044	0.043	0.948
		$\boldsymbol{\theta}_P$	−0.002	0.033	0.034	0.958	−0.001	0.036	0.034	0.940
90%	20%	$\boldsymbol{\theta}_R$	0.001	0.039	0.040	0.959	0.000	0.038	0.040	0.958
		$\boldsymbol{\theta}_C$	0.002	0.037	0.038	0.961	0.000	0.037	0.038	0.959
		$\boldsymbol{\theta}_P$	0.002	0.031	0.032	0.952	0.000	0.031	0.032	0.958
	50%	$\boldsymbol{\theta}_R$	0.001	0.050	0.050	0.939	0.000	0.050	0.050	0.952
		$\boldsymbol{\theta}_C$	0.001	0.044	0.044	0.937	0.000	0.044	0.044	0.951
		$\boldsymbol{\theta}_P$	0.001	0.031	0.032	0.954	0.000	0.032	0.032	0.947

Table 5.9: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.318$ (80^{th} percentile) and 2.183 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.001	0.040	0.040	0.948	−0.001	0.039	0.040	0.961
		$\boldsymbol{\theta}_C$	−0.001	0.038	0.038	0.946	−0.001	0.038	0.038	0.950
		$\boldsymbol{\theta}_P$	−0.001	0.034	0.033	0.945	−0.001	0.033	0.033	0.952
	50%	$\boldsymbol{\theta}_R$	0.001	0.050	0.050	0.945	0.001	0.052	0.050	0.947
		$\boldsymbol{\theta}_C$	0.001	0.045	0.045	0.948	0.002	0.045	0.045	0.943
		$\boldsymbol{\theta}_P$	0.002	0.035	0.034	0.943	0.003	0.035	0.034	0.941
90%	20%	$\boldsymbol{\theta}_R$	0.001	0.040	0.040	0.950	0.001	0.040	0.040	0.944
		$\boldsymbol{\theta}_C$	0.001	0.039	0.039	0.950	0.001	0.039	0.039	0.947
		$\boldsymbol{\theta}_P$	0.000	0.031	0.031	0.953	0.000	0.031	0.031	0.957
	50%	$\boldsymbol{\theta}_R$	0.000	0.051	0.050	0.948	−0.002	0.051	0.050	0.946
		$\boldsymbol{\theta}_C$	−0.001	0.046	0.046	0.949	−0.002	0.046	0.046	0.942
		$\boldsymbol{\theta}_P$	0.000	0.032	0.031	0.934	0.000	0.031	0.031	0.953

Table 5.10: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.231$ (80^{th} percentile) and 2.030 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.005	0.083	0.079	0.949	0.501	0.079	0.079	0.956
		$\boldsymbol{\theta}_C$	0.005	0.078	0.075	0.941	0.501	0.076	0.075	0.950
		$\boldsymbol{\theta}_P$	0.007	0.070	0.068	0.935	0.502	0.067	0.068	0.946
	50%	$\boldsymbol{\theta}_R$	0.001	0.102	0.100	0.949	0.506	0.103	0.100	0.938
		$\boldsymbol{\theta}_C$	0.003	0.087	0.085	0.945	0.505	0.085	0.085	0.955
		$\boldsymbol{\theta}_P$	0.002	0.070	0.069	0.938	0.504	0.071	0.069	0.940
90%	20%	$\boldsymbol{\theta}_R$	0.003	0.077	0.079	0.957	0.502	0.080	0.079	0.943
		$\boldsymbol{\theta}_C$	0.004	0.073	0.075	0.954	0.504	0.076	0.076	0.950
		$\boldsymbol{\theta}_P$	0.003	0.062	0.065	0.958	0.503	0.066	0.066	0.951
	50%	$\boldsymbol{\theta}_R$	−0.001	0.099	0.100	0.945	0.498	0.103	0.100	0.941
		$\boldsymbol{\theta}_C$	−0.001	0.088	0.087	0.943	0.499	0.088	0.086	0.940
		$\boldsymbol{\theta}_P$	0.001	0.066	0.065	0.943	0.501	0.067	0.065	0.940

Table 5.11: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.390$ (80^{th} percentile) and 2.262 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.004	0.079	0.079	0.947	0.503	0.080	0.079	0.940
		$\boldsymbol{\theta}_C$	0.004	0.077	0.076	0.952	0.503	0.076	0.076	0.948
		$\boldsymbol{\theta}_P$	0.003	0.066	0.067	0.952	0.503	0.065	0.067	0.950
	50%	$\boldsymbol{\theta}_R$	−0.001	0.096	0.100	0.950	0.497	0.100	0.100	0.953
		$\boldsymbol{\theta}_C$	0.001	0.084	0.088	0.958	0.500	0.086	0.088	0.950
		$\boldsymbol{\theta}_P$	0.000	0.068	0.068	0.952	0.499	0.069	0.068	0.950
90%	20%	$\boldsymbol{\theta}_R$	0.000	0.079	0.079	0.945	0.500	0.081	0.079	0.958
		$\boldsymbol{\theta}_C$	0.001	0.078	0.077	0.944	0.501	0.079	0.077	0.944
		$\boldsymbol{\theta}_P$	−0.001	0.065	0.064	0.938	0.499	0.066	0.064	0.934
	50%	$\boldsymbol{\theta}_R$	−0.003	0.101	0.100	0.944	0.497	0.103	0.100	0.943
		$\boldsymbol{\theta}_C$	−0.003	0.091	0.090	0.948	0.499	0.092	0.090	0.945
		$\boldsymbol{\theta}_P$	−0.003	0.063	0.063	0.949	0.498	0.063	0.063	0.946

Table 5.12: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.231$ (80^{th} percentile) and 2.030 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.001	0.040	0.040	0.947	0.499	0.039	0.040	0.963
		$\boldsymbol{\theta}_C$	−0.001	0.038	0.038	0.951	0.499	0.037	0.038	0.957
		$\boldsymbol{\theta}_P$	0.000	0.035	0.034	0.944	0.500	0.033	0.034	0.957
	50%	$\boldsymbol{\theta}_R$	0.001	0.049	0.050	0.954	0.500	0.049	0.050	0.958
		$\boldsymbol{\theta}_C$	0.000	0.041	0.042	0.959	0.499	0.041	0.042	0.957
		$\boldsymbol{\theta}_P$	0.001	0.034	0.034	0.958	0.500	0.034	0.034	0.954
90%	20%	$\boldsymbol{\theta}_R$	−0.001	0.040	0.040	0.945	0.498	0.040	0.040	0.954
		$\boldsymbol{\theta}_C$	−0.001	0.038	0.038	0.946	0.498	0.038	0.038	0.949
		$\boldsymbol{\theta}_P$	−0.001	0.033	0.033	0.947	0.498	0.033	0.033	0.949
	50%	$\boldsymbol{\theta}_R$	−0.001	0.050	0.050	0.951	0.500	0.051	0.050	0.949
		$\boldsymbol{\theta}_C$	0.000	0.043	0.043	0.947	0.501	0.045	0.043	0.943
		$\boldsymbol{\theta}_P$	−0.001	0.032	0.032	0.950	0.500	0.035	0.033	0.930

Table 5.13: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.390$ (80^{th} percentile) and 2.262 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	0.000	0.039	0.040	0.953	0.501	0.038	0.040	0.954
		$\boldsymbol{\theta}_C$	0.000	0.038	0.038	0.953	0.501	0.037	0.038	0.960
		$\boldsymbol{\theta}_P$	0.000	0.034	0.034	0.947	0.501	0.033	0.033	0.955
	50%	$\boldsymbol{\theta}_R$	0.000	0.050	0.050	0.953	0.501	0.050	0.050	0.950
		$\boldsymbol{\theta}_C$	0.001	0.044	0.044	0.939	0.501	0.044	0.044	0.949
		$\boldsymbol{\theta}_P$	0.002	0.033	0.034	0.956	0.501	0.034	0.034	0.952
90%	20%	$\boldsymbol{\theta}_R$	0.000	0.039	0.039	0.945	0.501	0.039	0.039	0.954
		$\boldsymbol{\theta}_C$	0.000	0.038	0.038	0.943	0.501	0.038	0.038	0.947
		$\boldsymbol{\theta}_P$	0.000	0.031	0.032	0.951	0.501	0.032	0.032	0.955
	50%	$\boldsymbol{\theta}_R$	0.003	0.052	0.050	0.945	0.502	0.051	0.050	0.944
		$\boldsymbol{\theta}_C$	0.003	0.045	0.045	0.954	0.502	0.044	0.045	0.955
		$\boldsymbol{\theta}_P$	0.002	0.032	0.032	0.945	0.501	0.031	0.032	0.955

Table 5.14: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.175$ (80^{th} percentile) and 1.958 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.501	0.081	0.080	0.948	0.692	0.081	0.080	0.952
		$\boldsymbol{\theta}_C$	−0.499	0.079	0.076	0.945	0.693	0.077	0.076	0.952
		$\boldsymbol{\theta}_P$	−0.499	0.068	0.068	0.953	0.693	0.067	0.068	0.951
	50%	$\boldsymbol{\theta}_R$	−0.500	0.105	0.101	0.941	0.690	0.099	0.101	0.959
		$\boldsymbol{\theta}_C$	−0.500	0.087	0.086	0.949	0.689	0.086	0.086	0.955
		$\boldsymbol{\theta}_P$	−0.500	0.068	0.069	0.951	0.690	0.069	0.069	0.942
90%	20%	$\boldsymbol{\theta}_R$	−0.495	0.080	0.079	0.955	0.698	0.082	0.079	0.937
		$\boldsymbol{\theta}_C$	−0.494	0.077	0.076	0.948	0.696	0.079	0.076	0.942
		$\boldsymbol{\theta}_P$	−0.497	0.065	0.064	0.940	0.693	0.067	0.064	0.944
	50%	$\boldsymbol{\theta}_R$	−0.500	0.102	0.102	0.954	0.691	0.102	0.102	0.949
		$\boldsymbol{\theta}_C$	−0.500	0.090	0.088	0.946	0.694	0.088	0.088	0.948
		$\boldsymbol{\theta}_P$	−0.500	0.065	0.064	0.941	0.694	0.063	0.064	0.951

Table 5.15: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.335$ (80^{th} percentile) and 2.192 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.499	0.075	0.080	0.961	0.693	0.078	0.080	0.958
		$\boldsymbol{\theta}_C$	−0.499	0.073	0.077	0.955	0.693	0.075	0.077	0.958
		$\boldsymbol{\theta}_P$	−0.500	0.064	0.067	0.957	0.692	0.066	0.067	0.962
	50%	$\boldsymbol{\theta}_R$	−0.497	0.099	0.101	0.956	0.694	0.101	0.101	0.944
		$\boldsymbol{\theta}_C$	−0.499	0.088	0.090	0.948	0.694	0.089	0.090	0.955
		$\boldsymbol{\theta}_P$	−0.499	0.065	0.068	0.968	0.694	0.065	0.068	0.960
90%	20%	$\boldsymbol{\theta}_R$	−0.498	0.080	0.080	0.954	0.696	0.081	0.080	0.946
		$\boldsymbol{\theta}_C$	−0.499	0.078	0.077	0.950	0.695	0.080	0.077	0.943
		$\boldsymbol{\theta}_P$	−0.501	0.064	0.063	0.940	0.693	0.065	0.063	0.941
	50%	$\boldsymbol{\theta}_R$	−0.502	0.100	0.101	0.952	0.682	0.102	0.101	0.943
		$\boldsymbol{\theta}_C$	−0.500	0.090	0.092	0.954	0.683	0.090	0.092	0.962
		$\boldsymbol{\theta}_P$	−0.496	0.061	0.062	0.953	0.697	0.062	0.062	0.950

Table 5.16: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $a = 1.175$ (80^{th} percentile) and 1.958 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.501	0.040	0.040	0.949	0.695	0.040	0.040	0.955
		$\boldsymbol{\theta}_C$	−0.501	0.038	0.038	0.959	0.694	0.038	0.038	0.948
		$\boldsymbol{\theta}_P$	−0.501	0.034	0.034	0.952	0.695	0.033	0.034	0.941
	50%	$\boldsymbol{\theta}_R$	−0.501	0.050	0.050	0.950	0.692	0.048	0.050	0.959
		$\boldsymbol{\theta}_C$	−0.500	0.042	0.043	0.956	0.693	0.042	0.043	0.952
		$\boldsymbol{\theta}_P$	−0.500	0.033	0.034	0.946	0.693	0.034	0.034	0.952
90%	20%	$\boldsymbol{\theta}_R$	−0.498	0.040	0.040	0.938	0.693	0.041	0.040	0.944
		$\boldsymbol{\theta}_C$	−0.498	0.038	0.038	0.947	0.694	0.039	0.038	0.938
		$\boldsymbol{\theta}_P$	−0.498	0.032	0.032	0.948	0.694	0.032	0.032	0.952
	50%	$\boldsymbol{\theta}_R$	−0.501	0.051	0.050	0.941	0.693	0.050	0.050	0.945
		$\boldsymbol{\theta}_C$	−0.500	0.044	0.044	0.946	0.694	0.044	0.044	0.944
		$\boldsymbol{\theta}_P$	−0.500	0.032	0.032	0.948	0.695	0.032	0.032	0.946

Table 5.17: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, $a = 1.335$ (80^{th} percentile) and 2.192 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.500	0.039	0.040	0.953	0.693	0.039	0.040	0.946
		$\boldsymbol{\theta}_C$	−0.500	0.037	0.038	0.952	0.693	0.038	0.038	0.947
		$\boldsymbol{\theta}_P$	−0.500	0.033	0.033	0.949	0.693	0.034	0.033	0.947
	50%	$\boldsymbol{\theta}_R$	−0.499	0.049	0.050	0.959	0.694	0.047	0.050	0.966
		$\boldsymbol{\theta}_C$	−0.498	0.043	0.045	0.960	0.695	0.042	0.045	0.962
		$\boldsymbol{\theta}_P$	−0.499	0.034	0.034	0.953	0.694	0.033	0.034	0.950
90%	20%	$\boldsymbol{\theta}_R$	−0.500	0.040	0.040	0.946	0.694	0.040	0.040	0.945
		$\boldsymbol{\theta}_C$	−0.500	0.039	0.038	0.943	0.694	0.039	0.038	0.946
		$\boldsymbol{\theta}_P$	−0.499	0.030	0.031	0.955	0.695	0.030	0.031	0.953
	50%	$\boldsymbol{\theta}_R$	−0.501	0.048	0.050	0.954	0.693	0.048	0.050	0.959
		$\boldsymbol{\theta}_C$	−0.500	0.044	0.046	0.962	0.694	0.045	0.046	0.952
		$\boldsymbol{\theta}_P$	−0.500	0.030	0.031	0.953	0.693	0.031	0.031	0.950

Table 5.18: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 0.5$, $\rho = 1.5$, $a = 0.451$ (80^{th} percentile) and 0.846 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.500	0.041	0.039	0.944	0.693	0.040	0.039	0.944
		$\boldsymbol{\theta}_C$	−0.500	0.038	0.037	0.937	0.694	0.038	0.038	0.941
		$\boldsymbol{\theta}_P$	−0.500	0.035	0.034	0.940	0.694	0.034	0.034	0.950
	50%	$\boldsymbol{\theta}_R$	−0.500	0.051	0.050	0.948	0.694	0.051	0.050	0.943
		$\boldsymbol{\theta}_C$	−0.499	0.045	0.043	0.940	0.693	0.043	0.043	0.949
		$\boldsymbol{\theta}_P$	−0.499	0.035	0.034	0.939	0.693	0.035	0.034	0.949
90%	20%	$\boldsymbol{\theta}_R$	−0.500	0.042	0.040	0.944	0.691	0.040	0.040	0.931
		$\boldsymbol{\theta}_C$	−0.501	0.039	0.038	0.947	0.692	0.038	0.038	0.941
		$\boldsymbol{\theta}_P$	−0.501	0.033	0.032	0.942	0.692	0.033	0.033	0.945
	50%	$\boldsymbol{\theta}_R$	−0.502	0.050	0.050	0.950	0.691	0.052	0.050	0.943
		$\boldsymbol{\theta}_C$	−0.502	0.043	0.044	0.954	0.691	0.044	0.044	0.944
		$\boldsymbol{\theta}_P$	−0.502	0.033	0.032	0.937	0.691	0.034	0.032	0.933

Table 5.19: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.85$, $a = 0.532$ (80^{th} percentile) and 0.965 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.501	0.040	0.039	0.957	0.692	0.039	0.040	0.952
		$\boldsymbol{\theta}_C$	−0.501	0.038	0.038	0.956	0.692	0.038	0.038	0.954
		$\boldsymbol{\theta}_P$	−0.501	0.034	0.033	0.953	0.693	0.033	0.033	0.950
	50%	$\boldsymbol{\theta}_R$	−0.500	0.052	0.050	0.950	0.694	0.052	0.050	0.943
		$\boldsymbol{\theta}_C$	−0.499	0.046	0.044	0.948	0.695	0.045	0.045	0.951
		$\boldsymbol{\theta}_P$	−0.498	0.035	0.034	0.943	0.695	0.035	0.034	0.936
90%	20%	$\boldsymbol{\theta}_R$	−0.490	0.041	0.040	0.942	0.695	0.040	0.040	0.947
		$\boldsymbol{\theta}_C$	−0.499	0.040	0.038	0.944	0.695	0.038	0.038	0.950
		$\boldsymbol{\theta}_P$	−0.499	0.032	0.032	0.948	0.695	0.031	0.032	0.960
	50%	$\boldsymbol{\theta}_R$	−0.500	0.050	0.050	0.942	0.682	0.050	0.050	0.947
		$\boldsymbol{\theta}_C$	−0.499	0.045	0.045	0.954	0.693	0.046	0.045	0.947
		$\boldsymbol{\theta}_P$	−0.500	0.031	0.031	0.958	0.693	0.032	0.031	0.942

Table 5.20: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.5$, $a = 0.451$ (80^{th} percentile) and 0.846 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.502	0.020	0.020	0.932	0.692	0.020	0.020	0.944
		$\boldsymbol{\theta}_C$	−0.501	0.019	0.019	0.938	0.692	0.019	0.019	0.942
		$\boldsymbol{\theta}_P$	−0.501	0.017	0.017	0.947	0.692	0.016	0.017	0.962
	50%	$\boldsymbol{\theta}_R$	−0.500	0.025	0.025	0.950	0.694	0.026	0.025	0.944
		$\boldsymbol{\theta}_C$	−0.500	0.021	0.021	0.954	0.694	0.022	0.021	0.943
		$\boldsymbol{\theta}_P$	−0.500	0.016	0.017	0.961	0.693	0.017	0.017	0.947
90%	20%	$\boldsymbol{\theta}_R$	−0.500	0.021	0.020	0.934	0.693	0.020	0.020	0.945
		$\boldsymbol{\theta}_C$	−0.500	0.020	0.019	0.931	0.694	0.019	0.019	0.950
		$\boldsymbol{\theta}_P$	−0.500	0.016	0.016	0.944	0.694	0.016	0.016	0.964
	50%	$\boldsymbol{\theta}_R$	−0.500	0.025	0.025	0.943	0.693	0.026	0.025	0.946
		$\boldsymbol{\theta}_C$	−0.500	0.021	0.022	0.960	0.694	0.022	0.022	0.945
		$\boldsymbol{\theta}_P$	−0.500	0.016	0.016	0.956	0.693	0.016	0.016	0.939

Table 5.21: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.85$, $a = 0.532$ (80^{th} percentile) and 0.965 (90^{th} percentile) and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
a	γ	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
80%	20%	$\boldsymbol{\theta}_R$	−0.501	0.020	0.020	0.954	0.692	0.020	0.020	0.943
		$\boldsymbol{\theta}_C$	−0.501	0.019	0.019	0.956	0.692	0.020	0.019	0.940
		$\boldsymbol{\theta}_P$	−0.500	0.016	0.017	0.960	0.692	0.017	0.017	0.952
	50%	$\boldsymbol{\theta}_R$	−0.500	0.025	0.025	0.950	0.693	0.025	0.025	0.947
		$\boldsymbol{\theta}_C$	−0.500	0.022	0.022	0.943	0.693	0.022	0.022	0.947
		$\boldsymbol{\theta}_P$	−0.500	0.018	0.017	0.941	0.693	0.018	0.017	0.940
90%	20%	$\boldsymbol{\theta}_R$	−0.500	0.020	0.020	0.953	0.694	0.020	0.020	0.950
		$\boldsymbol{\theta}_C$	−0.500	0.019	0.019	0.947	0.694	0.019	0.019	0.948
		$\boldsymbol{\theta}_P$	−0.500	0.016	0.016	0.953	0.694	0.016	0.016	0.943
	50%	$\boldsymbol{\theta}_R$	−0.500	0.025	0.025	0.948	0.694	0.025	0.025	0.948
		$\boldsymbol{\theta}_C$	−0.500	0.023	0.023	0.946	0.694	0.023	0.023	0.939
		$\boldsymbol{\theta}_P$	−0.500	0.015	0.016	0.964	0.693	0.016	0.016	0.956

Table 5.22: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $X_1 = X_2 \sim N(0,1)$

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	80%	20%	1.21	1.16	1.18	1.06
		50%	1.04	1.08	1.11	0.95
	90%	20%	1.22	1.31	1.24	1.35
		50%	1.17	1.20	1.21	1.26
0.85	80%	20%	1.07	1.11	1.12	1.17
		50%	1.05	1.08	0.98	0.94
	90%	20%	1.32	1.25	1.36	1.28
		50%	1.35	1.35	1.35	1.46

Table 5.23: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$ and $X_1 = X_2 \sim N(0,1)$.

