UNIVERSITY OF ALBERTA

SOME CONTRIBUTIONS TO BOOTSTRAP AND EMPIRICAL LIKELIHOOD METHODS TO SOME NON I.I.D. MODELS

BY

THUAN QUOC THACH ©

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

STATISTICS

DEPARTMENT OF MATHEMATICAL SCIENCES

EDMONTON, ALBERTA

SPRING 1998

0-612-29119-7

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Some Contributions to Bootstrap and Empirical Likelihood Methods to Some Non I.I.D. Models** submitted by **Thuan Quoc Thach** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in statistics.

N. G. N. Prasad (Supervisor)

R. Karunamuni

Doug Kelker

D. Kelker

Y. Wu

F. Yeh

for Y. P. Chaubey (External Examiner)

Date: January 28, 1998

TO MY BELOVED WIFE

Abstract

In this dissertation, we study the application of empirical likelihood approach and bootstrap technique in four problems. The empirical likelihood approach is applied to get an improved estimator of the regression parameter in a logistic regression model when values of a covariate for a subset of the study subjects are missing at random. We also applied empirical likelihood technique in a finite population framework to obtain a stable variance estimator of a ratio estimator under two-phase sampling.

We develop two bootstrap sampling algorithms to draw robust inference on the regression parameter under measurement error models with known error variance ratio. A weighted bootstrap procedure is also suggested to draw inference on modeling exceedances over a threshold under one-way random effects model. The thesis includes results from simulation studies for all four problems.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Professor N. G. N. Prasad for his valuable suggestions and comments throughout the entire process of this research and the final preparation of this dissertation. I will always be grateful for his unwavering support, patience and great conscientiousness which have been ideal to the kind of supervision I really wanted. Without his guidance, this thesis would have been impossible.

I wish to express my utmost appreciation to my thesis committee members: Professors Y. P. Chaubey, R. Karunamuni, D. Kelker, F. Yeh and Y. Wu for taking the time to read this thesis and their untiring efforts in coming up with many helpful suggestions and raising innovative questions which provided me with insights on how to improve this thesis further.

I thank my wife, Kin Cheung, for her unfailing support, understanding and encouragement.

My heartfelt thanks to Professor R. Arnab for providing significant input to Chapters 2 and 3. and to Professor P. J. Solomon for providing the data set of Chapter 5.

L. McCann also deserves special mention for extending assistance in some technical and computing related problems.

Contents

Appendices

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation deals with the following problems:

(I) Inference on the slope parameter in a logistic regression when some values of the covariate are missing but values of a surrogate variable are available.

(II) Variance estimation for ratio and regression estimator under two-phase sampling.

(III) Proposed two bootstrap procedures for the slope parameter under measurement error model with known error variance ratio.

(IV) Use a weighted bootstrap procedure to estimate exceedances over a threshold under one-way random effects model.

1.0.1 Problem I

Logistic models are often used to model the conditional mean of a binary response Y with a covariate of interest X. That is, the conditional mean of $Y|X$ is modeled as $F(\beta_0+\beta_1 X)$ where β_0 and β_1 are unknown parameters and $F(x) = (1+e^{-x})^{-1}$.

In such applications, values of the covariate X for a subset of the study subjects may be missing; either the measurement is difficult or expensive to obtain. A closely related variable Z may be used as a surrogate for the covariate X. For example, in the Nurses Health Study described by Rosner, Willett, and Spiegelman (1989), the relationship between breast cancer Y, a binary variable and long-term dietary saturated fat X was examined prospectively. The primary data set consisted of a cohort of 89538 women, but instead of observing X, a surrogate Z was observed, namely a self-administered questionnaire. To understand the relationship between X and Z, 173 nurses became part of a validation study, in which Y, X and Z were observed. The covariate X was not observed directly, but diet was measured sufficiently often in the validation data set, for one week at four different points in the year.