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	80%	20%	0.96	1.09	0.95	1.11
		50%	1.06	1.03	1.16	1.10
	90%	20%	1.31	1.10	1.19	1.12
		50%	1.22	1.23	1.15	1.11
0.85	80%	20%	1.17	1.20	1.04	1.15
		50%	1.10	1.09	1.08	1.08
	90%	20%	1.25	1.18	1.26	1.29
		50%	1.24	1.32	1.29	1.39

Table 5.24: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$ and $X_1 = X_2 \sim N(0,1)$.

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	80%	20%	1.08	1.08	1.02	1.07
		50%	1.17	1.10	1.25	1.15
	90%	20%	1.24	1.11	1.32	1.29
		50%	1.27	1.33	1.22	1.29
0.85	80%	20%	1.30	1.22	1.12	1.08
		50%	1.33	1.25	1.12	1.16
	90%	20%	1.32	1.25	1.44	1.35
		50%	1.38	1.29	1.43	1.36

Table 5.25: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$ and $X_1 = X_2 \sim N(0,1)$.

			$n = 200$		$n = 800$	
ρ	a	γ	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	80%	20%	1.02	1.08	1.21	1.26
		50%	1.03	1.03	1.18	1.05
	90%	20%	1.10	1.17	1.07	1.24
		50%	1.15	1.06	1.30	1.14
0.85	80%	20%	1.20	1.23	1.15	1.12
		50%	1.10	1.07	1.05	1.10
	90%	20%	1.17	1.33	1.29	1.25
		50%	1.26	1.19	1.34	1.25

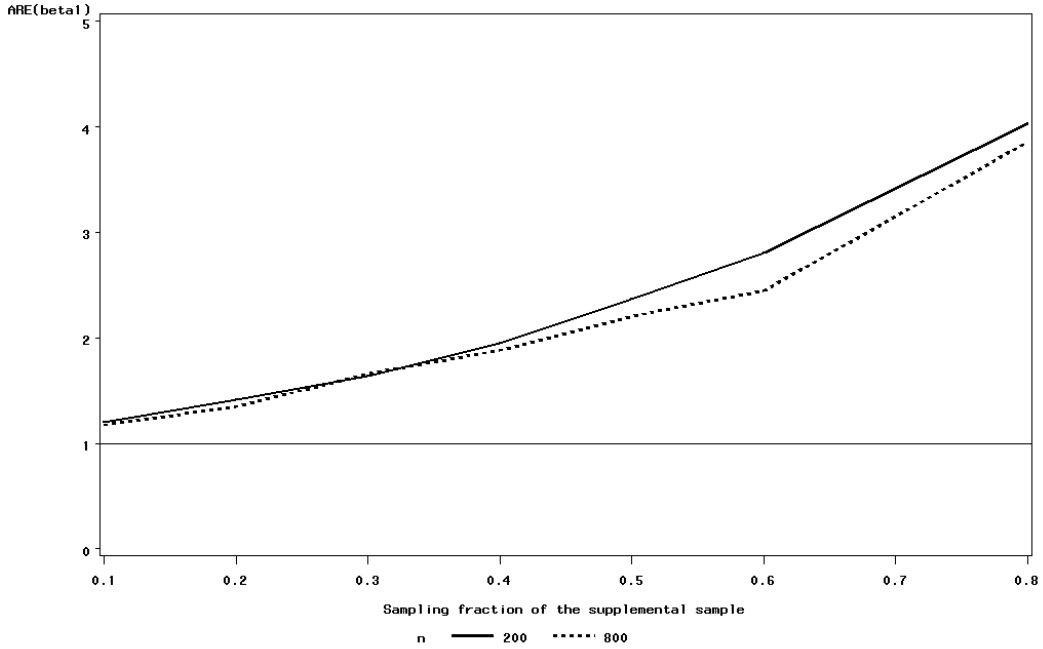

Figure 5.1: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_R$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample, under the models in Tables 5.14 and 5.16 with $a = 80\%$.

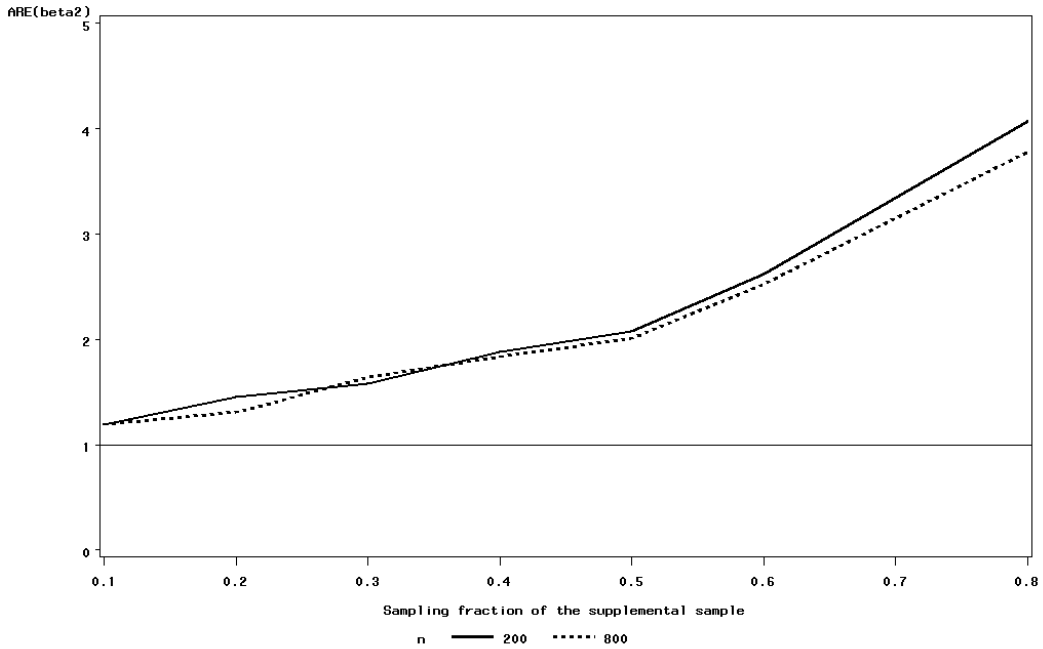

Figure 5.2: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_R$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample, under the models in Tables 5.14 and 5.16 with $a = 80\%$.

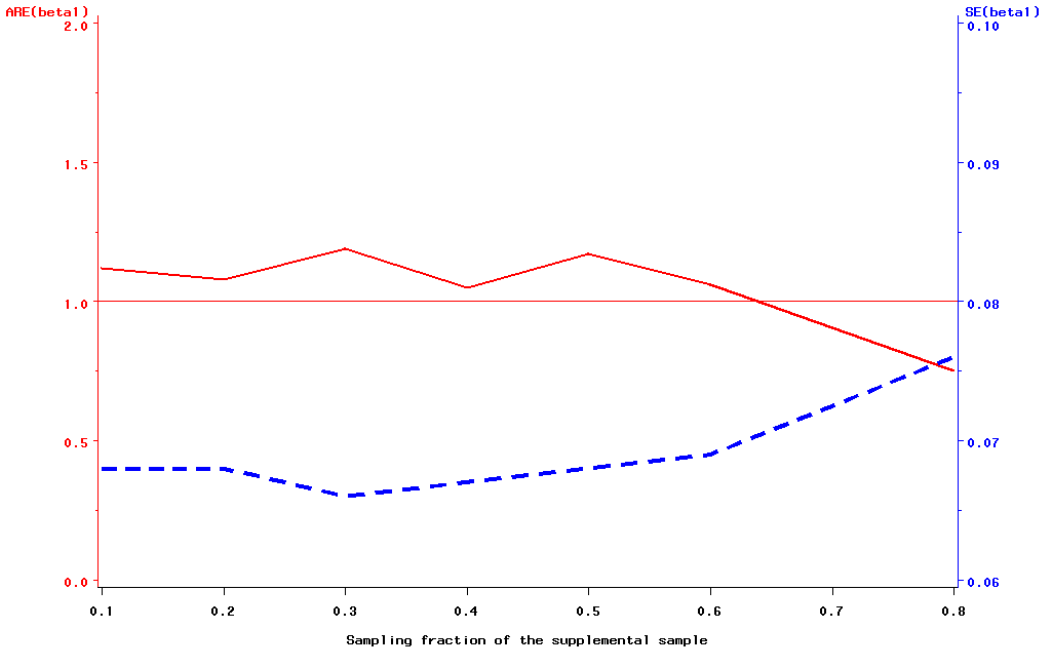

Figure 5.3: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample, under the model in Table 5.14 with $a = 80\%$.

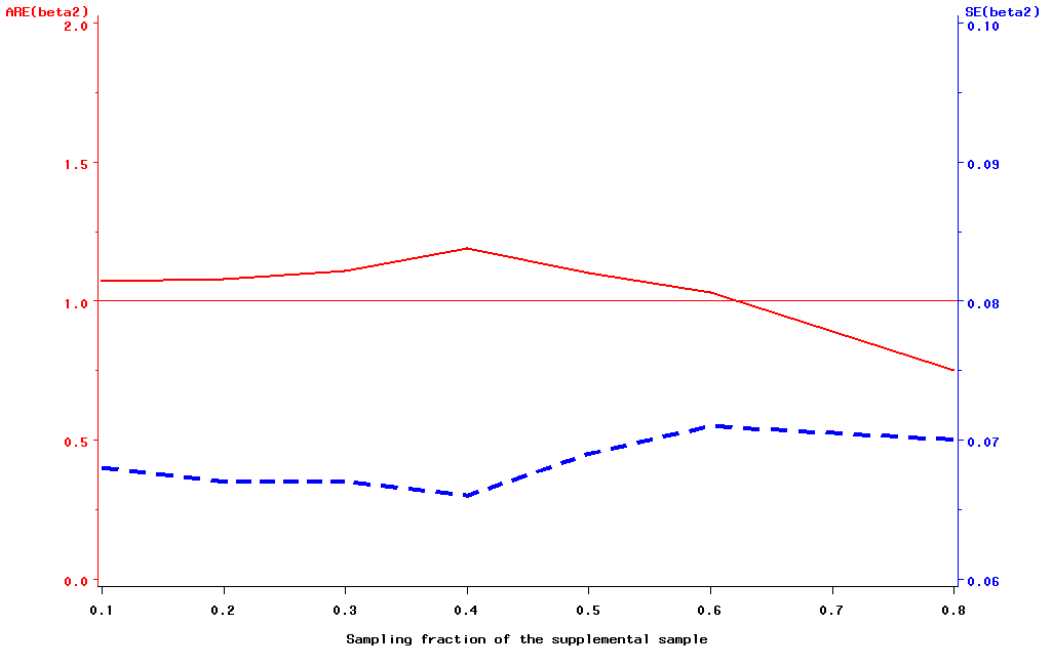

Figure 5.4: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample, under the model in Table 5.14 with $a = 80\%$.

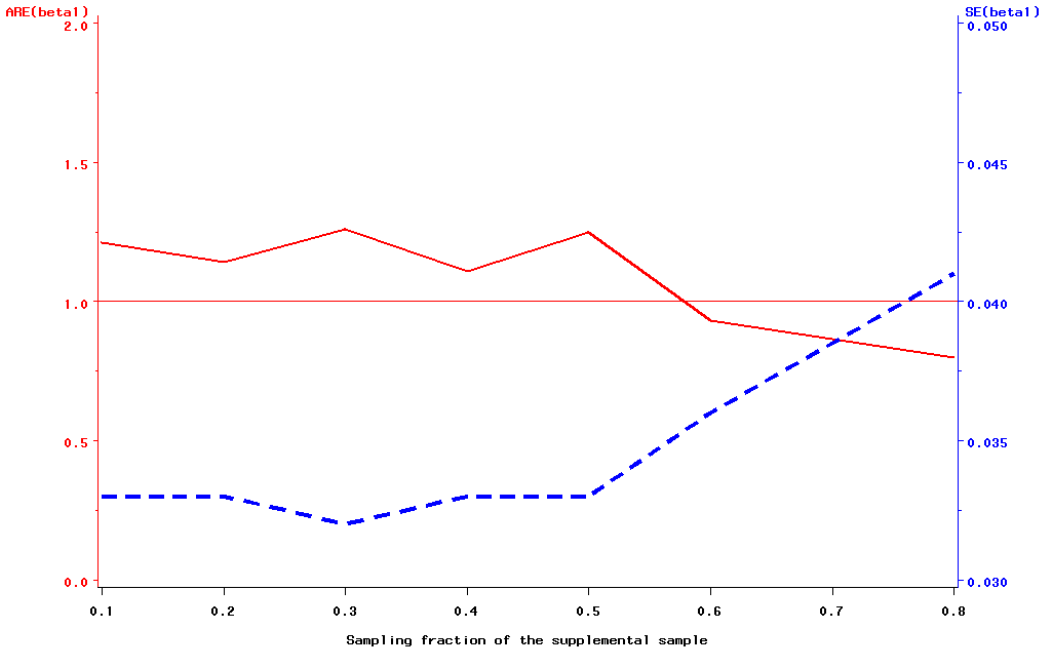

Figure 5.5: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample under the model in Table 5.16 with $a = 80\%$.

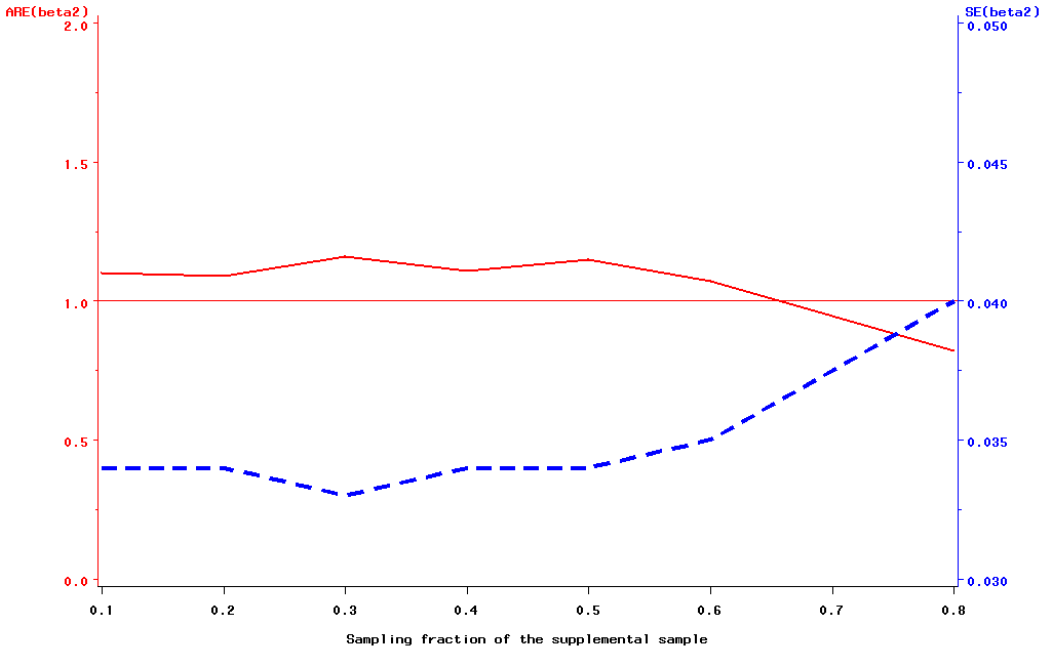

Figure 5.6: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample under the model in Table 5.16 with $a = 80\%$.

APPENDIX: ASYMPTOTIC RESULTS

For any function $h(\boldsymbol{Y}, \boldsymbol{X})$, $\mathrm{E}\Big[h(\boldsymbol{Y}, \boldsymbol{X})\Big]$ denotes expectation conditional on $\{\sum \boldsymbol{Y} < a\}$,

$$\mathrm{E}\Big[h(\boldsymbol{Y}, \boldsymbol{X})\Big] = \int_{\mathbb{X}} \frac{1}{\pi^0} \int \ldots \int_{\sum \boldsymbol{Y} < a} h(\boldsymbol{y}, \boldsymbol{x}) f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}^0)\ dy_1 \ldots dy_k\ dG_{\boldsymbol{X}}(\boldsymbol{x})\ .$$

We assume the following regularity conditions:

A1. As $n \to \infty$, $\dfrac{n_1}{n} \to \gamma > 0$ and $\dfrac{n_0}{n} \to 1 - \gamma > 0$, where γ represents the supplemental sampling fraction.

A2. The parameter space, $\boldsymbol{\Theta}$, is a compact subset of $\mathbb{R}^p$; $\boldsymbol{\theta}^0$ lies in the interior of $\boldsymbol{\Theta}$; the covariate space, $\mathbb{X}$, is a compact subset of $\mathbb{R}^q$, for some $q \geq 1$.

A3. $f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})$ is continuous in both $\boldsymbol{y}$ and $\boldsymbol{\theta}$ and is strictly positive for all $\boldsymbol{y} \in \mathbb{Y}$, $\boldsymbol{x} \in \mathbb{X}$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Furthermore, the partial derivatives, $\partial f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})/\partial\theta_i$ and $\partial^2 f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})/\partial\theta_i\partial\theta_j$, for $i,\ j = 1, \ldots, p$, exist and are continuous for all $\boldsymbol{y} \in \mathbb{Y}$, $\boldsymbol{x} \in \mathbb{X}$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

A4. Interchanges of differentiation and integration of $f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})$ are valid for the first and second partial derivatives with respect to $\boldsymbol{\theta}$.

A5. The expected value matrix, $\mathrm{E}\left[-\dfrac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]$, is finite and positive definite at $\boldsymbol{\theta}^0$.

A6. There exists a $\delta > 0$ such that for the set $A = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}^0| \leq \delta\}$,

$$\mathrm{E}\left[\sup_A \left|\frac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right|\right] < \infty,$$

for $i,\ j = 1, ..., p$.

A7. The derivatives, $\dfrac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\theta_j}$, $j = 1, \ldots, p$, are linearly independent. That is, suppose $\boldsymbol{t}$ is any $(p \times 1)$ vector such that

$$\sum_{j=1}^{p} t_j \frac{\partial P_0(\boldsymbol{x};\boldsymbol{\theta}^0)}{\theta_j} = 0$$

for almost all $\boldsymbol{x} \in \mathbb{X}$ if $\boldsymbol{t} = \mathbf{0}$.

Proof of Theorem 1 (Consistency) Using Assumption A1 and the Law of Large Numbers, we have

$$\frac{1}{n}\frac{\partial l_S(\boldsymbol{\phi}_{SM})}{\partial \boldsymbol{\theta}} \xrightarrow{\ p\ } \frac{\partial \widetilde{l_S}(\boldsymbol{\phi}_{SM})}{\partial \boldsymbol{\theta}} \ ,$$

where

$$\frac{\partial \widetilde{l_S}(\boldsymbol{\phi}_{SM})}{\partial \boldsymbol{\theta}} = \mathrm{E}\left[\frac{\partial\ \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\lambda \dfrac{\partial P_0(\boldsymbol{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{1 + \lambda\Big[P_0(\boldsymbol{X};\boldsymbol{\theta}) - \pi\Big]}\right] \ .$$

Since it is straightforward to see that

$$\frac{\partial \widetilde{l_S}(\boldsymbol{\phi}_{SM})}{\partial \boldsymbol{\phi}_{SM}} = \mathbf{0}$$

at the true parameter values, we know that the profile log-likelihood function converges in probability to a continuous, vector-valued function and a root of the likelihood equations exists; i.e.,

$$\frac{1}{n}\frac{\partial l_S(\boldsymbol{\phi}^0_{SM})}{\partial \boldsymbol{\phi}_{SM}} \xrightarrow{\ p\ } \mathbf{0} \ .$$

Again using the Law of Large Numbers, we can demonstrate that the convergence in

probability of

$$\frac{1}{n}\frac{\partial^2 l_S(\boldsymbol{\phi}_{SM})}{\partial \boldsymbol{\phi}_{SM}\partial \boldsymbol{\phi}_{SM}^T} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_S(\boldsymbol{\phi}_{SM})}{\partial \boldsymbol{\phi}_{SM}\partial \boldsymbol{\phi}_{SM}^T}$$

is uniform for $\boldsymbol{\phi}_{SM}$ in an open neighborhood for $\boldsymbol{\phi}_{SM}^0$, and at the true parameter values,

$$-\frac{\partial^2 \widetilde{l}_S(\boldsymbol{\phi}_{SM}^0)}{\partial \boldsymbol{\phi}_{SM}\partial \boldsymbol{\phi}_{SM}^T} = \boldsymbol{J} \ ,$$

which can be shown to be invertible. Finally, by applying Theorem 2 in Foutz' (1977) which showed the existence of a consistent solution to the likelihood equations and its uniqueness by using the Inverse Function Theorem, and weakening the requirement of the matrix of second derivatives of the log likelihood function to be negative definite, the result in Theorem follows.

Proof of Theorem 2 (Asymptotic Normality)

We first start from a Taylor series expansion of the estimated score function around the true parameter $\boldsymbol{\phi}_{SM}^0$ evaluated at $\widehat{\boldsymbol{\phi}}_{SM}$,

$$\frac{\partial l_S(\widehat{\boldsymbol{\phi}}_{SM})}{\partial \boldsymbol{\phi}_{SM}} = \frac{\partial l_S(\boldsymbol{\phi}_{SM}^0)}{\partial \boldsymbol{\phi}_{SM}} + \frac{\partial^2 l_S(\widetilde{\boldsymbol{\phi}}_{SM})}{\partial \boldsymbol{\phi}_{SM}\partial \boldsymbol{\phi}_{SM}^T}(\widehat{\boldsymbol{\phi}}_{SM} - \boldsymbol{\phi}_{SM}^0) \ ,$$

where $\widetilde{\boldsymbol{\phi}}_{SM} = \kappa\boldsymbol{\phi}_{SM}^0 + (1-\kappa)\widehat{\boldsymbol{\phi}}_{SM}$ for some $\kappa \in [0,1]$, as in Cosslett (1981b). The left-hand side of the above equation is equal to zero since our estimator $\widehat{\boldsymbol{\phi}}_{SM}$ has been shown to be a consistent solution to $\partial l_S(\boldsymbol{\phi}_{SM})/\partial \boldsymbol{\phi}_{SM} = \mathbf{0}$; after rearranging,

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_{SM} - \boldsymbol{\phi}_{SM}^0) = \left[-\frac{1}{n}\frac{\partial^2 l_S(\widetilde{\boldsymbol{\phi}}_{SM})}{\partial \boldsymbol{\phi}_{SM}\partial \boldsymbol{\phi}_{SM}^T}\right]^{-1}\left[\frac{1}{\sqrt{n}}\frac{\partial l_S(\boldsymbol{\phi}_{SM}^0)}{\partial \boldsymbol{\phi}_{SM}}\right] \ .$$

To prove the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\phi}}_{SM} - \boldsymbol{\phi}^0_{SM})$, it is sufficient to show that $-(1/n)\partial^2 l_S(\widetilde{\boldsymbol{\phi}}_{SM})/\partial\boldsymbol{\phi}_{SM}\partial\boldsymbol{\phi}^T_{SM}$ converges to an invertible matrix in probability and $(1/\sqrt{n})\partial l_S(\boldsymbol{\phi}^0_{SM})/\partial\boldsymbol{\phi}_{SM}$ has an asymptotically normal distribution.

From Theorem 1, we have known that $\widehat{\boldsymbol{\phi}}_{SM} \xrightarrow{p} \boldsymbol{\phi}^0_{SM}$, which implies that $\widetilde{\boldsymbol{\phi}}_{SM} \xrightarrow{p} \boldsymbol{\phi}^0_{SM}$. And we also have shown that

$$\frac{1}{n}\frac{\partial^2 l_S(\boldsymbol{\phi}_{SM})}{\partial\boldsymbol{\phi}_{SM}\partial\boldsymbol{\phi}^T_{SM}} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_S(\boldsymbol{\phi}_{SM})}{\partial\boldsymbol{\phi}_{SM}\partial\boldsymbol{\phi}_{SM}}$$

uniformly for $\boldsymbol{\phi}_{SM} \in \boldsymbol{U}$. According to Lemma 4 in Amemiya (1973), we can see that

$$-\frac{1}{n}\frac{\partial^2 l_S(\widetilde{\boldsymbol{\phi}}_{SM})}{\partial\boldsymbol{\phi}_{SM}\partial\boldsymbol{\phi}^T_{SM}} \xrightarrow{p} -\frac{\partial^2 \widetilde{l}_S(\boldsymbol{\phi}^0_{SM})}{\partial\boldsymbol{\phi}_{SM}\partial\boldsymbol{\phi}_{SM}} = \boldsymbol{J} \ .$$

Since $\boldsymbol{J}$ is shown to be positive definite, it follows that its inverse exists. By the Central Limit Theorem, we have

$$\frac{1}{\sqrt{n}}\frac{\partial l_S(\boldsymbol{\phi}^0_{SM})}{\partial\boldsymbol{\phi}_{SM}} \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{V}) \ ,$$

where

$$\boldsymbol{V} = \mathrm{Var}\left[\frac{\partial l_S(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}^0_{SM})}{\partial\boldsymbol{\phi}_{SM}}\right] \ .$$

Finally, we can apply Slutsky's Theorem (Sen and Singer, 1993) to conclude that $\sqrt{n}(\widehat{\boldsymbol{\phi}}_{SM} - \boldsymbol{\phi}^0_{SM}) \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\phi}^0_{SM}))$, where $\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{V}\boldsymbol{J}$, the asymptotic covariance matrix of $\widehat{\boldsymbol{\phi}}_{SM}$.

Proof of Theorem 3 (A Consistent Estimator for the Asymptotic Variance-Covariance Matrix)

It is noted that the observations from our *Multivariate-ODS* design are *i.i.d.*; thus, the sample covariance matrix over the observed values is consistent for $\boldsymbol{\Sigma}(\boldsymbol{\phi}_{SM})$. Then, it is straightforward to see that

$$\widehat{\boldsymbol{V}}(\boldsymbol{\phi}_{SM}) = \frac{1}{n}\,\widehat{\mathrm{Var}}_{\{i\}}\left[\frac{\partial l_S(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\phi}_{SM})}{\partial \boldsymbol{\phi}_{SM}}\right] \xrightarrow{p} \boldsymbol{V}(\boldsymbol{\phi}_{SM}) \ .$$

By Assumption 3, the components of $\boldsymbol{V}(\boldsymbol{\phi}_{SM})$ are continuous in $\boldsymbol{\phi}_{SM}$. We can then use the triangle inequality to obtain that

$$\|\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_{SM}) - \boldsymbol{V}(\boldsymbol{\phi}^0_{SM})\| \leq \|\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_{SM}) - \boldsymbol{V}(\widehat{\boldsymbol{\phi}}_{SM})\| + \|\boldsymbol{V}(\widehat{\boldsymbol{\phi}}_{SM}) - \boldsymbol{V}(\boldsymbol{\phi}^0_{SM})\| \xrightarrow{p} 0$$

as n goes to ∞. Furthermore, in the proof of Theorem 2, we have shown that

$$\widehat{\boldsymbol{J}}(\widehat{\boldsymbol{\phi}}_{SM}) = -\frac{1}{n}\frac{\partial^2 l_S(\widehat{\boldsymbol{\phi}}_{SM})}{\partial \boldsymbol{\phi}_{SM} \partial \boldsymbol{\phi}^T_{SM}} \xrightarrow{p} \boldsymbol{J}(\boldsymbol{\phi}^0_{SM}) \ ,$$

It then follows that $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\phi}}_{SM})$ is a consistent estimator of the asymptotic covariance matrix.

CHAPTER 6

STATISTICAL INFERENCES FOR MULTIVARIATE-ODS GENERAL SELECTION CRITERION

6.1 Introduction

To investigate the relationships between a disease outcome and an exposure given other characteristics, epidemiology and other biomedical studies often rely on the observational study designs. Cohort and case-control studies are most commonly used designs. The cohort study is to observe several individual exposures and the individual disease occurrence on the basis of a follow-up period and could take a long time to obtain the results. It could cost a lot to conduct a study especially when the disease is rare. Case-control design, on the other hand, is retrospective and studying the patients already having a disease to yield more information on risk factors of this group of people that differ from those who are free of disease (Cornfield, 1951). The case-control study in epidemiology or the choice-based sampling in econometrics are examples of a general scheme, *outcome-dependent sampling* (ODS) design, where the individuals are selected with probabilities depending on their observed outcome variables. The ODS design is appealing in practice because it allows the researchers to concentrate resources

on observations with the greatest amount of information of primary interest (Anderson, 1972).

Much work for studying dichotomous outcomes under an ODS setting has been continuously developed (e.g., White, 1982; Prentice, 1986; Brewlow and Cain, 1988; Lawless et al., 1999; Zhao and Lipsitz, 1992; Schill et al., 1993; Wacholder and Weinberg, 1994; Breslow and Holubkov, 1997; Wang and Zhou, 2006, 2008). The approach to dichotomize or categorize the outcome variable is commonly applied when the outcome is continuous and then one can conduct available statistical methods on the categorical outcomes. However, a selection bias often occurs since such a simplification for the outcome would induce a loss of efficiency and information and increase the risk for misclassification (Suissa, 1991; Zhou et al., 2002; Weaver and Zhou, 2005), especially when the results are sensitive to the choice of the cutpoints.

To directly apply the continuous scale of the outcome variable without losing information on dichotomization, Zhou et al. (2002) considered a general ODS scheme where (i) an overall simple random sample was drawn from the base population (the prospective component); and (ii) additional supplement samples were randomly selected from segments of the outcome space of particular interest (the retrospective component). They proposed a maximum semiparametric empirical likelihood inference procedure without specifying the underlying distribution for the covariates. Weaver and Zhou (2005) further developed a maximum estimated likelihood estimator (MELE) for the continuous outcome under a two-stage ODS scheme. These methods, however, were developed for the case with the univariate continuous outcome.

In practice, multivariate data arise in many contexts, for example, in epidemiological

cohort studies where the outcomes are recorded for members within families, in animal experiments in which treatments are applied to samples of littermates, or in most clinical trials where study subjects are experiencing multiple events. Among these studies, the correlation between the responses cannot be neglected. An increasing number of studies are indeed performed using the *Multivariate-ODS* design, a further generalization of the biased sampling, which is built on the idea of the ODS design with an aggregate of the responses in the multivariate form and at the same preserves the advantages of the ODS. An example of the ongoing study will be given to illustrate this idea in the next paragraph. The usual statistical method for analyzing the multivariate data if accounting for the *Multivariate-ODS* design is no longer appropriate. A statistical inference procedure is needed to take advantage of the *Multivariate-ODS* setting.

We are motivated by the Collaborative Perinatal Project (CPP), a prospective cohort study designed to identify determinants of neurodevelopmental deficits in children (Niswander and Gordon, 1972; Gray et al., 2000). Longnecker et al. (2004) studied the association in humans between maternal third trimester serum polychlorinated biphenyls (PCBs) levels and audiometry results in offsprings at approximately 8 years old. The sample selected by the investigators was according to an ODS scheme: 726 having an 8-year audiometric evaluation of 1200 subjects were selected at random from the underlying population and a supplemental sample of 200 eligible children was randomly selected from the 440 children whose 8-year audiometric evaluation showed sensorineural hearing loss (SNHL). It was anticipated that a sampling design where children with SNHL were oversampled was to enhance the study efficiency relative to an SRS design of the same size. The outcome variable discussed in the paper was whether the child had hearing loss,

defined from each individual's mean hearing level across both ears and then dichotomized by a threshold. Our goal is to develop a proper inference procedure by considering the continuous hearing measures from both ears simultaneously under the *Multivariate-ODS* design to achieve greater efficiency than only considering a simple random sample with the univariate outcome or alternatively simply dichotomizing the continuous outcome.

In this chapter we consider statistical inferences on regression models under a *Multivariate-ODS* design with a general selection criterion for drawing supplemental samples in addition to an overall simple random sample. Specifically, we model the underlying distributions of covariates nonparametrically using the empirical likelihood methods. A novelty of the proposed method is that one can make inferences on the regression parameters without postulating any of the distributions for the covariates by combining a nonparametric component with a parametric regression model. We show that the proposed estimator with the outcome-dependent nature accounted for is more efficient and statistically powerful than other alternative methods. We also investigate that the sampling strategies under the *Multivariate-ODS* framework can be used to design a cost-effective study. The remainder of this chapter is as follows. Section 6.2 presents the notation and the data structure under the *Multivariate-ODS* design with multivariate continuous outcomes. We then demonstrate the likelihood approaches and derive the asymptotic properties. Section 6.3 describes the simulation studies and the small sample properties of our proposed estimator and compares with other methods. We thereafter apply the proposed method to analyze the data in Collaborative Perinatal Project study in Section 6.4 and Section 6.5 gives a brief discussion and suggests some possible extensions of the proposed method in future research.

6.2 The Multivariate-ODS Design and Inference

6.2.1 The Multivariate-ODS Data Structure and Likelihood

To fix notation, let Y_{ij} be the jth continuous outcome for the subject i, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$ ($p \geq 2$), and $\boldsymbol{X}_i$ be a vector of covariates for the ith subject, which can include both discrete and continuous components. Let $\boldsymbol{a} = \{a_j,\ j = 1, \ldots, p\}$ and $\boldsymbol{b} = \{b_j,\ j = 1, \ldots, p\}$, where a_j and b_j are known constants and satisfying $\{a_j > b_j, \forall j\}$, are the fixed cutpoints on the domain of $\boldsymbol{Y}_j = \{Y_{ij},\ \forall i\}$. The data structure of the *Multivariate-ODS* design consists of three components: an overall *simple random sample* (SRS) of size n_0 (≥ 0), a *supplemental sample* of size n_1 (≥ 0) conditional on $\{Y_{i1} > a_1, Y_{i2} > a_2, \ldots, Y_{ip} > a_p\}$, and another *supplemental sample* of size n_2 conditional on $\{Y_{i1} < b_1, Y_{i2} < b_2, \ldots, Y_{ip} < b_p\}$:

(i) SRS Component: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i\right\}$, $i = 1, \cdots, n_0$;

(ii) Supplemental Component 1: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid \{Y_{i1} > a_1, Y_{i2} > a_2, \ldots, Y_{ip} > a_p\}\right\}$, $i = 1, \ldots, n_1$ and $j = 1, \ldots, p$;

(iii) Supplemental Component 2: $\left\{\boldsymbol{Y}_i, \boldsymbol{X}_i \mid \{Y_{i1} < b_1, Y_{i2} < b_2, \ldots, Y_{ip} < b_p\}\right\}$, $i = 1, \ldots, n_2$ and $j = 1, \ldots, p$;

the total sample size in the *Multivariate-ODS* is $n = \sum_{k=0}^{2} n_k$.

Without loss of generality, we assume that $p = 2$, i.e., each individual has two responses, and the cutpoints are set to be a_1, a_2, b_1 and b_2. The joint density of $(\boldsymbol{Y}, \boldsymbol{X})$ can be written as $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})g_{\boldsymbol{X}}(\boldsymbol{X})$, where $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})$ is the conditional density function of $\boldsymbol{Y}$ given $\boldsymbol{X}$, $\boldsymbol{\theta}$ is a vector of the regression coefficients of interest, and $g_{\boldsymbol{X}}(\boldsymbol{X})$ is the

marginal density of $\boldsymbol{X}$, which is independent of $\boldsymbol{\theta}$. The corresponding unknown distribution function of $\boldsymbol{X}$ can be denoted as $G_{\boldsymbol{X}}(\boldsymbol{X})$. We can then write the joint likelihood function, $L_{GL}(\boldsymbol{\theta}, G_{\boldsymbol{X}})$, for $(\boldsymbol{Y}, \boldsymbol{X})$ drawn under the *Multivariate-ODS* design as

$$\begin{aligned} L_{GL}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) &= \left[\prod_{i=1}^{n_0} f(Y_{i1}, Y_{i2}, \boldsymbol{X}_i; \boldsymbol{\theta})\right] \left[\prod_{i=1}^{n_1} f(Y_{i1}, Y_{i2}, \boldsymbol{X}_i; \boldsymbol{\theta} | Y_{i1} > a_1, Y_{i2} > a_2)\right] \\ &\quad \times \left[\prod_{i=1}^{n_2} f(Y_{i1}, Y_{i2}, \boldsymbol{X}_i; \boldsymbol{\theta} | Y_{i1} < b_1, Y_{i2} < b_2)\right] , \end{aligned} \tag{6.1}$$

where the first component is the likelihood from the SRS in the *Multivariate-ODS* while the last two parts are contributions from the two supplemental samples. For simplicity, we define that

$$P_1(\boldsymbol{X}; \boldsymbol{\theta}) = \Pr\{Y_1 > a_1, Y_2 > a_2 | \boldsymbol{X}\} = \int_{a_1}^{\infty} \int_{a_2}^{\infty} f(\boldsymbol{Y} | \boldsymbol{X}; \boldsymbol{\theta}) dY_1 dY_2 \tag{6.2}$$

and

$$\pi_1 = \pi_1(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \int_{\mathbb{X}} P_1(\boldsymbol{x}; \boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}) d\boldsymbol{X} \tag{6.3}$$

are the conditional and marginal probabilities that Y_1 and Y_2 satisfy $\{Y_1 > a_1, Y_2 > a_2\}$;

$$P_2(\boldsymbol{X}; \boldsymbol{\theta}) = \Pr\{Y_1 < b_1, Y_2 < b_2 | \boldsymbol{X}\} = \int_{-\infty}^{b_1} \int_{\infty}^{b_2} f(\boldsymbol{Y} | \boldsymbol{X}; \boldsymbol{\theta}) dY_1 dY_2 \tag{6.4}$$

and

$$\pi_2 = \pi_2(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \int_{\mathbb{X}} P_2(\boldsymbol{x}; \boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}) d\boldsymbol{X} \ , \tag{6.5}$$

are the conditional and marginal probabilities for $\{Y_1 < b_1, Y_2 < b_2\}$. Using Bayes' Law, we can further rewrite the likelihood function in (6.1) as

$$\begin{aligned}
L_{GL}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) &= \left[\prod_{i=1}^{n_0} f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} \frac{f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{\Pr(Y_{i1} > a_1, Y_{i2} > a_2)}\right] \\
&\quad \times \left[\prod_{i=1}^{n_2} \frac{f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{\Pr(Y_{i1} < b_1, Y_{i2} < b_2)}\right] \\
&= \left[\prod_{i=1}^{n_0} f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right] \left[\prod_{i=1}^{n_1} \frac{f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{\pi_1(\boldsymbol{\theta}, G_{\boldsymbol{X}})}\right] \\
&\quad \times \left[\prod_{i=1}^{n_2} \frac{f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta}) g_{\boldsymbol{X}}(\boldsymbol{X}_i)}{\pi_2(\boldsymbol{\theta}, G_{\boldsymbol{X}})}\right] \\
&= \left[\prod_{i=1}^{n} f(Y_{i1}, Y_{i2}|\boldsymbol{X}_i;\boldsymbol{\theta})\right] \left[\left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \pi_1^{-n_1} \pi_2^{-n_2}\right] \\
&= L_{GL1}(\boldsymbol{\theta}) \times L_{GL2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) \ , \qquad (6.6)
\end{aligned}$$

where

$$L_{GL1}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) \qquad (6.7)$$

and

$$L_{GL2}(\boldsymbol{\theta}, G_{\boldsymbol{X}}) = \left(\prod_{i=1}^{n} g_{\boldsymbol{X}}(\boldsymbol{X}_i)\right) \pi_1^{-n_1} \pi_2^{-n_2} \ . \qquad (6.8)$$