In Chapter 2 of the thesis, we consider a logistic regression with surrogate covariate. Let Y denote a binary outcome variable, X be a covariate of interest and $P_\beta(Y|X) = F(\beta X)$ be the logistic regression model for the conditional distribution of Y given X, where β is a scalar parameter. The objective is to obtain inference on the parameter β when some units of X are missing. The data available for analysis consists of m observations $\{(Y_i, X_i, Z_i), i \in S_m\}$ and $n-m$ observations $\{(Y_i, Z_i), i \in S_{n-m}\}$, where the variables $\{Z_i, i \in S_n\}$ are the surrogate variables for $\{X_i, i \in S_n\}$ with $S_n = S_m \cup S_{n-m}$. We assume that the validation set, S_{n-m}, is a simple random sample from S_n.

1.0.2 Problem II

It is often convenient and economical to collect certain items of information from all units in the sample and other items of information from a subsample of the units in the original sample. This technique known as two-phase sampling or double sampling is useful in several ways. It is generally employed when it is proposed to utilize the information collected in the first phase as supplementary information in order to improve the precision of the information to be collected in the second phase. Thus, in a survey that estimates the total wheat yield in a given locality in Canada, one might use a large sample of n' farms to estimate the total area under wheat cultivation and a subsample of n farms to determine the actual yield.

In this thesis, ratio estimation under two phase sampling is studied. Suppose a simple random sample s' of size n' is taken without replacement from a population of N elements and x_i alone is observed for all elements $i \in s'$. A simple random subsample s of size n is then drawn without replacement from s' and y_i is observed for $i \in s$. A ratio estimator of $\bar{Y}$ is $\bar{y}_{ts} = (\bar{y}_n/\bar{x}_n)\bar{x}_{n'} = r\bar{x}_{n'}$, where $\bar{y}_n$ and $\bar{x}_n$ are the means for s and $\bar{x}_{n'}$ is the mean for s'. Rao and Sitter (1995) proposed a new linearization variance estimator that made a better use of the sample data than the standard formula. They also obtained a jackknife variance estimator and showed that these variance estimators performed well in tracking the conditional mean squared error. The use of the empirical likelihood, Owen (1990, 1991), has become attractive in unifying methods involving the use of auxiliary information in survey sampling. Chen and Qin (1993) employed the empirical likelihood to use summary information on the auxiliary variables in improving the customary estimator under simple random sampling. They showed that the empirical likelihood estimator of the population mean was asymptotically equivalent to the linear regression estimator when the population mean of the auxiliary variable was known (see also Hartley and Rao, 1968). In Chapter 3 of the thesis, we propose two alternative variance estimators. The first one is based on a modification of the standard variance estimator and the second is suggested by the empirical likelihood principle. A Monte Carlo comparison of the proposed variance estimators with other estimators is also given.

1.0.3 Problem III

In educational and social studies, measurements are often subject to measurement error, in the sense that a repetition of the measurements on the same subject does not produce identical results. It is well documented in the literature that the use of statistical methods that ignore such measurement errors can lead to wrong conclusions. For example, Goldstein (1979) demonstrated that a conclusion could be reversed when a correction for the measurement error was introduced in analyzing the data on social class differences in the educational attainment of children aged 11 years. For a good review of the methods for dealing with measurement errors, see Fuller (1987). Woodhouse, Yang, Goldstein, and Rasbash (1996) proposed an adjustment for measurement error in multilevel analysis.

In this thesis, we consider the structural equations model for n random vectors $\mathbf{Z}_i = (X_i, Y_i)^T$. It is assumed that for each $i = 1 \ldots n$, we have

$$\mathbf{Z}_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} U_{1i} \\ U_{2i} \end{pmatrix} + \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} = \mathbf{U}_i + \boldsymbol{\xi}_i, \tag{1.0.1}$$

where

$$U_{2i} = \alpha + \beta U_{1i}, \tag{1.0.2}$$

and the $\mathbf{U}_i$'s are independently distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Gamma}_{\mathbf{U}}$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}_{\mathbf{U}} = \begin{pmatrix} \sigma^2_{U_1} & \beta\sigma^2_{U_1} \\ & \beta^2\sigma^2_{U_1} \end{pmatrix}. \tag{1.0.3}$$

The $\boldsymbol{\xi}_i$'s are i.i.d. with mean vector $\mathbf{0}$ and variance-covariance matrix

$$\boldsymbol{\Gamma}_{\boldsymbol{\xi}} = \begin{pmatrix} \sigma^2_\delta & 0 \\ & \sigma^2_\varepsilon \end{pmatrix}. \tag{1.0.4}$$