There are several possible approaches that could be used to make inferences about $\boldsymbol{\theta}$. Without knowing $G_{\boldsymbol{X}}$, one of the naive approaches is to take the observations in the SRS portion of the *Multivariate-ODS* and derive a maximum likelihood estimator for $\boldsymbol{\theta}$. However, ignoring the information from the supplemental sample would lose accuracy and efficiency. Or, one could obtain $\boldsymbol{\theta}$ by maximizing the conditional likelihood based on the complete data in the *Multivariate-ODS*. Clearly, these two estimators are not

the most efficient since the information regarding the supplemental sample is not fully accounted. If $G_{\boldsymbol{X}}(\boldsymbol{X})$ is parameterized to a parameter vector, say ξ, one could maximize the resulting $L_{GL}(\boldsymbol{\theta}, \widehat{G}_{\boldsymbol{X}})$ subject to $(\boldsymbol{\theta}, \xi)$. However, misspecification of $G_{\boldsymbol{X}}$ could lead to erroneous conclusions so that such approach will be limited only if the form of $G_{\boldsymbol{X}}$ is correctly specified. As a result, a nonparametric modeling of $G_{\boldsymbol{X}}$ is desirable in this case. Nevertheless, $G_{\boldsymbol{X}}$ is an infinite-dimensional nuisance parameter and cannot be easily factored out of $L_{GL2}(\boldsymbol{\theta}, G_{\boldsymbol{X}})$. Thus, to incorporate all the available information in the *Multivariate-ODS* data without specifying $G_{\boldsymbol{X}}$, one needs a new method that will be tractable both theoretically and computationally. We next describe a semiparametric empirical likelihood estimator, where $G_{\boldsymbol{X}}$ is left unspecified.

6.2.2 A Semiparametric Likelihood Approach for the Multivariate-ODS

Our plan for estimating $\boldsymbol{\theta}$ is to develop a profile log likelihood function for $\boldsymbol{\theta}$ by first fixing $\boldsymbol{\theta}$ and obtaining the empirical likelihood function of $G_{\boldsymbol{X}}$ in (6.6) (NPMLE) (Vardi, 1985), which will be a function of $\boldsymbol{\theta}$, π_1, and π_2. Then we can obtain the semiparametric empirical maximum estimator $\widehat{\boldsymbol{\theta}}$ by maximizing the resulting profile log likelihood function over $\boldsymbol{\theta}$. The procedure is detailed in the following.

We first maximize $L_{GL}(\boldsymbol{\theta}, G_{\boldsymbol{X}})$, with $\boldsymbol{\theta}$ fixed, over all discrete distributions whose support includes the observed values by considering a discrete distribution function (i.e. a step function) which has all of its probability located at the observed data points (Vardi, 1985). Let $p_i = dG_{\boldsymbol{X}}(\boldsymbol{X}_i) = g_{\boldsymbol{X}}(\boldsymbol{X}_i)$, $i = 1, \ldots, n$, be the probability mass for the ith covariate vector. We want to find values $\{\widehat{p}_i, \forall i\}$, which maximize the log likelihood

function corresponding to (6.6)

$$l_{GL}(\boldsymbol{\theta}, \{p_i\}) = \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) + \sum_{i=1}^{n} \ln p_i - n_1 \ln \pi_1 - n_2 \ln \pi_2 \ , \tag{6.9}$$

subject to the following constraints:

$$\left\{ \{p_i\} \geq 0 \ \forall i, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i \Big(P_1(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_1 \Big) = 0, \ \sum_{i=1}^{n} p_i \Big(P_2(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_2 \Big) = 0 \right\} . \tag{6.10}$$

The above conditions reflect the fact that $G_{\boldsymbol{X}}$ is a discrete distribution function. For a fixed $\boldsymbol{\theta}$, there exists a unique maximum for $\{p_i\}$ in (6.9) subject to the constraints in (6.10) if 0 is inside the convex hull of the points $\{P_1(\boldsymbol{X}_i;\boldsymbol{\theta}), \ \forall\, i\}$ and $\{P_2(\boldsymbol{X}_i;\boldsymbol{\theta}), \ \forall\, i\}$ (Qin and Lawless, 1994). We use the Lagrange multiplier argument to maximize $l_{GL}(\boldsymbol{\theta}, \{p_i\})$ over all $\{p_i, \forall i\}$,

$$\begin{aligned} H_{GL}(\boldsymbol{\theta}, \{p_i\}, \delta, \lambda_1, \lambda_2) \ &= \ \sum_{i=1}^{n} \ln p_i - n_1 \ln \pi_1 - n_2 \ln \pi_2 - \delta \Big(\sum_{i=1}^{n} p_i - 1 \Big) \\ &\quad -n\lambda_1 \sum_{i=1}^{n} p_i \Big(P_1(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_1 \Big) - n\lambda_2 \sum_{i=1}^{n} p_i \Big(P_2(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_2 \Big) \ , \end{aligned}$$

where the restrictions that $\pi_1 = \sum_{i=1}^{n} p_i P_1(\boldsymbol{X}_i;\boldsymbol{\theta})$ and $\pi_2 = \sum_{i=1}^{n} p_i P_2(\boldsymbol{X}_i;\boldsymbol{\theta})$ are reflected; δ, λ_1 and λ_2 are Lagrange multipliers corresponding to the normalized restriction on the $\{\widehat{p}_i, \forall i\}$. After taking the derivative of H_{GL} with respect to p_i and applying the constraints in (6.10), we obtain $\widehat{\delta} = n$ and

$$\widehat{p}_i = \left\{ n \ \left[1 + \lambda_1 \Big(P_1(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_1 \Big) + \lambda_2 \Big(P_2(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_2 \Big) \right] \right\}^{-1} , \tag{6.11}$$

where $i = 1, \ldots, n$. We can then substitute $\widehat{p}_i$ into to (6.9) to obtain a function of $\boldsymbol{\theta}$, π_1, π_2, λ_1 and λ_2. Define $\boldsymbol{\phi}_{GL}^T = (\boldsymbol{\theta}^T, \pi_1, \pi_2, \lambda_1, \lambda_2)$, representing the combined parameter vector and note that we are treating λ_1, λ_2, π_1 and π_2 as parameters independent of $\boldsymbol{\theta}$. Thus, the resulting profile log likelihood function for $\boldsymbol{\phi}_{GL}$ is

$$\begin{aligned} l_{GL}(\boldsymbol{\phi}_{GL}) &= \sum_{i=1}^{n} \ln f(\boldsymbol{Y}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) \\ &\quad - \sum_{i=1}^{n} \ln\ n\left[1 + \lambda_1\Big(P_1(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_1\Big) + \lambda_2\Big(P_2(\boldsymbol{X}_i;\boldsymbol{\theta}) - \pi_2\Big)\right] - n_1 \ln \pi_1 - n_2 \ln \pi_2\ . \end{aligned} \tag{6.12}$$

From (6.12), we can then obtain the proposed estimator, $\widehat{\boldsymbol{\phi}}_{GL}$, which is a maximizer of (6.12). We refer $\widehat{\boldsymbol{\phi}}_{GL}$ as the semiparametric empirical maximum likelihood estimator (SPMLE). The Newton-Raphson algorithm will be used to solve the score equations with respect to (6.12).

6.2.3 Asymptotic Properties of the SEMLE

The main results for the SEMLE regarding the existence and consistency, asymptotic normality, and a consistent estimator for the asymptotic variance-covariance matrix are demonstrated as three theorems, respectively. Outlines of the proofs of the main results are provided in the Appendix.

We indicate $\boldsymbol{\phi}_{GL}^0$ as the true parameter vector of interest containing $\boldsymbol{\theta}^0$, π_1^0, π_2^0, λ_1^0 and λ_2^0, where π_1^0 and π_1^0 are the true marginal probability that $\{Y_1 > a_1, Y_2 > a_2\}$ and $\{Y_1 < b_1, Y_2 < b_2\}$, respectively; λ_1^0 and λ_2^0 are the true Lagrange multiplier.

Theorem 6.1 (Consistency of the SEMLE): *With probability going to 1 as* $N \to \infty$*, there exists a sequence* $\{\widehat{\boldsymbol{\phi}}_{GL}\}$ *of solutions to the score equations from (6.12) such that* $\widehat{\boldsymbol{\phi}}_{GL} \xrightarrow{p} \boldsymbol{\phi}^0_{GL}$*, where* $\boldsymbol{\phi}^0_{GL}$ *is the true parameter vector of interest. If another sequence* $\{\bar{\boldsymbol{\phi}}_{GL}\}$ *of solutions to the score equations exists such that* $\bar{\boldsymbol{\phi}}_{GL} \xrightarrow{p} \boldsymbol{\phi}^0_{GL}$*, then* $\bar{\boldsymbol{\phi}}_{GL} = \widehat{\boldsymbol{\phi}}_{GL}$ *with probability going to 1 as* $n \to \infty$*.*

Theorem 6.2 (Asymptotic Normality of the SEMLE): *The SEMLE has the following asymptotic normal distribution:*

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_{GL} - \boldsymbol{\phi}^0_{GL}) \xrightarrow{D} N_{(p+2)}\Big(\mathbf{0},\ \boldsymbol{\Sigma}(\boldsymbol{\phi}^0_{GL})\Big) \ ,$$

with the asymptotic variance-covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{V}\boldsymbol{J}^{-1} \ , \tag{6.13}$$

where

$$\boldsymbol{J} = -\frac{\partial^2 \widetilde{l}_{GL}(\boldsymbol{\phi}^0_{GL})}{\partial \boldsymbol{\phi}_{GL} \partial \boldsymbol{\phi}^T_{GL}}$$

and

$$\boldsymbol{V} = \mathrm{Var}\left[\frac{\partial l_{GL}(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}^0_{GL})}{\partial \boldsymbol{\phi}_{GL}}\right] \ ,$$

where $\widetilde{l}_{GL}$ is the limiting form of l_{GL}.

Theorem 6.3 (A Consistent Estimator for the Asymptotic Variance-Covariance Matrix): *A consistent estimator for the variance-covariance matrix shown in Equation*

(6.13) is

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\phi}}_{GL}) = \widehat{\boldsymbol{J}}^{-1}(\widehat{\boldsymbol{\phi}}_{GL})\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_{GL})\widehat{\boldsymbol{J}}^{-1}(\widehat{\boldsymbol{\phi}}_{GL}),$$

where

$$\widehat{\boldsymbol{J}}(\boldsymbol{\phi}_{GL}) = -\frac{1}{n}\frac{\partial^2 l_{GL}(\boldsymbol{\phi}_{GL})}{\partial \boldsymbol{\phi}_{GL}\partial \boldsymbol{\phi}_{GL}^T}$$

and

$$\widehat{\boldsymbol{V}}(\boldsymbol{\phi}_{GL}) = \frac{1}{n}\,\widehat{\mathrm{Var}}_{\{i\}}\left[\frac{\partial l_{GL}(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\phi}_{GL}^0)}{\partial \boldsymbol{\phi}_{GL}}\right].$$

6.3 Simulation Studies

In this section, we evaluate the performance of the proposed estimator in the small samples, by means of simulation studies. We then compare our proposed estimator $\widehat{\boldsymbol{\theta}}_P$ to three competing estimators: (i) the maximum likelihood estimator by maximizing the likelihood from the SRS portion of the *Multivariate-ODS* data ($\widehat{\boldsymbol{\theta}}_R$), (ii) the maximum likelihood estimator by maximizing the conditional likelihood based on the complete *Multivariate-ODS* data ($\widehat{\boldsymbol{\theta}}_C$), and (iii) the maximum likelihood estimator obtained from a random sample of the same size as the *Multivariate-ODS* sample ($\widehat{\boldsymbol{\theta}}_S$). Comparing $\widehat{\boldsymbol{\theta}}_P$ with $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$ will give us an insight of the impact on ignoring the part of the information from the *Multivariate-ODS* sample. The comparison between $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_S$ will demonstrate the efficiency gain of the *Multivariate-ODS* design over the simple random sample of the same size. All simulation studies were conducted using programs written in R.

We consider the following bivariate normal model to generate the simulated data:

$$\boldsymbol{Y}|\boldsymbol{X} \sim N\left[\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \\ \mu_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} {\sigma_1}^2 & \rho\sigma_1\sigma_2 \\ \\ \rho\sigma_1\sigma_2 & {\sigma_2}^2 \end{pmatrix}\right],$$

where $\boldsymbol{Y} = \left(Y_1, Y_2\right)^T$, $\boldsymbol{X} = \left(X_1, X_2\right)^T$, $\mu_1 = \alpha_1 + \beta_1 X_1$ and $\mu_2 = \alpha_2 + \beta_2 X_2$; i.e., the conditional distributions of Y_1 given X_1 and Y_2 given X_2 are normally distributed with means $\alpha_1 + \beta_1 X$ and $\alpha_2 + \beta_2 X$, variances σ_1^2 and σ_2^2, respectively, and the correlation coefficient ρ. Our goal is to estimate the parameter vector $\boldsymbol{\theta}_P = (\alpha_1, \beta_1, \alpha_2, \beta_2, \sigma_1, \sigma_2, \rho)^T$. In particular, we will investigate the behavior of β_1 and β_2 by fixing $\alpha_1 = 0.5$, $\alpha_2 = -0.8$, and $\sigma_1^2 = \sigma_2^2 = 1$ or $\sigma_1^2 = \sigma_2^2 = 1.5$, and allowing $\boldsymbol{\beta}$ to take different values for β_1 and β_2. Then the same models are applied to $\rho = 0.5$ and $\rho = 0.85$ to see how the magnitude of association between outcome variables affects the parameter estimates.

The study *Multivariate-ODS* sample sizes for investigation were $n = 200$ and $n = 800$. The *Multivariate-ODS* design consisted an overall SRS of size n_0 supplemented with two additional samples of sizes n_1 and n_2 separately from individuals whose outcome values fall in the two tails of the outcome distributions. For $n = 200$, we considered (i) $n_0 = 160$, $n_1 = n_2 = 20$ and (ii) $n_0 = 100$, $n_1 = n_2 = 50$; for $n = 800$, (i) $n_0 = 640$, $n_1 = n_2 = 80$ and (ii) $n_0 = 400$, $n_1 = n_2 = 200$. We also considered two settings of the cutpoints: (i) the upper tails of the 90th percentiles from the distributions of $\{Y_{i1},\ \forall i\}$ and $\{Y_{i2},\ \forall i\}$ and the lower tails of the 10th percentiles of the distributions, and (ii) the upper tails of the 70th percentiles and the lower tails of the 30th percentiles. For each experiment in which the independent 1,000 data sets were generated, we computed the parameter estimates and

the estimated standard errors for the proposed method and other competing methods, and the nominal 95% confidence intervals were calculated based on their asymptotic normal distributions.

The simulation results were presented in Tables 6.1 through 6.15. The results in the tables were presented for different combinations of $\boldsymbol{\beta}$, ρ, various cutpoints, allocations of the SRS and the supplemental samples, and the sample sizes n, with three methods. Within each table, the sampling specifications and the covariate distribution were fixed. Tables 6.1 - 6.12 included the small sample properties of the proposed estimator $\widehat{\boldsymbol{\theta}}_P$ and the competing estimators, $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$. Tables 6.13 - 6.15 presented the relative efficiencies of $\widehat{\boldsymbol{\theta}}_S$ versus $\widehat{\boldsymbol{\theta}}_P$ based on the models in Tables 6.1 - 6.12.

6.3.1 The Unbiasedness, the Normality and the Variance Estimator

Tables 6.1 through 6.4 contained simulation results for $\beta_1 = \beta_2 = 0$: $n = 200$ in Tables 6.1 and 6.2 with the correlation coefficients of $\rho = 0.5$ and $\rho = 0.85$, respectively; the same models were considered in Tables 6.3 and 6.4 but with $n = 800$. We make the following observations concerning the results presented in Tables 6.1 - 6.4.

1. The proposed method $\widehat{\boldsymbol{\theta}}_P$ along with $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$ produced unbiased estimates compared with the "true" parameter values under four settings. As the sample size n increased, the bias was even hardly observed.

2. The proposed method $\widehat{\boldsymbol{\theta}}_P$ produced the smallest standard errors for estimating the model parameters whereas $\widehat{\boldsymbol{\theta}}_R$ always provided the least efficient estimators. The standard errors were smaller as the sample size n increased.

3. The proposed estimator $\widehat{\boldsymbol{\theta}}_P$ provided a very good estimate of the true variability; for $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$, the means of the standard error estimates were close to the simulation standard errors as well.

4. The confidence intervals based on the proposed estimator $\widehat{\boldsymbol{\theta}}_P$ provided good coverage close to the nominal 95% level. The same findings were observed for both $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\theta}}_C$.

5. In Table 6.1, within the same sampling design across two settings of cutpoints, the standard errors of $\widehat{\boldsymbol{\theta}}_P$ decreased as the percentiles of the cutpoints increased, indicating that our proposed method was even more efficient and favored when the supplemental samples included more extreme observations. Similar results were observed in Tables 6.2 - 6.4.

6. With the cutpoints fixed, as the proportions of the supplementals samples out of the *Multivariate-ODS* increased, the standard errors of $\widehat{\boldsymbol{\theta}}_P$ decreased, suggesting that $\widehat{\boldsymbol{\theta}}_P$ was more efficient as the supplemental sample sized increased.

7. Above observations were true for both $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

6.3.2 Additional Results for the Unbiasedness, the Normality and the Variance Estimator

Tables 6.5 through 6.8 presented the results for $\beta_1 = -0.5$ and $\beta_2 = \ln(2)$ with the same sampling specifications as those in Tables 6.1 - 6.4 respectively. We observed similar tendencies exhibited in Tables 6.1 - 6.4. The proposed estimator $\widehat{\boldsymbol{\theta}}_P$ continued to outperform the competing estimators and provided consistency and good variance

estimates.

Tables 6.9 through 6.12 presented the results using the same models as Table 6.5 - 6.8 except that now σ_1 and σ_2 increased to be $\sigma_1 = \sigma_2 = 1.5$. The small sample properties observed were similar to those in Tables 6.5 - 6.8 and held well. Note that as the variances increased, the standard errors were larger, which was expected.