We assume that for each i,

$$\mathbf{U}_i \text{ and } \boldsymbol{\xi}_i \text{ are independent} \tag{1.0.5}$$

and

$$\lambda^2 = \sigma^2_\varepsilon / \sigma^2_\delta \text{ is known.} \tag{1.0.6}$$

Let F denote the common distribution of the $\mathbf{Z}_i$'s. By (1.0.1)-(1.0.6), the mean vector $\boldsymbol{\mu}_F$ and covariance matrix $\boldsymbol{\Gamma}_F$ are, respectively, given by

$$\boldsymbol{\mu}(F) = \begin{pmatrix} \mu_X(F) \\ \mu_Y(F) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \alpha + \beta\mu_1 \end{pmatrix} \tag{1.0.7}$$

and

$$\Gamma(F) = \begin{pmatrix} \sigma_{XX}(F) & \sigma_{XY}(F) \\ & \sigma_{YY}(F) \end{pmatrix} = \begin{pmatrix} \sigma_{U_1}^2 + \sigma_\delta^2 & \beta\sigma_{U_1}^2 \\ & \beta^2\sigma_{U_1}^2 + \sigma_\varepsilon^2 \end{pmatrix}. \qquad (1.0.8)$$

Expressions (1.0.1)–(1.0.6) represent the structural linear relationship with known error variance ratio.

Gleser (1983) showed that the asymptotic results in a normal error functional model, with the additional assumption that the first and second order sample moments of the U_{1i} converge to finite limits as $n \to \infty$, were identical to those of a bivariate normal structural model with normal distribution in the U_{1i}. Kelly (1984) showed that the maximum likelihood estimators of the slope parameter β and of the intercept parameter α under normal error model were also the method of moments estimators for non-normal structural models when λ^2 was known. However, the sampling distribution of the slope parameter β was skewed (see Anderson and Sawa, 1982). Large sample normal approximations methods in obtaining inferences are not satisfactory for small samples. In view of this, a bootstrap technique is proposed as an alternative method in dealing with moderate sample sizes. Babu and Singh (1983, 1984) and Babu and Bai (1992), among others, showed that the bootstrap sampling distributions incorporated the skewness of the true distributions. Linder and Babu (1994) studied small sample behavior of the model-based bootstrap in the context of obtaining inference about the slope parameter. In Chapter 4 of the thesis, we propose two bootstrap procedures that incorporate skewness of the true sampling distributions. Theoretical justifications and Monte Carlo comparisons of the proposed methods with the existing ones are also given.

1.0.4 Problem IV

Random effects models are widely used in epidemiologic research to study the degree of familial resemblance with respect to biological characteristics (see Elston, 1977) and in genetics to study heritability of selected traits in animal and plant populations (see Smith, 1980). Solomon (1989) used the random effects model to estimate the expected number of exceedances in systolic blood pressure over a given threshold from a sample of 16 individuals. She adopted the following balanced one-way random effects model:

$$y_{ij} = \mu + u_{ij}, \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, n \tag{1.0.9}$$

where

$$u_{ij} = v_i + e_{ij}, \tag{1.0.10}$$

with y_{ij} being the j-th observation in the i-th class, μ being an unknown parameter to be estimated and u_{ij} being the random error associated with y_{ij}. Here u_{ij} is assumed to be the sum of the random effects, v_i, associated with i-th class and random errors, e_{ij}, associated with j-th observation for the i-th class. The random errors e_{ij} are independent, identically distributed with mean 0 and variance σ_e^2 and the random effects v_i are independent, identically distributed with mean 0 and variance σ_v^2. Further, v_i and e_{ij} are uncorrelated so that the variance-covariance structure of u_{ij} is given by

$$E(u_{ij}u_{i'j'}) = \begin{cases} \sigma_v^2 + \sigma_e^2, & \text{for } i = i' \text{ and } j = j' \\ \sigma_v^2, & \text{for } i = i' \text{ and } j \neq j' \\ 0 & \text{otherwise.} \end{cases} \tag{1.0.11}$$