6.3.3 The Performance of $\widehat{ARE}$ ($= Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$)

We further investigated the amount of information gained by the use of the *Multivariate-ODS* design over a simple random sample of the same size, and the results of the relative efficiencies (ratios of variances, $Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$) were summarized in Tables 6.13 through 6.15 with different model settings. Throughout the three tables, $\widehat{AREs}$ were greater than one, except for only one case in Table 6.14 which was indeed closer to one. We make the following observations concerning the results in Tables 6.13 through 6.15.

1. The estimates of $\boldsymbol{\beta}$ from the proposed method $\widehat{\boldsymbol{\theta}}_P$ were more efficient than $\widehat{\boldsymbol{\theta}}_S$, indicating that the supplemental sample contained substantial information and the proposed method led to more efficiency gains. Among three tables, the greatest efficiency gains were seen in Table 6.13, where $\rho = 0.85$, the cutpoints were 90% and 10%, and the allocation of $n_0 = 50\%$ and $n_1 = n_2 = 25\%$.

2. With the correlation coefficient and the sampling design fixed, the efficiency gains of $\widehat{\boldsymbol{\theta}}_P$ over $\widehat{\boldsymbol{\theta}}_S$ increased as the cutpoints chosen were located further out in the two tails of the distributions.

3. With the cutpoints and the sampling designs fixed, there was generally an increase in

the relative efficiencies as the correlation coefficient increased from 0.5 to 0.85.

4. With the correlation coefficient and the cutpoints fixed, the efficiency gains of $\widehat{\boldsymbol{\theta}}_P$ over $\widehat{\boldsymbol{\theta}}_S$ generally increased as the proportions of the supplemental samples in the *Multivariate-ODS* increased.

5. As the sample size n increased from 200 to 800, the above observations held.

6. Comparing the results in Tables 6.14 and 6.15, for most cases there was an increase in efficiency gains as the variances of β_1 and β_2 increased.

Overall we can see that the observed efficiency gains for $\widehat{\boldsymbol{\theta}}_P$ obtained by using the *Multivariate-ODS* design were noticeably larger than $\widehat{\boldsymbol{\theta}}_S$ from a simple random sample with the same sample size.

6.3.4 The Effect of Changing Supplemental Sampling Fractions on $\widehat{ARE}$

To investigate the effect of changing the supplemental sampling fractions on the improvement of the *Multivariate-ODS* design over other simple random sample designs, we conducted several simulation experiments using the same simulation models used in Tables 6.5 and 6.7 but with the cutpoints located at the 10th and 90th percentiles of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$. Figures 6.1 and 6.2 presented the relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ over $\widehat{\boldsymbol{\theta}}_R$. Clearly, the efficiency gains of the *Multivariate-ODS* design over the simple random sample design increased with the supplemental sampling fractions, agreed by both sample size considerations, and $\widehat{\boldsymbol{\theta}}_P$ was consistently more efficient than $\widehat{\boldsymbol{\theta}}_R$ regardless of the sampling fractions. Although the efficiency gains increased as the supplemental sample size

increased, it was not practical in reality since it may not be easy to have enough individuals in the extreme tails. We suggested the possible remedy for an appropriate proportion of the supplemental sample to be in the region from 0.3 to 0.6. Figures 6.3 through 6.6 illustrated the standard errors of $\widehat{\boldsymbol{\theta}}_P$, and the relative efficiency of the *Multivariate-ODS* design to a simple random sample of the same sample size across various supplemental sampling fractions γ. The increase in the relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ was not monotone over the fractions although $\widehat{\boldsymbol{\theta}}_P$ was substantially more efficient than $\widehat{\boldsymbol{\theta}}_S$ regardless of the sampling fractions and the sample sizes. We observed that the most efficiency gain for $\widehat{\beta}_1$ was when $\gamma = 0.5$, and for $\widehat{\beta}_2$, the greatest efficiency gains were when $\gamma = 0.3$ as $n = 800$ and $\gamma = 0.5$ as $n = 200$. As γ was larger than 60%, there was a slight decrease in the relative efficiency. Thus, these results suggested that a great efficiency gain can be achieved when γ was between 0.3 and 0.6.

Table 6.1: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	0.000	0.080	0.080	0.947	−0.001	0.079	0.080	0.953
		$\boldsymbol{\theta}_C$	−0.001	0.078	0.077	0.947	0.000	0.076	0.077	0.959
		$\boldsymbol{\theta}_P$	−0.002	0.061	0.060	0.954	−0.002	0.059	0.060	0.953
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.001	0.102	0.101	0.948	−0.003	0.104	0.101	0.951
		$\boldsymbol{\theta}_C$	−0.001	0.091	0.091	0.951	−0.004	0.092	0.091	0.942
		$\boldsymbol{\theta}_P$	0.000	0.054	0.056	0.960	−0.003	0.058	0.056	0.946
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	0.002	0.077	0.080	0.954	0.004	0.078	0.080	0.956
		$\boldsymbol{\theta}_C$	0.001	0.074	0.077	0.957	0.003	0.075	0.077	0.955
		$\boldsymbol{\theta}_P$	0.000	0.065	0.065	0.960	0.002	0.066	0.066	0.950
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	0.000	0.101	0.101	0.954	0.004	0.102	0.101	0.946
		$\boldsymbol{\theta}_C$	0.001	0.087	0.088	0.950	0.004	0.089	0.088	0.943
		$\boldsymbol{\theta}_P$	−0.001	0.064	0.063	0.954	0.001	0.064	0.063	0.946

Table 6.2: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	0.000	0.079	0.080	0.956	0.002	0.079	0.080	0.948
		$\boldsymbol{\theta}_C$	0.000	0.078	0.078	0.948	0.001	0.078	0.078	0.950
		$\boldsymbol{\theta}_P$	0.000	0.057	0.058	0.952	0.001	0.056	0.057	0.962
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	0.000	0.103	0.101	0.944	0.002	0.102	0.101	0.953
		$\boldsymbol{\theta}_C$	−0.002	0.094	0.092	0.939	0.000	0.093	0.092	0.952
		$\boldsymbol{\theta}_P$	−0.001	0.049	0.049	0.941	0.001	0.048	0.049	0.961
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.001	0.080	0.080	0.948	0.001	0.081	0.080	0.946
		$\boldsymbol{\theta}_C$	−0.001	0.077	0.077	0.952	0.001	0.078	0.077	0.954
		$\boldsymbol{\theta}_P$	0.001	0.066	0.066	0.957	0.003	0.066	0.066	0.958
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.001	0.101	0.101	0.954	0.000	0.102	0.101	0.949
		$\boldsymbol{\theta}_C$	0.000	0.091	0.089	0.952	0.002	0.090	0.089	0.947
		$\boldsymbol{\theta}_P$	0.001	0.061	0.060	0.950	0.002	0.061	0.060	0.951

Table 6.3: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	0.000	0.041	0.040	0.947	0.001	0.039	0.040	0.947
		$\boldsymbol{\theta}_C$	0.000	0.040	0.039	0.948	0.000	0.038	0.039	0.945
		$\boldsymbol{\theta}_P$	0.000	0.030	0.030	0.943	0.000	0.030	0.030	0.949
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	0.002	0.050	0.050	0.951	0.003	0.049	0.050	0.962
		$\boldsymbol{\theta}_C$	0.001	0.046	0.045	0.948	0.001	0.044	0.045	0.959
		$\boldsymbol{\theta}_P$	0.000	0.029	0.028	0.939	0.000	0.028	0.028	0.946
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.001	0.038	0.040	0.963	−0.001	0.039	0.040	0.948
		$\boldsymbol{\theta}_C$	−0.001	0.037	0.038	0.959	−0.001	0.038	0.038	0.956
		$\boldsymbol{\theta}_P$	−0.001	0.033	0.033	0.943	−0.001	0.034	0.033	0.947
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.003	0.051	0.050	0.948	−0.003	0.050	0.050	0.953
		$\boldsymbol{\theta}_C$	−0.004	0.044	0.044	0.954	−0.003	0.044	0.044	0.950
		$\boldsymbol{\theta}_P$	−0.002	0.031	0.031	0.951	−0.001	0.031	0.031	0.958

Table 6.4: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	0.000	0.039	0.040	0.949	−0.001	0.038	0.040	0.954
		$\boldsymbol{\theta}_C$	−0.001	0.038	0.039	0.953	−0.001	0.037	0.039	0.954
		$\boldsymbol{\theta}_P$	−0.001	0.028	0.029	0.954	−0.002	0.028	0.029	0.956
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	0.000	0.051	0.050	0.948	0.000	0.050	0.050	0.948
		$\boldsymbol{\theta}_C$	−0.001	0.047	0.046	0.947	0.000	0.047	0.046	0.943
		$\boldsymbol{\theta}_P$	0.000	0.024	0.024	0.947	0.000	0.025	0.024	0.941
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.001	0.038	0.040	0.954	−0.001	0.039	0.040	0.958
		$\boldsymbol{\theta}_C$	0.000	0.037	0.038	0.956	0.000	0.038	0.038	0.952
		$\boldsymbol{\theta}_P$	0.001	0.033	0.033	0.954	0.001	0.033	0.033	0.945
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.001	0.051	0.050	0.941	−0.002	0.051	0.050	0.948
		$\boldsymbol{\theta}_C$	−0.001	0.045	0.044	0.940	−0.002	0.045	0.044	0.942
		$\boldsymbol{\theta}_P$	−0.001	0.030	0.030	0.947	−0.002	0.030	0.030	0.951

Table 6.5: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.497	0.081	0.080	0.948	0.692	0.079	0.080	0.949
		$\boldsymbol{\theta}_C$	−0.498	0.081	0.079	0.942	0.693	0.076	0.078	0.955
		$\boldsymbol{\theta}_P$	−0.498	0.064	0.063	0.940	0.694	0.062	0.064	0.954
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.492	0.100	0.101	0.946	0.696	0.104	0.101	0.948
		$\boldsymbol{\theta}_C$	−0.494	0.095	0.096	0.949	0.693	0.094	0.093	0.943
		$\boldsymbol{\theta}_P$	−0.496	0.058	0.059	0.955	0.693	0.063	0.061	0.934
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.497	0.081	0.080	0.953	0.694	0.079	0.080	0.953
		$\boldsymbol{\theta}_C$	−0.497	0.078	0.078	0.955	0.695	0.077	0.077	0.949
		$\boldsymbol{\theta}_P$	−0.497	0.068	0.068	0.956	0.695	0.068	0.068	0.958
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.504	0.103	0.102	0.940	0.691	0.102	0.102	0.949
		$\boldsymbol{\theta}_C$	−0.504	0.096	0.093	0.933	0.690	0.090	0.091	0.954
		$\boldsymbol{\theta}_P$	−0.501	0.066	0.066	0.949	0.693	0.070	0.066	0.934

Table 6.6: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.503	0.083	0.079	0.938	0.690	0.080	0.079	0.949
		$\boldsymbol{\theta}_C$	−0.503	0.082	0.079	0.941	0.689	0.079	0.078	0.952
		$\boldsymbol{\theta}_P$	−0.501	0.063	0.063	0.950	0.691	0.062	0.063	0.952
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.494	0.104	0.101	0.942	0.697	0.102	0.101	0.945
		$\boldsymbol{\theta}_C$	−0.494	0.099	0.097	0.945	0.697	0.096	0.096	0.947
		$\boldsymbol{\theta}_P$	−0.500	0.055	0.056	0.946	0.694	0.055	0.057	0.959
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.499	0.080	0.080	0.952	0.693	0.080	0.080	0.943
		$\boldsymbol{\theta}_C$	−0.499	0.078	0.079	0.953	0.692	0.078	0.078	0.952
		$\boldsymbol{\theta}_P$	−0.497	0.068	0.067	0.943	0.694	0.068	0.067	0.938
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.503	0.099	0.101	0.956	0.690	0.101	0.101	0.950
		$\boldsymbol{\theta}_C$	−0.503	0.091	0.095	0.954	0.691	0.091	0.094	0.946
		$\boldsymbol{\theta}_P$	−0.503	0.063	0.063	0.954	0.691	0.061	0.063	0.959

Table 6.7: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.501	0.040	0.040	0.948	0.694	0.040	0.040	0.947
		$\boldsymbol{\theta}_C$	−0.501	0.039	0.039	0.949	0.694	0.040	0.039	0.938
		$\boldsymbol{\theta}_P$	−0.501	0.031	0.031	0.951	0.695	0.033	0.032	0.939
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.501	0.049	0.050	0.959	0.692	0.052	0.050	0.949
		$\boldsymbol{\theta}_C$	−0.501	0.047	0.047	0.955	0.691	0.047	0.046	0.946
		$\boldsymbol{\theta}_P$	−0.500	0.029	0.029	0.958	0.693	0.031	0.030	0.944
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.501	0.040	0.040	0.956	0.692	0.040	0.040	0.943
		$\boldsymbol{\theta}_C$	−0.501	0.039	0.039	0.956	0.691	0.039	0.038	0.945
		$\boldsymbol{\theta}_P$	−0.501	0.035	0.034	0.948	0.692	0.035	0.034	0.936
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.498	0.051	0.050	0.954	0.695	0.050	0.050	0.953
		$\boldsymbol{\theta}_C$	−0.499	0.046	0.046	0.954	0.694	0.044	0.044	0.954
		$\boldsymbol{\theta}_P$	−0.499	0.032	0.033	0.960	0.694	0.032	0.033	0.952

Table 6.8: Simulation Results: Bivariate normal with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.85$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.500	0.040	0.040	0.942	0.693	0.041	0.040	0.942
		$\boldsymbol{\theta}_C$	−0.500	0.040	0.039	0.948	0.693	0.041	0.039	0.934
		$\boldsymbol{\theta}_P$	−0.500	0.032	0.031	0.960	0.693	0.033	0.032	0.940
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.499	0.049	0.050	0.959	0.693	0.051	0.050	0.948
		$\boldsymbol{\theta}_C$	−0.499	0.048	0.048	0.939	0.693	0.048	0.048	0.949
		$\boldsymbol{\theta}_P$	−0.499	0.028	0.028	0.949	0.694	0.029	0.028	0.947
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.502	0.040	0.040	0.956	0.692	0.040	0.040	0.943
		$\boldsymbol{\theta}_C$	−0.502	0.040	0.039	0.952	0.692	0.040	0.039	0.950
		$\boldsymbol{\theta}_P$	−0.501	0.034	0.034	0.949	0.693	0.034	0.034	0.950
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.499	0.049	0.050	0.960	0.694	0.049	0.050	0.956
		$\boldsymbol{\theta}_C$	−0.499	0.046	0.047	0.949	0.694	0.045	0.047	0.954
		$\boldsymbol{\theta}_P$	−0.499	0.031	0.031	0.951	0.694	0.031	0.031	0.953

Table 6.9: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.5$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.495	0.119	0.120	0.943	0.702	0.122	0.120	0.950
		$\boldsymbol{\theta}_C$	−0.494	0.116	0.117	0.951	0.702	0.118	0.117	0.947
		$\boldsymbol{\theta}_P$	−0.500	0.091	0.092	0.963	0.697	0.095	0.093	0.938
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.491	0.157	0.152	0.943	0.702	0.153	0.152	0.952
		$\boldsymbol{\theta}_C$	−0.492	0.144	0.141	0.949	0.699	0.141	0.138	0.948
		$\boldsymbol{\theta}_P$	−0.497	0.088	0.086	0.957	0.696	0.086	0.087	0.951
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.496	0.120	0.120	0.954	0.696	0.122	0.120	0.940
		$\boldsymbol{\theta}_C$	−0.496	0.116	0.116	0.953	0.696	0.119	0.115	0.944
		$\boldsymbol{\theta}_P$	−0.495	0.100	0.101	0.956	0.697	0.103	0.101	0.948
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.505	0.156	0.152	0.946	0.690	0.153	0.152	0.948
		$\boldsymbol{\theta}_C$	−0.504	0.141	0.136	0.942	0.690	0.134	0.134	0.953
		$\boldsymbol{\theta}_P$	−0.498	0.100	0.096	0.939	0.693	0.096	0.096	0.949

Table 6.10: Simulation Results: Bivariate normal model with $n = 200$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.85$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.499	0.121	0.120	0.950	0.692	0.122	0.120	0.955
		$\boldsymbol{\theta}_C$	−0.499	0.120	0.118	0.952	0.691	0.120	0.117	0.961
		$\boldsymbol{\theta}_P$	−0.502	0.093	0.092	0.951	0.690	0.092	0.091	0.946
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.498	0.160	0.152	0.946	0.692	0.155	0.152	0.945
		$\boldsymbol{\theta}_C$	−0.500	0.150	0.143	0.950	0.690	0.146	0.142	0.950
		$\boldsymbol{\theta}_P$	−0.499	0.081	0.079	0.944	0.692	0.080	0.079	0.946
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.505	0.118	0.119	0.954	0.691	0.117	0.119	0.957
		$\boldsymbol{\theta}_C$	−0.507	0.115	0.117	0.953	0.690	0.113	0.116	0.956
		$\boldsymbol{\theta}_P$	−0.503	0.100	0.099	0.949	0.693	0.099	0.099	0.955
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.498	0.153	0.152	0.950	0.693	0.153	0.152	0.954
		$\boldsymbol{\theta}_C$	−0.500	0.139	0.139	0.957	0.692	0.111	0.138	0.955
		$\boldsymbol{\theta}_P$	−0.498	0.094	0.092	0.948	0.694	0.092	0.092	0.947

Table 6.11: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.5$, and $X_1 = X_2 \sim N(0,1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.501	0.059	0.059	0.958	0.691	0.058	0.059	0.949
		$\boldsymbol{\theta}_C$	−0.501	0.058	0.058	0.955	0.691	0.056	0.058	0.952
		$\boldsymbol{\theta}_P$	−0.500	0.046	0.046	0.949	0.691	0.044	0.046	0.953
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.503	0.073	0.075	0.955	0.694	0.076	0.075	0.951
		$\boldsymbol{\theta}_C$	−0.503	0.067	0.070	0.954	0.692	0.068	0.069	0.948
		$\boldsymbol{\theta}_P$	−0.501	0.042	0.043	0.950	0.693	0.044	0.043	0.945
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.501	0.060	0.059	0.946	0.693	0.058	0.059	0.955
		$\boldsymbol{\theta}_C$	−0.502	0.059	0.058	0.946	0.693	0.056	0.057	0.951
		$\boldsymbol{\theta}_P$	−0.501	0.050	0.050	0.941	0.694	0.049	0.050	0.952
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.502	0.073	0.075	0.959	0.696	0.073	0.075	0.954
		$\boldsymbol{\theta}_C$	−0.500	0.066	0.067	0.958	0.695	0.065	0.066	0.948
		$\boldsymbol{\theta}_P$	−0.502	0.048	0.048	0.946	0.693	0.047	0.047	0.946

Table 6.12: Simulation Results: Bivariate normal model with $n = 800$, $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$, $\rho = 0.85$, and $X_1 = X_2 \sim N(0, 1)$.

			$\widehat{\beta}_1$				$\widehat{\beta}_2$			
Cutpoints (U, L)	Design	Method	Mean	SE	$\widehat{\text{SE}}$	95% CI	Mean	SE	$\widehat{\text{SE}}$	95% CI
90%, 10%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.500	0.060	0.059	0.954	0.693	0.059	0.059	0.952
		$\boldsymbol{\theta}_C$	−0.500	0.059	0.059	0.957	0.693	0.057	0.058	0.948
		$\boldsymbol{\theta}_P$	−0.500	0.045	0.045	0.952	0.693	0.045	0.045	0.946
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.497	0.074	0.075	0.948	0.695	0.074	0.075	0.950
		$\boldsymbol{\theta}_C$	−0.498	0.071	0.071	0.954	0.694	0.070	0.071	0.940
		$\boldsymbol{\theta}_P$	−0.499	0.040	0.039	0.942	0.693	0.040	0.039	0.948
70%, 30%	$n_1 = n_2 = 10\%$	$\boldsymbol{\theta}_R$	−0.499	0.059	0.059	0.957	0.695	0.058	0.059	0.951
		$\boldsymbol{\theta}_C$	−0.499	0.058	0.058	0.954	0.695	0.057	0.058	0.947
		$\boldsymbol{\theta}_P$	−0.500	0.050	0.050	0.945	0.695	0.049	0.050	0.950
	$n_1 = n_2 = 25\%$	$\boldsymbol{\theta}_R$	−0.497	0.075	0.075	0.952	0.695	0.073	0.075	0.961
		$\boldsymbol{\theta}_C$	−0.496	0.069	0.069	0.951	0.695	0.068	0.069	0.957
		$\boldsymbol{\theta}_P$	−0.497	0.046	0.046	0.954	0.694	0.046	0.046	0.945

Table 6.13: Simulation Results of Relative Efficiencies ($Var_{\hat{\boldsymbol{\theta}}_S}/Var_{\hat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = 0$, $\alpha_2 = -0.8$, $\beta_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $X_1 = X_2 \sim N(0,1)$.