Let h be a given threshold, and define

$$I_{ij}(h) = \begin{cases} 1 & y_{ij} > h \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad T = \sum_{i=1}^{a} T_i, \text{ where } T_i = \sum_{j=1}^{n} I_{ij}(h). \quad (1.0.12)$$

The goal is to estimate the average number of exceedances, $E(T)$, defined as the expected number of values that exceed a given threshold, its variance, $Var(T)$ and the probability of no exceedance, $Pr(T = 0)$. In Chapter 5 of the thesis, we propose a weighted bootstrap procedure to estimate these quantities. Theoretical justifications and finite sample properties of the proposed bootstrap procedure are also provided.

In this thesis, we are going to investigate (i) the empirical likelihood approach to take advantage of information contained in the entire sample for problems I and II and (ii) the bootstrap approach to obtain robust methods for problems III and IV. In the next two sections, we describe these two approaches.

1.1 The bootstrap

Efron (1979) introduced the bootstrap technique as a very general resampling procedure for estimating the distributions of statistics based on independent observations. The procedure is more widely applicable and has a more profound theoretical basis than the Quenouille-Tukey jackknife (see Efron, 1982). Efron (1979, 1982) considered a number of applications of the bootstrap method.

A formal description of the bootstrap is as follows: A random sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ from an unknown probability distribution F has been observed and we wish to estimate a parameter of interest $T(\boldsymbol{X}, F)$ on the basis of $\boldsymbol{X}$. Let

$\hat{F}$ be the empirical distribution putting probability n^{-1} on each of the observed values $\boldsymbol{X}_i$, $i = 1, 2, \ldots, n$. The bootstrap method consists of approximating the sampling distribution of $T(\boldsymbol{X}, F)$ under F by $T(\boldsymbol{X}^*, \hat{F})$ under $\hat{F}$, where $\boldsymbol{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)$ denotes the random sample of size n from $\hat{F}$. The difficult part of the bootstrap procedure is often the calculation of the true bootstrap distribution. Efron (1979) suggested a Monte Carlo approximation: repeat realizations of $\boldsymbol{X}^*$ by taking B independent random samples of size n from $\hat{F}$, say, $\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \ldots, \boldsymbol{X}_B^*$. The sampling distribution of the corresponding values $T(\boldsymbol{X}_1^*; F), T(\boldsymbol{X}_2^*; F), \ldots, T(\boldsymbol{X}_B^*; F)$ is taken as an approximation to the actual bootstrap distribution of $T(\boldsymbol{X}^*; F)$. This approximation can be made arbitrarily accurate by taking B sufficiently large.

For a wide class of statistics, T, and a wide class of distribution functions F this approximation has a high degree of accuracy. It essentially corrects for the skewness of the sampling distribution. See, for example, Bickel and Freedman (1981), and Babu and Singh (1983, 1984).

Wu (1986) proposed a weighted bootstrap method in the context of classical regression. Generally, the method entails first taking i.i.d. samples $\{t_i, i = 1, 2, ..., n\}$ from an external population having mean 0 and variance 1 and then generating bootstrap data by setting

$$y_i^* = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} + t_i e_i, \quad i = 1, 2, \ldots, n, \tag{1.1.1}$$

where $\boldsymbol{x}_i$ is a $p \times 1$ deterministic vector, $\hat{\boldsymbol{\beta}}$ is the $p \times 1$ vector of least squares estimators of $\boldsymbol{\beta}$ and $e_i = y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$. Liu (1988) suggested that another restriction needed to be imposed on t_i, namely, $E(t_i^3) = 1$, to modify Wu's bootstrap procedure so that it shared the usual second order asymptotic properties of the classical bootstrap.