	Cutpoints		Design	$n = 200$		$n = 800$	
ρ	Upper	Lower	$n_1 = n_2$	$\widehat{ARE}_{\hat{\beta}_1}$	$\widehat{ARE}_{\hat{\beta}_2}$	$\widehat{ARE}_{\hat{\beta}_1}$	$\widehat{ARE}_{\hat{\beta}_2}$
0.5	90%	10%	10%	1.37	1.53	1.40	1.37
			25%	1.72	1.44	1.39	1.67
	70%	30%	10%	1.31	1.18	1.13	1.09
			25%	1.24	1.30	1.29	1.21
0.85	90%	10%	10%	1.48	1.60	1.66	1.68
			25%	2.25	2.30	2.11	1.93
	70%	30%	10%	1.07	1.10	1.26	1.18
			25%	1.29	1.30	1.39	1.43

Table 6.14: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1$ and $X_1 = X_2 \sim N(0,1)$.

	Cutpoints		Design	$n = 200$		$n = 800$	
ρ	Upper	Lower	$n_1 = n_2$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	90%	10%	10%	1.21	1.38	1.35	1.17
			25%	1.55	1.44	1.67	1.34
	70%	30%	10%	1.10	1.14	0.98	1.02
			25%	1.13	1.10	1.21	1.07
0.85	90%	10%	10%	1.23	1.22	1.21	1.15
			25%	1.71	1.69	1.58	1.50
	70%	30%	10%	1.09	1.08	1.19	1.10
			25%	1.33	1.36	1.28	1.23

Table 6.15: Simulation Results of Relative Efficiencies ($Var_{\widehat{\boldsymbol{\theta}}_S}/Var_{\widehat{\boldsymbol{\theta}}_P}$): Bivariate normal model with $\alpha_1 = 0.5$, $\beta_1 = -0.5$, $\alpha_2 = -0.8$, $\beta_2 = \ln(2)$, $\sigma_1 = \sigma_2 = 1.5$ and $X_1 = X_2 \sim N(0,1)$.

	Cutpoints		Design	$n = 200$		$n = 800$	
ρ	Upper	Lower	$n_1 = n_2$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$	$\widehat{ARE}_{\widehat{\beta}_1}$	$\widehat{ARE}_{\widehat{\beta}_2}$
0.5	90%	10%	10%	1.53	1.25	1.28	1.37
			25%	1.44	1.54	1.64	1.44
	70%	30%	10%	1.11	1.08	1.07	1.11
			25%	1.13	1.29	1.20	1.25
0.85	90%	10%	10%	1.35	1.41	1.22	1.36
			25%	1.80	1.78	1.70	1.73
	70%	30%	10%	1.13	1.17	1.16	1.24
			25%	1.39	1.44	1.18	1.12

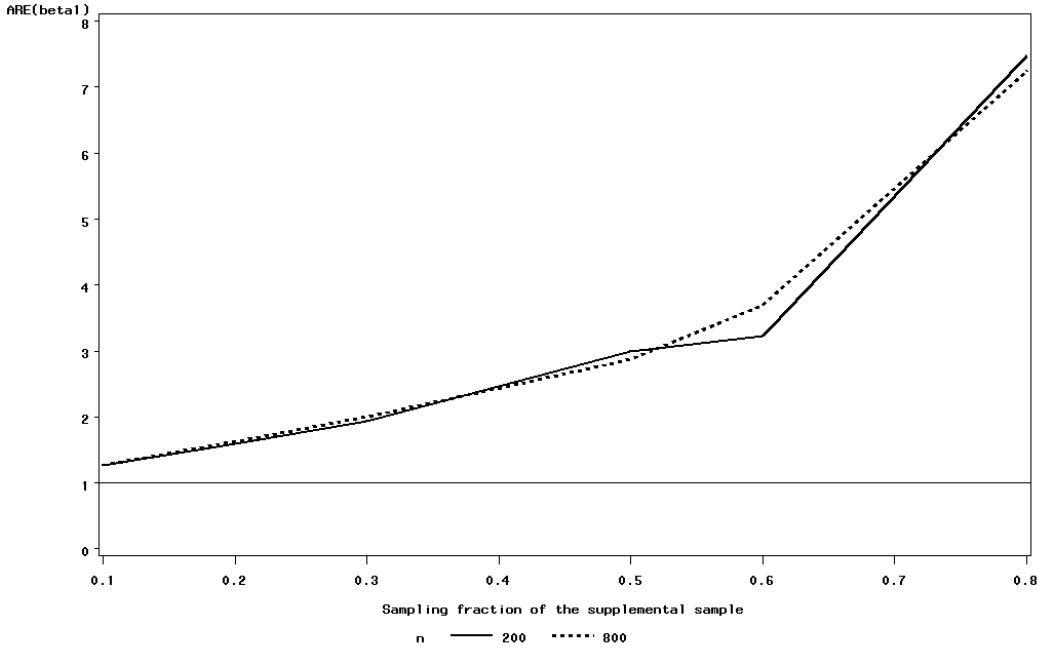

Figure 6.1: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_R$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample, under the models in Tables 6.5 and 6.7.

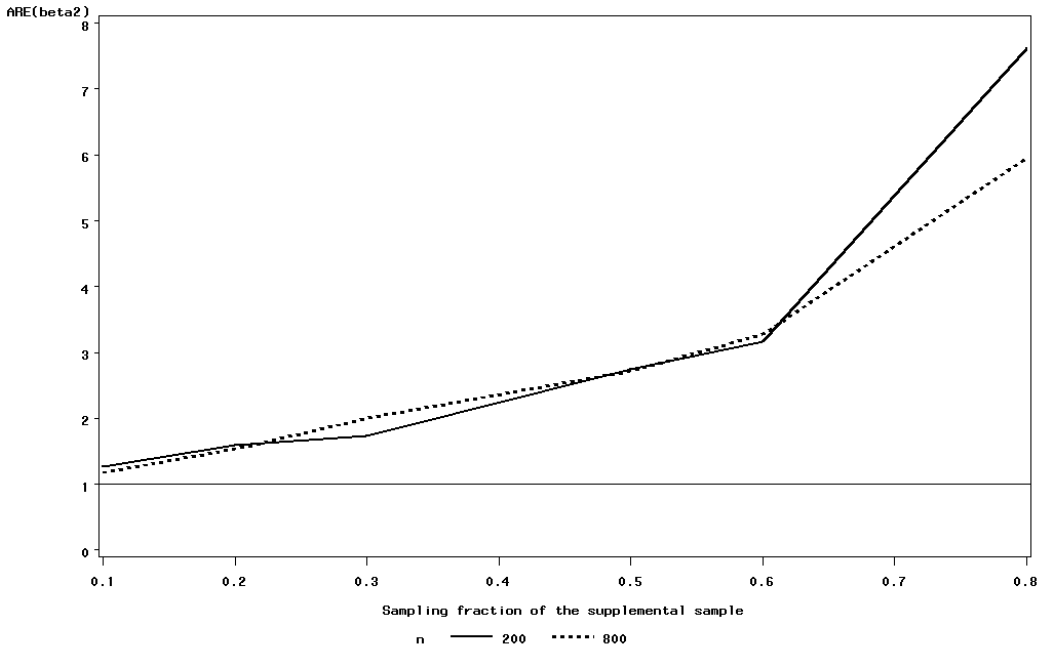

Figure 6.2: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_R$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample, under the models in Tables 6.5 and 6.7.

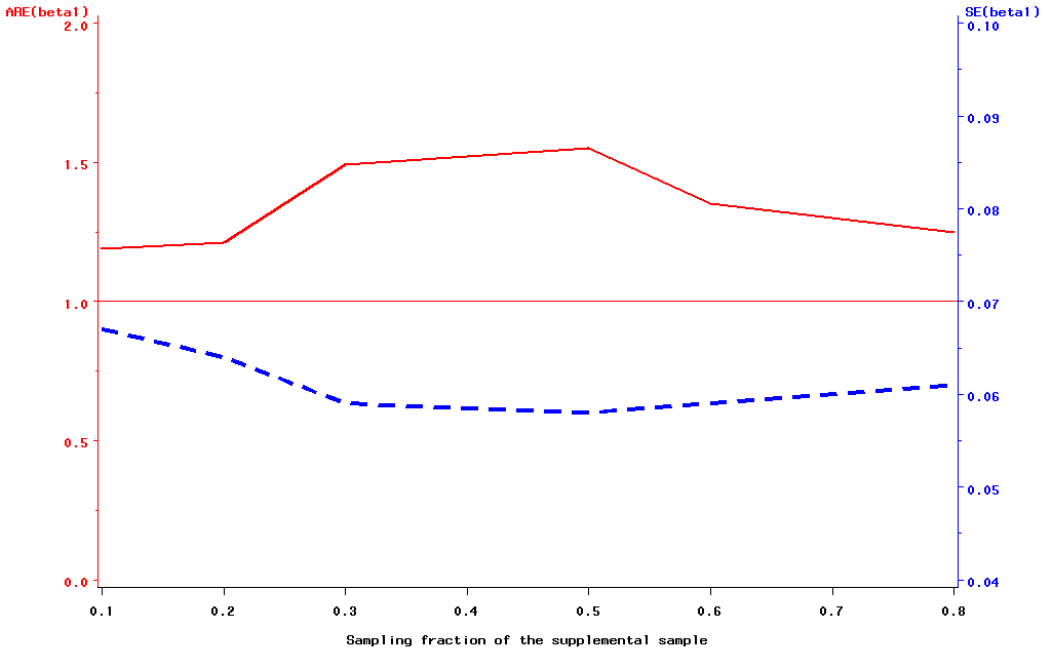

Figure 6.3: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample, under the model in Table 6.5 with the cutpoints = (90%, 10%).

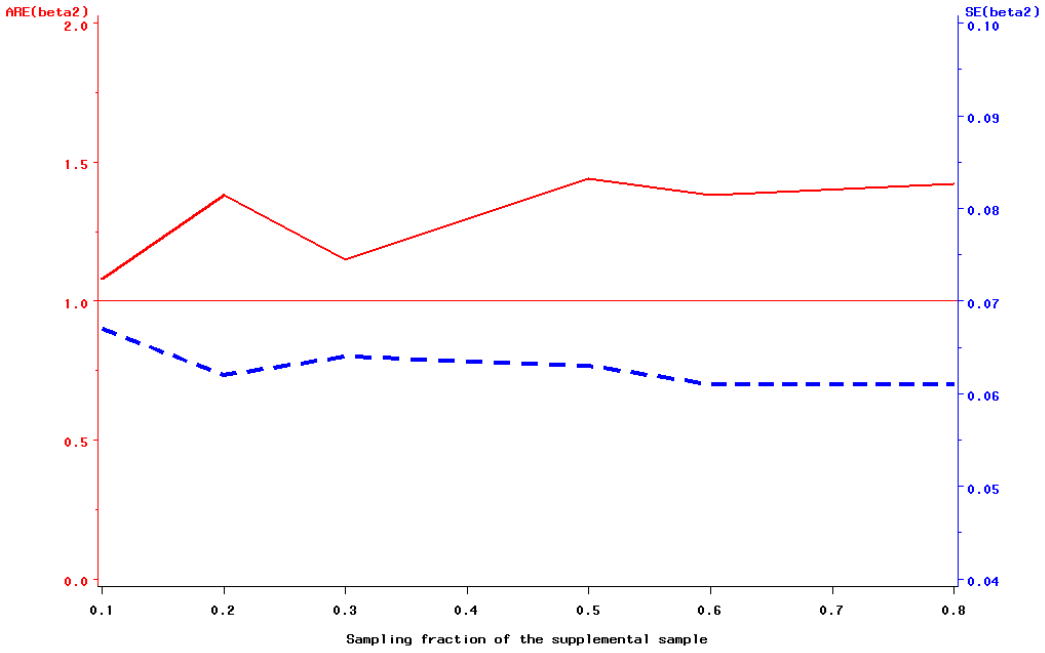

Figure 6.4: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample, under the model in Table 6.5 with the cutpoints = (90%, 10%).

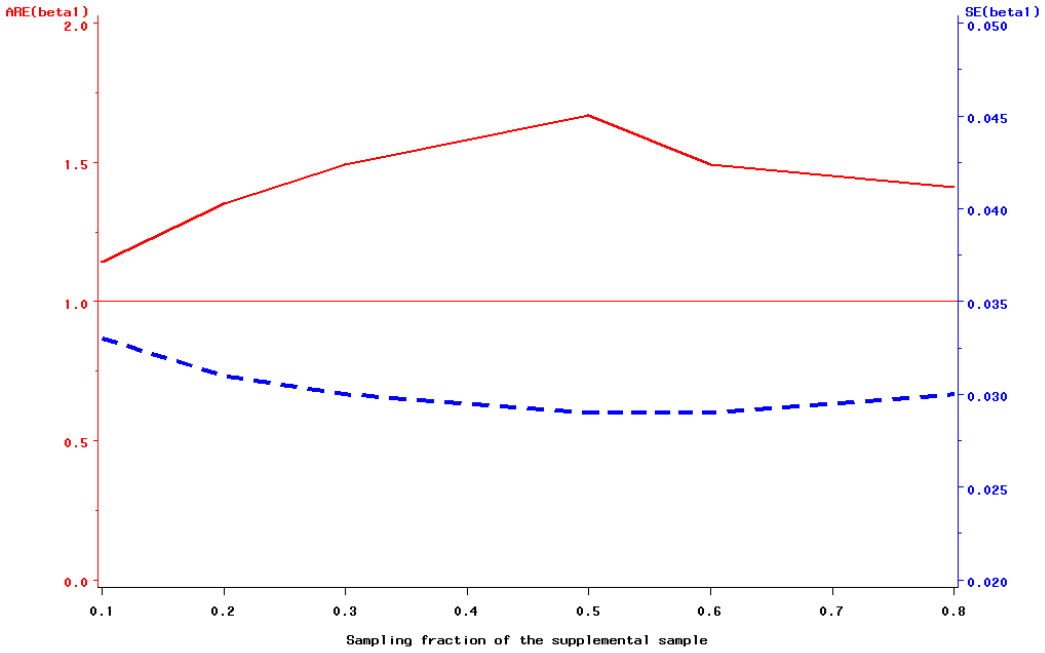

Figure 6.5: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_1$ across the sampling fraction of the supplemental sample under the model in Table 6.7 with the cutpoints = (90%, 10%).

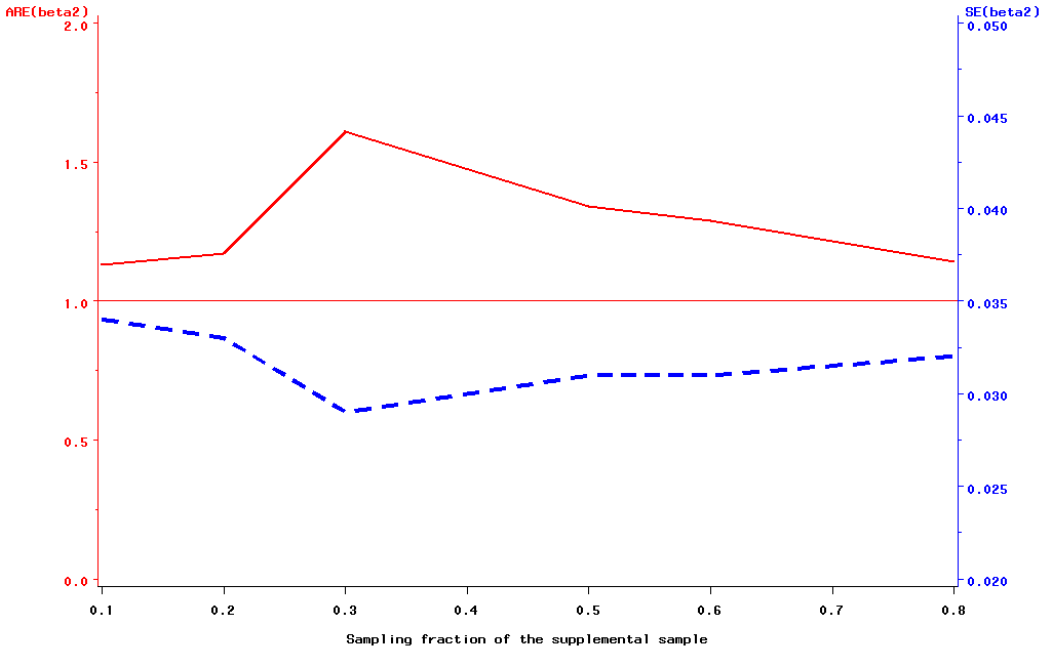

Figure 6.6: Relative efficiency of $\widehat{\boldsymbol{\theta}}_P$ to $\widehat{\boldsymbol{\theta}}_S$ for $\widehat{\beta}_2$ across the sampling fraction of the supplemental sample under the model in Table 6.7 with the cutpoints = (90%, 10%).

6.4 Analysis of the Collaborative Prenatal Project Data

We applied the proposed method to analyze the Collaborative Perinatal Project (CPP) data to study the effect of the third trimester maternal pregnancy serum level of polychlorinated biphenyls (PCBs) on hearing loss children. The CPP was a prospective study designed to identify determinants of neurodevelopmental deficits in children. Details were described in Section 1.2.1. Nearly 56,000 pregnant women were recruited into the CPP study from 1959 through 1966 through 12 study centers across the United States. Women were enrolled, usually at their first prenatal visit; it resulted in 55,908 pregnancies. Data were collected on the mothers at each prenatal visit and at delivery and when the children were 24 hours, 4 and 8 months, and 1, 3, 4, 7, and 8 years.

In a recent environmental epidemiologic study (Longnecker et al., 2001 and 2004), the researchers were interested in studying the relationship between the audiometric evaluation, which was done when the children were approximately 8 years old, and *in utero* exposure to polychlorinated biphenyls (PCBs) measured as the rhird trimester maternal serum PCB level. The study subjects were children born into the CPP. There were 44,075 eligible children who met the following criteria: (1) live born singleton, and (2) a 3-ml third trimester maternal serum specimen was available. The investigators obtained exposure measurements for an outcome-dependent subsample from the population. In particular, the planned sampling design included an SRS of 1,200 subjects from eligible children, of whom 726 had an 8-year audiometric evaluation and a supplemental sample of 200 children whose audiometric evaluation showed sensorineural hearing loss (SNHL), defined defined by a hearing threshold $\geq$ 13.3 dB according to the average across both

ears at 1000, 2000, and 4000 Hz, without any evidence of conductive hearing loss. Evidence of conductive hearing loss exists when the air-bone difference in hearing threshold is ≥ 10 dB again based on the average across both ears. It was anticipated that a sampling design where children with SNHL were oversampled was to enhance the study efficiency relative to an SRS design of the same size.

In our analysis, we took the average measurements at frequencies 1000, 2000, and 4000 Hz for each ear separately to be the continuous outcome variables. The exposure variable of interest was the third trimester maternal serum PCB level (PCB) measured in $\mu g/L$. Additional factors considered potentially confounding included, for the mother, the age (AGE), the socioeconomic index (SEI) score and the highest education level attained when giving birth (EDUC), and the race (RACE) and the gender (SEX) of the child. The covariate of RACE was coded to have two levels: 1 = "White", 0 = "Black and Others". The covariate SEX was coded 1 for males and 0 for females.

We considered the subjects who did not have missing observations for the variables selected into the model fitting and we assumed that missing data were missing completely at random. Of the 44,075 eligible children, 1,256 subjects were selected at random, of which 729 had complete data for the variables mentioned above and will then represent the study population in our data analysis. In order to adjust for our selection criterion described in the previous section, we considered the first and third quartiles of the distributions of hearing levels for each ear as the cutpoints. Hence, 100 out of 729 subjects were those whose hearing level measurements were both above the third quartiles, and 122 children had hearing measurements both below the first quartiles. To illustrate our proposed method with the application of real data, we considered the following two

designs with the total sample size $n = 200$ under the *Multivariate-ODS* design: (i) an overall simple random sample of size $n_0 = 100$ from 729 supplemented with additional samples of $n_1 = 50$ and $n_2 = 50$ separately drawn from the remaining subjects in each group, and (ii) $n_0 = 150$ and $n_1 = n_2 = 25$.

6.4.1 The Conditional Model

After examining the distributions of the hearing levels across three frequencies for each ear, we transformed the outcome variables on the natural log scale in order to exploit the normal properties. We therefore fitted the following linear model to the CPP *Multivariate-ODS* data,

$$\begin{aligned} \ln(Hearing_{ij}) &= \beta_{0j} + \beta_{1j}PCB_i + \beta_{2j}SEX_{ij} + \beta_{3j}RACE_{ij} + \beta_{4j}AGE_{ij} + \beta_{5j}EDUC_{ij} \\ &\quad + \beta_{6j}SEI_{ij} + \epsilon_j \ , \end{aligned} \tag{6.14}$$

where $\epsilon_j \sim N(0, \sigma_j^2)$, $i = 1, \ldots, 200$ and $j = 1$ representing the hearing level across three frequencies from the left ear and $j = 2$ from the right ear. We assumed that $f(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\theta})$ is bivariate normal, where $\boldsymbol{\theta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \sigma_1^2, \sigma_2^2)$ and $\boldsymbol{\beta}_j^T = (\beta_{0j}, \ldots, \beta_{6j})$ and $j = 1, 2$. We estimated the parameters using the methods considered in the simulation studies: the proposed estimator $\boldsymbol{\theta}_P$ and the competing estimators, $\boldsymbol{\theta}_R$ and $\boldsymbol{\theta}_S$.

6.4.2 Results

Tables 6.16 and 6.17 presented the results of the parameter estimates, the estimated standard errors and the 95% confidence intervals calculated based on the asymptotic normal distributions for the proposed method $\widehat{\boldsymbol{\theta}}_P$ and the competing methods $\widehat{\boldsymbol{\theta}}_R$ and $\boldsymbol{\theta}_S$. In Table 6.16, the *Multivariate-ODS* design consisted of an SRS of $n_0 = 100$, and two supplemental samples of sizes $n_1 = n_2 = 50$. Three methods all showed that the corresponding 95% confidence intervals for the PCB effect included 0. Thus, we concluded that *in utero* PCB exposure did not have a significant effect on hearing levels for both ears. Observing the confidence intervals for other confounding parameters for the left ear, the covariate RACE showed a significant effect at the nominal level of 0.05, agreed by the three methods; however, for the right ear, the significance was detected only in $\boldsymbol{\theta}_S$ and $\boldsymbol{\theta}_P$. The results suggested that white children had negative impact on hearing loss; in other words, white children were more likely to have better hearing ability than black and other children. Observing the confidence intervals for other covariates, AGE showed a significance on the borderline for the right ear with $\widehat{\boldsymbol{\theta}}_R$. Table 6.17, where $n_0 = 150$ and $n_1 = n_2 = 25$, exhibited similar results with slightly different estimates and also concluded that RACE was a significant factor by the three methods.

Although PCB was not significant, we could still see some efficiency gains from the results; the observed 95% confidence intervals for PCB provided by the proposed estimator $\widehat{\boldsymbol{\theta}}_P$ were narrower for both ears, compared with the CIs obtained by $\widehat{\boldsymbol{\theta}}_R$; for example, for the left ear in Table 6.16, the CI was $(-0.037, 0.067)$ for $\widehat{\boldsymbol{\theta}}_P$ versus $(-0.063, 0.084)$ for $\widehat{\boldsymbol{\theta}}_R$ and $(-0.058, 0.073)$ for $\widehat{\boldsymbol{\theta}}_S$. It indicated that the proposed estimator provides more

precise estimates. Moreover, $\widehat{\boldsymbol{\theta}}_P$ obtained relatively smaller standard error estimates for all the variables in the model for both ears than those from $\widehat{\boldsymbol{\theta}}_R$. Comparing Table 6.16 with 6.17, we observed that the standard errors for $\widehat{\boldsymbol{\theta}}_P$ decreased as the proportion of the supplemental sample out of the total *Multivariate-ODS* sample size increased, for which our simulation studies also exhibited the same tendency. Hence, there were observable benefits of using the proposed method and taking the advantage of the *Multivariate-ODS* design.

Table 6.16: Results of modeling fitting for the CPP data with $n_0 = 100$, $n_1 = n_2 = 50$, and the total*Multivariate-ODS* sample size $n = 200$.

		$\boldsymbol{\theta}_R$ ($n_0 = 150$)			$\boldsymbol{\theta}_S$ ($n = 200$)			$\boldsymbol{\theta}_P$ ($n = 200$)		
		$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI	$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI	$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI
Left Ear	Int	1.730	0.543	(0.666, 2.795)	1.618	0.384	(0.865, 2.371)	1.651	0.334	(0.997, 2.305)
	PCB	0.011	0.037	(−0.063, 0.084)	0.008	0.033	(−0.058, 0.073)	0.015	0.027	(−0.037, 0.067)
	SEX	−0.069	0.185	(−0.432, 0.294)	0.030	0.126	(−0.216, 0.276)	0.098	0.110	(−0.117, 0.313)
	RACE	−0.701	0.225	(−1.142, −0.260)	−0.887	0.140	(−1.161, −0.612)	−0.800	0.133	(−1.061, −0.540)
	AGE	0.017	0.014	(−0.011, 0.045)	0.012	0.011	(−0.009, 0.032)	0.003	0.009	(−0.014, 0.020)
	EDUC	−0.014	0.043	(−0.099, 0.070)	0.014	0.032	(−0.049, 0.077)	−0.007	0.027	(−0.060, 0.047)
	SEI	0.014	0.056	(−0.096, 0.125)	0.019	0.040	(−0.059, 0.097)	0.034	0.035	(−0.035, 0.104)
Right Ear	Int	1.804	0.570	(0.687, 2.922)	1.688	0.403	(0.897, 2.478)	1.840	0.342	(1.169, 2.511)
	PCB	−0.009	0.039	(−0.086, 0.068)	−0.050	0.035	(−0.118, 0.019)	−0.014	0.027	(−0.067, 0.039)
	SEX	−0.329	0.194	(−0.710, 0.052)	−0.100	0.132	(−0.359, 0.158)	−0.070	0.112	(−0.289, 0.150)
	RACE	−0.304	0.236	(−0.767, 0.159)	−0.852	0.147	(−1.141, −0.564)	−0.459	0.138	(−0.730, −0.188)
	AGE	0.031	0.015	(0.001, 0.061)	0.006	0.011	(−0.016, 0.027)	0.007	0.009	(−0.010, 0.025)
	EDUC	−0.027	0.045	(−0.116, 0.062)	0.026	0.034	(−0.040, 0.092)	−0.011	0.028	(−0.066, 0.044)
	SEI	−0.041	0.059	(−0.157, 0.075)	0.033	0.042	(−0.049, 0.115)	0.006	0.036	(−0.066, 0.077)

Table 6.17: Results of modeling fitting for the CPP data with $n_0 = 150$, $n_1 = n_2 = 25$, and the total*Multivariate-ODS* sample size $n = 200$.

		$\boldsymbol{\theta}_R$ ($n_0 = 150$)			$\boldsymbol{\theta}_S$ ($n = 200$)			$\boldsymbol{\theta}_P$ ($n = 200$)		
		$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI	$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI	$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI
Left Ear	Int	1.413	0.527	(0.380, 2.445)	1.618	0.408	(0.818, 2.417)	1.267	0.398	(0.487, 2.047)
	PCB	0.035	0.042	(−0.046, 0.116)	0.028	0.041	(−0.052, 0.109)	0.037	0.034	(−0.029, 0.104)
	SEX	0.103	0.145	(−0.181, 0.387)	−0.051	0.124	(−0.293, 0.191)	0.115	0.114	(−0.108, 0.339)
	RACE	−0.817	0.174	(−1.158, −0.476)	−0.707	0.138	(−0.978, −0.436)	−0.760	0.136	(−1.027, −0.493)
	AGE	0.018	0.013	(−0.007, 0.044)	0.006	0.011	(−0.015, 0.027)	0.013	0.009	(−0.005, 0.032)
	EDUC	−0.009	0.039	(−0.086, 0.069)	0.026	0.031	(−0.036, 0.087)	−0.006	0.029	(−0.063, 0.052)
	SEI	0.040	0.049	(−0.057, 0.136)	0.008	0.038	(−0.067, 0.083)	0.061	0.037	(−0.012, 0.135)
Right Ear	Int	2.051	0.491	(1.089, 3.014)	1.210	0.380	(0.464, 1.955)	1.802	0.364	(1.088, 2.516)
	PCB	0.007	0.039	(−0.069, 0.083)	0.040	0.038	(−0.036, 0.115)	0.019	0.031	(−0.042, 0.079)
	SEX	−0.101	0.135	(−0.366, 0.163)	−0.032	0.115	(−0.257, 0.194)	−0.113	0.104	(−0.317, 0.091)
	RACE	−0.673	0.162	(−0.991, −0.355)	−0.770	0.129	(−1.023, −0.517)	−0.653	0.125	(−0.898, −0.408)
	AGE	0.011	0.012	(−0.012, 0.035)	0.018	0.010	(−0.002, 0.037)	0.009	0.009	(−0.008, 0.026)
	EDUC	−0.033	0.037	(−0.105, 0.039)	0.036	0.029	(−0.022, 0.093)	−0.023	0.026	(−0.075, 0.029)
	SEI	0.035	0.046	(−0.055, 0.124)	0.006	0.036	(−0.064, 0.075)	0.050	0.034	(−0.016, 0.117)

6.5 Discussion

Much research has been discussed for multivariate continuous data, of which is a common and important form; nevertheless, the methods accounting for the *Multivariate-ODS* design are lacking. Throughout previous sections, we have demonstrated the need for developing the statistical inferences on the *Multivariate-ODS* and proposed a semiparametric empirical likelihood method for multivariate continuous outcomes. The proposed estimator is semiparametric in nature that the underlying distributions of the covariates are modeled nonparametrically using the empirical likelihood methods. We have shown that the proposed estimator is consistent and asymptotically normally distributed and a consistent estimator for the asymptotic variance-covariance exists, by incorporating additional information into such *Multivariate-ODS* design process. We used simulated data generated from a standard linear regression model with Normal errors to examine the performance and the small-sample properties of our proposed estimator. Our limited simulation results indicated that the proposed estimator, $\boldsymbol{\theta}_P$, holds well for all the properties and is more efficient than $\boldsymbol{\theta}_R$, which only takes the simple random sample into consideration, and $\boldsymbol{\theta}_C$, the conditional estimator, using the complete *Multivariate-ODS* data but ignoring additional information in the supplemental sample. For the relative efficiency studies, we observed that $\boldsymbol{\theta}_P$ exhibits more efficiency gains than $\boldsymbol{\theta}_S$, using a simple random sample of the same size as the *Multivariate-ODS* from the underlying population, in terms of different correlation coefficients between the outcomes, the allocations of the cutpoints and the the supplemental fractions. We conclude that the *Multivariate-ODS* design, combined with an appropriate analysis, can provide a cost-effective approach to

further improve study efficiency, for a given sample size. Finally, we applied the proposed method to the Collaborative Perinatal Project data, where the researchers are interested in studying the association between a child's hearing loss and *in utero* exposure to PCBs as well as other covariates of interest. Our results showed that the estimator obtained using the proposed method produced substantially smaller standard errors for both ears than those from the competing methods; moreover, the estimator obtained by $\boldsymbol{\theta}_P$ clearly gained more efficiency and was more precise than the other competing estimators, $\boldsymbol{\theta}_R$ and $\boldsymbol{\theta}_S$, although PCBs could not be concluded as a significant effect.

Our simulated studies also suggest that the higher proportion of the sample sizes of the supplemental samples over the *Multivariate-ODS* sample, the greater the gains of efficiency are, which was similar to the guidance suggested by Zhou et al. (2002) in using the ODS design concerning these issues under one continuous outcome variable. Further investigation for the sample size determination, the optimal sample allocations, the optimal correlation coefficient between the outcomes and power analyses aimed at multivariate outcomes under the *Multivariate-ODS* is required. We considered two-dimensional multivariate data in this dissertation; the future work may include the flexibility of incorporating the covariance structures for higher-dimensional data. Our proposed method can also be applied to the quantitative genetics studies, in which the quantitative trait is modeled as a continuous variable; in fact, more and more studies in order to limit the expenses on the DNA analysis are actually adopting the form of the ODS design. We believe that the proposed methods can be a useful tool toward such studies.

APPENDIX: ASYMPTOTIC RESULTS

For any function $h(\boldsymbol{Y}, \boldsymbol{X})$, let $\mathrm{E}_1\Big[h(\boldsymbol{Y}, \boldsymbol{X})\Big]$ and $\mathrm{E}_2\Big[h(\boldsymbol{Y}, \boldsymbol{X})\Big]$ denote expectations conditional on $\{Y_1 > a_1, Y_2 > a_2\}$ and $\{Y_1 < b_1, Y_2 < b_2\}$, respectively, that

$$\mathrm{E}_1\Big[h(\boldsymbol{Y}, \boldsymbol{X})\Big] = \int_{\mathbb{X}} \frac{1}{\pi_1^0} \int_{a_1}^{\infty} \int_{a_2}^{\infty} h(\boldsymbol{y}, \boldsymbol{x}) f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}^0) d\boldsymbol{y} dG_{\boldsymbol{X}}(\boldsymbol{x})$$

and

$$\mathrm{E}_2\Big[h(\boldsymbol{Y}, \boldsymbol{X})\Big] = \int_{\mathbb{X}} \frac{1}{\pi_2^0} \int_{\infty}^{b_1} \int_{\infty}^{b_2} h(\boldsymbol{y}, \boldsymbol{x}) f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}^0) d\boldsymbol{y} dG_{\boldsymbol{X}}(\boldsymbol{x}) \ .$$

We assume the following regularity conditions:

A1. As $n \to \infty$, $\frac{n_1}{n} \to \gamma_1 > 0$, $\frac{n_2}{n} \to \gamma_2 > 0$ and $\frac{n_0}{n} \to 1 - \gamma_1 - \gamma_2$, where γ_1 is the sampling fraction of the supplemental sample drawn conditional on $\{Y_1 > a_1, Y_2 > a_2\}$ and γ_2 represents the allocation of the supplemental sample conditional on $\{Y_1 < b_1, Y_2 < b_2\}$ to the *Multivariate-ODS* sample.

A2. The parameter space, $\boldsymbol{\Theta}$, is a compact subset of $\mathbb{R}^p$; $\boldsymbol{\theta}^0$ lies in the interior of $\boldsymbol{\Theta}$; the covariate space, $\mathbb{X}$, is a compact subset of $\mathbb{R}^q$, for some $q \geq 1$.

A3. $f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})$ is continuous in both $\boldsymbol{y}$ and $\boldsymbol{\theta}$ and is strictly positive for all $\boldsymbol{y} \in \mathbb{Y}$, $\boldsymbol{x} \in \mathbb{X}$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Furthermore, the partial derivatives, $\partial f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})/\partial\theta_i$ and $\partial^2 f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})/\partial\theta_i \partial\theta_j$, for $i, j = 1, ..., p$, exist and are continuous for all $\boldsymbol{y} \in \mathbb{Y}$, $\boldsymbol{x} \in \mathbb{X}$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

A4. Interchanges of differentiation and integration of $f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})$ are valid for the first and second partial derivatives with respect to $\boldsymbol{\theta}$.

A5. The expected value matrix, $\mathrm{E}\left[-\dfrac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]$, is finite and positive definite at $\boldsymbol{\theta}^0$.

A6. There exists a $\delta > 0$ such that for the set $A = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}^0| \leq \delta\}$,

$$\mathrm{E}\left[\sup_A \left|\frac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right|\right] < \infty,$$

for i, $j = 1, ..., p$.

A7. The derivatives, $\dfrac{\partial P_1(\boldsymbol{x};\boldsymbol{\theta}^0)}{\theta_j}$ and $\dfrac{\partial P_2(\boldsymbol{x};\boldsymbol{\theta}^0)}{\theta_j}$, $j = 1, \ldots, p$, are linearly independent. That is, suppose $\boldsymbol{t}$ and $\boldsymbol{s}$ are any $(p \times 1)$ vectors such that

$$\sum_{j=1}^{p} t_j \frac{\partial P_1(\boldsymbol{x};\boldsymbol{\theta}^0)}{\theta_j} = 0$$

and

$$\sum_{j=1}^{p} s_j \frac{\partial P_2(\boldsymbol{x};\boldsymbol{\theta}^0)}{\theta_j} = 0$$

for almost all $\boldsymbol{x} \in \mathbb{X}$ if $\boldsymbol{t} = \boldsymbol{0}$ and $\boldsymbol{s} = \boldsymbol{0}$.

Proof of Theorem 1 (Consistency)

Using Assumption A1 and the Law of Large Numbers, we have

$$\frac{1}{n}\frac{\partial l_{GL}(\boldsymbol{\phi}_{GL})}{\partial\boldsymbol{\theta}} \xrightarrow{p} \frac{\partial \widetilde{l}_{GL}(\boldsymbol{\phi}_{GL})}{\partial\boldsymbol{\theta}} \ ,$$

where

$$\frac{\partial \widetilde{l}_{GL}(\boldsymbol{\phi}_{GL})}{\partial \boldsymbol{\theta}} = \mathrm{E}\left[\frac{\partial \ \ln f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\lambda\frac{\partial P_1(\boldsymbol{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial P_2(\boldsymbol{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{1 + \lambda_1\Big(P_1(\boldsymbol{X};\boldsymbol{\theta}) - \pi_1\Big) + \lambda_2\Big(P_2(\boldsymbol{X};\boldsymbol{\theta}) - \pi_2\Big)}\right] .$$

Since it is straightforward to see that

$$\frac{\partial \widetilde{l}_{GL}(\boldsymbol{\phi}_{GL})}{\partial \boldsymbol{\phi}_{GL}} = \mathbf{0}$$

at the true parameter values, we know that the profile log-likelihood function converges in probability to a continuous, vector-valued function and a root of the likelihood equations exists; i.e.,

$$\frac{1}{n}\frac{\partial l_{GL}(\boldsymbol{\phi}^0_{GL})}{\partial \boldsymbol{\phi}_{GL}} \xrightarrow{p} \mathbf{0} .$$

Again using the Law of Large Numbers, we can demonstrate that the convergence in probability of

$$\frac{1}{n}\frac{\partial^2 l_{GL}(\boldsymbol{\phi}_{GL})}{\partial \boldsymbol{\phi}_{GL}\partial \boldsymbol{\phi}^T_{GL}} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_{GL}(\boldsymbol{\phi}_{GL})}{\partial \boldsymbol{\phi}_{GL}\partial \boldsymbol{\phi}^T_{GL}}$$

is uniform for $\boldsymbol{\phi}_{GL}$ in an open neighborhood for $\boldsymbol{\phi}^0_{GL}$, and at the true parameter values,

$$-\frac{\partial^2 \widetilde{l}_{GL}(\boldsymbol{\phi}^0_{GL})}{\partial \boldsymbol{\phi}_{GL}\partial \boldsymbol{\phi}^T_{GL}} = \boldsymbol{J} ,$$

which can be shown to be invertible. Finally, by applying Theorem 2 in Foutz' (1977) which showed the existence of a consistent solution to the likelihood equations and its uniqueness by using the Inverse Function Theorem, and weakening the requirement of

the matrix of second derivatives of the log likelihood function to be negative definite, the result in Theorem follows.

Proof of Theorem 2 (Asymptotic Normality)

We first start from a Taylor series expansion of the estimated score function around the true parameter $\boldsymbol{\phi}_{GL}^0$ evaluated at $\widehat{\boldsymbol{\phi}}_{GL}$,

$$\frac{\partial l_{GL}(\widehat{\boldsymbol{\phi}}_{GL})}{\partial \boldsymbol{\phi}_{GL}} = \frac{\partial l_{GL}(\boldsymbol{\phi}_{GL}^0)}{\partial \boldsymbol{\phi}_{GL}} + \frac{\partial^2 l_{GL}(\widetilde{\boldsymbol{\phi}}_{GL})}{\partial \boldsymbol{\phi}_{GL} \partial \boldsymbol{\phi}_{GL}^T}(\widehat{\boldsymbol{\phi}}_{GL} - \boldsymbol{\phi}_{GL}^0) \ ,$$

where $\widetilde{\boldsymbol{\phi}}_{GL} = \kappa \boldsymbol{\phi}_{GL}^0 + (1-\kappa)\widehat{\boldsymbol{\phi}}_{GL}$ for some $\kappa \in [0,1]$, as in Cosslett (1981b). The left-hand side of the above equation is equal to zero since our estimator $\widehat{\boldsymbol{\phi}}_{GL}$ has been shown to be a consistent solution to $\partial l_{GL}(\boldsymbol{\phi}_{GL})/\partial \boldsymbol{\phi}_{GL} = \mathbf{0}$; after rearranging,

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_{GL} - \boldsymbol{\phi}_{GL}^0) = \left[-\frac{1}{n} \frac{\partial^2 l(\widetilde{\boldsymbol{\phi}}_{GL})}{\partial \boldsymbol{\phi}_{GL} \partial \boldsymbol{\phi}_{GL}^T} \right]^{-1} \left[\frac{1}{\sqrt{n}} \frac{\partial l_{GL}(\boldsymbol{\phi}_{GL}^0)}{\partial \boldsymbol{\phi}_{GL}} \right] .$$

To prove the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\phi}}_{GL} - \boldsymbol{\phi}_{GL}^0)$, it is sufficient to show that $-(1/n)\partial^2 l(\widetilde{\boldsymbol{\phi}}_{GL})/\partial \boldsymbol{\phi}_{GL} \partial \boldsymbol{\phi}_{GL}^T$ converges to an invertible matrix in probability and $(1/\sqrt{n})\partial l_{GL}(\boldsymbol{\phi}_{GL}^0)/\partial \boldsymbol{\phi}_{GL}$ has an asymptotic normal distribution.

From Theorem 1, we have known that $\widehat{\boldsymbol{\phi}}_{GL} \xrightarrow{p} \boldsymbol{\phi}_{GL}^0$, which implies that $\widetilde{\boldsymbol{\phi}}_{GL} \xrightarrow{p} \boldsymbol{\phi}_{GL}^0$. And we also have shown that

$$\frac{1}{n} \frac{\partial^2 l_{GL}(\boldsymbol{\phi}_{GL})}{\partial \boldsymbol{\phi}_{GL} \partial \boldsymbol{\phi}_{GL}^T} \xrightarrow{p} \frac{\partial^2 \widetilde{l}_{GL}(\boldsymbol{\phi}_{GL})}{\partial \boldsymbol{\phi}_{GL} \partial \boldsymbol{\phi}_{GL}}$$

uniformly for $\boldsymbol{\phi}_{GL} \in \boldsymbol{U}$. According to Lemma 4 in Amemiya (1973), we can see that

$$-\frac{1}{n}\frac{\partial^2 l_{GL}(\widetilde{\boldsymbol{\phi}}_{GL})}{\partial \boldsymbol{\phi}_{GL}\partial \boldsymbol{\phi}_{GL}^T} \xrightarrow{p} -\frac{\partial^2 \widetilde{l}_{GL}(\boldsymbol{\phi}_{GL}^0)}{\partial \boldsymbol{\phi}_{GL}\partial \boldsymbol{\phi}_{GL}} = \boldsymbol{J} \ .$$

Since $\boldsymbol{J}$ is shown to be positive definite, it follows that its inverse exists. By the Central Limit Theorem, we have

$$\frac{1}{\sqrt{n}}\frac{\partial l_{GL}(\boldsymbol{\phi}_{GL}^0)}{\partial \boldsymbol{\phi}_{GL}} \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{V}) \ ,$$

where

$$\boldsymbol{V} = \mathrm{Var}\left[\frac{\partial l_{GL}(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\phi}_{GL}^0)}{\partial \boldsymbol{\phi}_{GL}}\right] \ .$$

Finally, we can apply Slutsky's Theorem (Sen and Singer, 1993) to conclude that $\sqrt{n}(\widehat{\boldsymbol{\phi}}_{GL} - \boldsymbol{\phi}_{GL}^0) \xrightarrow{D} N(\boldsymbol{0},\ \boldsymbol{\Sigma}(\boldsymbol{\phi}_{GL}^0))$, where $\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{V}\boldsymbol{J}$, the asymptotic covariance matrix of $\widehat{\boldsymbol{\phi}}_{GL}$.

Proof of Theorem 3 (A Consistent Estimator for the Asymptotic Variance-Covariance Matrix) It is noted that the observations from our *Multivariate-ODS* design are *i.i.d.*; thus, the sample covariance matrix over the observed values is consistent for $\boldsymbol{\Sigma}(\boldsymbol{\phi})_{GL}$. Then, it is straightforward to see that

$$\widehat{\boldsymbol{V}}(\boldsymbol{\phi}_{GL}) = \frac{1}{n}\ \widehat{\mathrm{Var}}_{\{i\}}\left[\frac{\partial l(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\phi}_{GL})}{\partial \boldsymbol{\phi}_{GL}}\right] \xrightarrow{p} \boldsymbol{V}(\boldsymbol{\phi}_{GL}) \ .$$

By Assumption 3, the components of $\boldsymbol{V}(\boldsymbol{\phi}_{GL})$ are continuous in $\boldsymbol{\phi}_{GL}$. We can then use

the triangle inequality to obtain that

$$\|\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_{GL}) - \boldsymbol{V}(\boldsymbol{\phi}^0_{GL})\| \leq \|\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\phi}}_{GL}) - \boldsymbol{V}(\widehat{\boldsymbol{\phi}}_{GL})\| + \|\boldsymbol{V}(\widehat{\boldsymbol{\phi}}_{GL}) - \boldsymbol{V}(\boldsymbol{\phi}^0_{GL})\| \xrightarrow{p} 0$$

as n goes to ∞. Furthermore, in the proof of Theorem 2, we have shown that

$$\widehat{\boldsymbol{J}}(\widehat{\boldsymbol{\phi}}_{GL}) = -\frac{1}{n}\frac{\partial^2 l_{GL}(\widehat{\boldsymbol{\phi}}_{GL})}{\partial \boldsymbol{\phi}_{GL} \partial \boldsymbol{\phi}^T_{GL}} \xrightarrow{p} \boldsymbol{J}(\boldsymbol{\phi}^0_{GL}) \ ,$$

It then follows that $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\phi}}_{GL})$ is a consistent estimator of the asymptotic covariance matrix.

CHAPTER 7

SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

7.1 Summary

In this dissertation, we have demonstrated the need for developing the statistical inferences on the *Multivariate-ODS* and proposed semiparametric empirical likelihood methods for multivariate continuous outcomes. The data structure under the *Multivariate-ODS* design considered here consists of an overall simple random sample and some supplemental samples from the segments of the space of the outcomes, which were believed to have greater information. The proposed estimators are semiparametric in the sense that a parametric form is specified for the conditional distribution of the outcome variables given the covariates while the underlying distributions of the covariates are left unspecified.

In Chapter 2, we proposed the semiparametric methods and derived the likelihood functions for estimating the regression parameters under three selection criteria. The first method, the *Multivariate-ODS* with a maximum selection criterion, selects the supplemental sample conditional on the maximum values of the outcomes from each individual exceeding the known cutpoint. The second method, the *Multivariate-ODS* summation

criterion, draws the supplemental sample from those whose sums of the outcome values are above the cutpoint. The third method, the *Multivariate-ODS* general criterion, is a more flexible method since the selection of the supplemental samples was based on each outcome value, instead of choosing one value from all of the outcomes.

In Chapter 3, we established the theoretical asymptotic properties for the estimator from the *Multivariate-ODS* with a maximum selection criterion. We showed that the estimator is consistent and asymptotically normally distributed. The asymptotic variance of the estimator is of a sandwich form and a consistent estimator for the corresponding asymptotic variance matrix is developed.

In Chapter 4, we studied the small sample properties of the proposed estimator from the *Multivariate-ODS* with a maximum selection criterion by using extensive simulation studies. We generated data from the standard linear regression conditional model with normal errors. The results of the simulation studies showed that the asymptotic properties derived in Chapter 3 are preserved well even in the samples of moderate sizes. Moreover, the proposed estimator is more efficient than other competing estimators in terms of small sample relative efficiency. We also applied the proposed method to analyze the data from an ongoing study, the Collaborative Perinatal Project and explore the association between the hearing levels and *in utero* exposure to PCB and other possible covariates. Although our results could not conclude that PCB was a significant factor, we still observed some benefits of our proposed method that the standard errors from the proposed estimator were clearly smaller than those from a simple random sample only.

Chapter 5 is in a form of the article, which discusses the proposed method with a

summation selection criterion. The small sample properties were studied through simulated data. The results also showed that the asymptotic properties of the proposed estimator hold well. Furthermore, the proposed estimator outperformed the competing estimator in terms of the relative efficiency and comparing with the estimator from a simple random sample of the same sample size showed that the *Multivariate-ODS* design is more favored.

Chapter 6 is in a form similar to Chapter 5 and developed the semiparametric empirical maximum likelihood estimator under the *Multivariate-ODS* with a general selection criterion. The asymptotic properties also hold well and the proposed estimator produced smaller standard errors and is more efficient than other competing methods. The proposed method was applied to the CPP data by adjusting for the general selection criterion and the results showed that the proposed estimator had gains in efficiency over the estimator from a simple random sample only.

The three proposed methods provide a cost-effective study for the researchers when the data are in a *Multivariate-ODS* design; three different design specifications are particularly useful since they cover all the needs of choosing the supplemental samples. In addition, the proposed methods are computationally straightforward. Therefore, the *Multivariate-ODS* design provides an approach to further improve the study efficiency and the methods accounting for such design provides a good benchmark in terms of practical performance.

7.2 Directions to Future Research

There is still much work to be done for the *Multivariate-ODS* with continuous multivariate outcomes and we describe several potential directions for future research in the following.

- Further investigations for the sample size determination, the optimal sample allocations, the optimal correlation coefficient between the outcomes and power analyses aimed at multivariate continuous outcomes under the *Multivariate-ODS* are required in order to make the *Multivariate-ODS* more practical for researchers.

- We considered two-dimensional multivariate data in this dissertation; the future work should include the flexibility of incorporating different covariance structures for higher-dimensional data.

- The criteria for model checking and evaluating the fit of the model to the data need to be developed.

- For the applications, our proposed method can be applied to the quantitative genetics studies, in which the quantitative trait is modeled as a continuous variable; in fact, more and more studies in order to limit the expenses on the DNA analysis are actually adopting the form of the ODS design. Moreover, our *Multivariate-ODS* scheme with different selection criteria can be readily applied to different scenarios and needs. We believe that the proposed methods accounting for the nature of the *Multivariate-ODS* design with multivariate continuous outcomes can be a useful tool toward such studies.

REFERENCES

Amemiya, T. (1973). Regression Analysis When the Dependent Variable is Truncated Normal. *Econometrica*, **41**, 997-1016.

Breslow, N. E. and Cain, K.C. (1988). Logistic Regression for Two-Stage Case-Control Data. *Biometrika*, **75**, 11-20.

Breslow, N. E. and Holubkov, R. (1997). Maximum Likelihood Estimation of Logistic Regression Parameters Under Two-Phase, Outcome-Dependent Sampling. *Journal of the Royal Statistical Society, Series B*, **59**, 447-461.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research* vol I: *The Analysis of Case-Control Studies*. Oxford: Oxford University Press.

Chatterjee, N., Chen, Y. H., and Breslow, N. E. (2003). A Pseudoscore Estimator for Regression Problems with Two-Phase Sampling. *Journal of the American Statistical Association*, **98**, 158-168.

Cornfield, J. (1951). A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast, and cervix. *Journal of the National Cancer Institute*, **11**, 1269-1275.

Cox, D. R. (1958). The Regression Analysis of Binary Sequences (with discussion). *Journal of the Royal Statistical Society, Series B*, **20**, 215-242.

Cox, D. R. (1970). *The Analysis of Binary Data*. London: Methuen.

Cox, D. R. and Small, N. J. H. (1978). Testing Multivariate Normality. *Biometrika*, **65**, 263-272.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*, Oxford: Clarendon Press.

Fitzmaurice, G. M., Laird, N. M., and Rotnitzky, A. (1993). Regression Models for Discrete Longitudinal Responses. *Statistical Science*, **8**, 284-309.

Foutz, R. V. (1977). On the Unique Consistent Solution to the Likelihood Equations. *Journal of the American Statistical Association*, **72**, 147-148.

Gray, K. A., Longnecker, M. P., Klebanoff, M. A., Brock, J. W., Zhou, H., and Needham, L. (2000). *In Utero* Exposure to Background Levels of Polychlorinated Biphenyls and Cognitive Functioning Among School-Aged Children. unpublished manuscript, NIEHS.

Higgins, M., Province, M., Heiss, G., Eckfeldt, J., Ellison, C., Folsom, A. R., Rao, D. C., Sprafka, J. M., and Willams, R. (1996). NHLBI Family Heart Study: Objectives and Design. *American Journal of Epidemiology*, **143**, 1219-1228.

Horton, N. J. and Lipsitz, S. R. (1999). Review of Software to Fit Generalized Estimating Equation Regression Models. *American Statistician*, **53**, 160-169.

Hsieh, D. A., Manski, C. F., and McFadden, D. (1985). Estimation of Response Probabilities from Augmented Retrospective Observations. *Journal of the American Statistical Association*, **80**, 651-662.

Jennrich, R. T. (1969). Asymptotic Properties of Non-linear Least Squares Estimators. *Annals of Mathematical Statistics*, **40**, 633-643.

Laird, N. M. and Ware, J. H. (1982). Random-effects Models for Longitudinal Data . *Biometrics*, **33**, 133-158.

Lawless, J. F., Kalbfleisch, J. D., and Wild, C. J. (1999). Semiparametric Methods for Response-Selective and Missing Data Problems in Regression. *Journal of the Royal Statistical Society, Series B*, **61**, 413-438.

Lehmann, E. L. (1999). *Elements of Large Sample Theory.* New York: Springer-Verlag.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal Data Anlysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22.

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). "Multivariate Regression Analyses for Categorical Data. *Journal of the Royal Statistical Society, Series B*, **54**, 3-40.

Liao, D., Myers, R., Hunt, S., Shahar, E., Paton, C. Burke, G., Province, M., and Heiss, G. (1997). Familial history of stroke risk: the Family Heart Study. *Stroke*, **28**, 1908-1912.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. *Journal of the Royal Statistical Society, Series B*, **62**, 711-730.

Longnecker, M., Klebanoff, M., Zhou, H., Wilcox, A., Berendes, H., and Hoffman, H. (1997). Proposal to Study in Utero Exposure to DDE and PCBs in Relation to Male Birth Defects and Neurodevelopmental Outcomes in the Collaborative Perinatal Project. Study Proposal, National Institute of Environmental Health Sciences, Washington, D.C.

Longnecker, M., Hoffman, H., Klebanoff, M. A., Brock, J. W., Zhou, H., Needham, L., Adera, T., Guo, X., and Gray, K. A. (2004). In Utero Exposure to Polychlorinated Biphenyls and Sensorineural Hearing Loss in 8-year-old Children. *Neurotoxicology and Teratology*, **26**, 629-637.

Manski, C. F. and McFadden, D. (1981). Alternative Estimators and Sample Designs for Discrete Choice Analysis. In it Structural Analysis of Discrete Data: with Econometric Applications (eds C. F. Manski and D. McFadden), pp. 2-50. Cambridge: Massachusetts Institute of Technology Press.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, New York: Chapman and Hall.

Niswander, K. R. and Gordon, M. (1972). The Women and Their Pregnancies. USD-HEW Publication No. (NIH) 73-379, USGPO, Washington, DC.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249.

Owen, A. B. (1990). Empirical likelihood for confidence regions. *Annals of Statistics*, **18**, 90-120.

Pepe, M. S. and Fleming, T. R. (1991). A Nonparametric Method for Dealing with Mismeasured Covariate Data. *Journal of the American Statistical Association*, **86**, 108-113.

Prentice, R. L. and Pyke, R. (1979). Logistic Disease Incidence Models and Case-Control Studies. *Biometrika*, **66**, 403-411.

Prentice, R. L. (1986). A case-cohort design for epidemiologic studies and disease prevention trials. *Biometrika*, **73**, 1-11.

Qin, J. and Lawless, J. F. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, **22**, 300-325.

Rao, R. R. (1962). Relations Between Weak and Uniform Convergence of Measures with Application. *Annals of Mathematical Statistics*, **33**, 659-680.

Rudin, W. (1964). *Principles of Mathematical Analysis*, 2nd. Ed. New York: McGraw-Hill.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction Applications.* London: Chapman and Hall.

Scott, A. J. and Wild, C. J. (1986). Fitting Logistic Models Under Case-Control or Choice Based Sampling. *Journal of the Royal Statistical Society, Series B*, **48**, 170-182.

——(1997). Fitting Regression Models to Case-control Data by Maximum Likelihood. *Biometrika*, **84**, 57-71.

Suissa, S. (1991). Binary Methods for Continuous Outcomes: A Parametric Alternative. *Journal of Clinical Epidemiology*, **44**, 241-248.

Vardi, Y. (1985). Empirical Distributions in Selection Bias Models. *The Annals of Statistics*, **13**, 178-203.

Wang, X. and Zhou, H. (2006). A Semiparametric Empirical Likelihood Model for Biased Sampling Schemes with Auxiliary Covariates. *Biometrics*, **62**, 1149-1160.

Ware, J. H. (1985). Linear Models for the Analysis of Longitudinal Studies. *The American Statistician*, **39**, 95-101.

Weaver, M. A. and Zhou, H. (2005). An Estimated Likelihood Method for Continuous Outcome Regression Models with Outcome-Dependent Sampling. *Journal of the American Statistical Association*, **100**, 459-469.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, **50**, 1-25.

White, J. E. (1982). A two stage design for the study of the relationship between a rare exposure and a rare disease. *American Journal of Epidemiology*, **115**, 119-128.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, **44**, 1049-1060.

Zeger, S. L., Liang, K.-Y., and Self, S. G. (1985). The Analysis of Binary Longitudinal Data with Time-Independent Covariates. *Biometrika*, **72**, 31-38.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal Data Anlysis for Discrete and Continuous Outcomes. *Biometrics*, **42**, 121-130.

Zhou, H. and Pepe, M. S. (1995). Auxiliary Covariate Data in Failure Time Regression. *Biometrika*, **82**, 139-149.

Zhou, H., Weaver, M. A., Qin, J., Longnecker, M., and Wang, M. C. (2002), A Semiparametric Empirical Likelihood Method for Data from an Outcome-Dependent Sampling Scheme with a Continuous Outcome. *Biometrics*, **58**, 413-421.