1.2 The empirical likelihood method

Empirical likelihood, introduced by Owen (1988, 1990), is a computer intensive statistical method, but not as intensive as the bootstrap. However, instead of applying an equal probability weight n^{-1} to all data values, empirical likelihood places arbitrary probabilities on the data points, say p_i on the i-th data value. The weights p_i's are chosen by profiling a multinomial likelihood under a set of constraints. The constraints should reflect some extra knowledge on distributions. If extra information on distributions is available and can be expressed as

$$E\{g_t(X)\} = 0, \quad t = 1, 2, ..., q, \tag{1.2.1}$$

where $g_t(\cdot), t = 1, \ldots, q$ are some known real functions. Then, the empirical likelihood determines the p_i's by maximizing a multinomial likelihood $\prod_{i=1}^n p_i$ subject to

$$\sum_{i=1}^n p_i = 1 \quad \text{and} \quad \sum_{i=1}^n p_i g_t(X_i) = 0, \quad t = 1, 2, \ldots, q. \tag{1.2.2}$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_q$ be the Lagrange multipliers corresponding to the q constraints. Define $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_q)^T$ and $\boldsymbol{g}(X_i) = \{g_1(X_i), g_1(X_i), \ldots, g_q(X_i)\}^T$. The optimal weights are

$$p_i = n^{-1}\{1 + \boldsymbol{\lambda}^T \boldsymbol{g}(X_i)\}^{-1}, \quad i = 1, 2, \ldots, n \tag{1.2.3}$$

where $\boldsymbol{\lambda}$ is the solution of

$$\sum_{i=1}^{n} \frac{g_t(X_i)}{1 + \boldsymbol{\lambda}^T g_t(X_i)} = 0, \quad t = 1, 2, \ldots, q. \tag{1.2.4}$$

An attractive feature of the empirical likelihood approach is that it produces confidence regions whose shapes and orientations are entirely determined by the data and which have coverage accuracy at least comparable with those of bootstrap confidence regions. Its coverage properties have been examined by Hall and Scala (1990) and Diciccio, Hall, and Romano (1989) for the case of a smooth function of a mean of i.i.d. random variables, by Owen (1991) and Chen (1994) for the regression case and by Kolaczyk (1994) for the case of generalized linear models. Note that the use of additional information in the bootstrap is not straightforward.

Chapter 2

Logistic regression model with surrogate covariate

2.1 Introduction

Logistic models are often used to model the conditional mean of a binary response Y with a covariate of interest X. That is, the conditional mean of $Y|X$ is modeled as $F(\beta_0 + \beta_1 X)$ where β_0 and β_1 are unknown parameters and $F(x) = (1 + e^{-x})^{-1}$. In such applications, values of the covariate X for a subset study subjects may be missing; either the measurement is difficult or expensive to obtain. A closely related variable Z may be used as a surrogate for the covariate X. For example, consider the study done by Gladen and Rogan (1979). They examine the disease risk due to body burden of accumulated chemical pollutants in body tissues. Two classes of environmental pollutants which exhibit this "accumulation" phenomenon are the metals, such as DDT's, PCB's and PBB's. Body burden is measured by the levels of the chemicals. For the metals, depot tissues are teeth and bones. For the halogenated hydrocarbon, fat is the depot tissue. The depot tissue is usually impossible or difficult to obtain from living subjects. As a result, a surrogate measurement is necessary, such as blood levels. However, blood levels are usually lower than depot tissue and are affected by nutritional or metabolic state of the individual. Use of such measurements is, therefore, subject to criticism. One approach to this situation is to obtain the data in the form of two independent samples. In the first sample (validation data set) only the information on the response variable Y and the surrogate variable Z are measured. While in the second one (primary data set) information on the covariate X is measured in addition to the information on the response Y and surrogate variable Z. The use of surrogate variables is common, particularly in medical research, and there has been considerable discussion to identify "valid" surrogates. For a review of the use of surrogate variables in clinical trials, see Prentice (1989) and Wittes, Lakatos, and Probstfield (1989).

Although relatively few papers have addressed the missing value problem specifically in the context of logistic regression, there are four general methods for the analysis of incomplete data with surrogate variable that can be widely used, namely, partial case, imputation, maximum likelihood and semi-parametric methods. Perhaps the simplest approach to this problem is the partial case method, which discards cases with missing values. This is the default method used by most statistical software packages such as SAS and SPSS. Since we can measure the outcome and surrogate variables for the discarded units, these units still carry some information on the effect of the covariate. Hence, partial case analysis is not efficient for not using all the available information. Especially, large missing rate in the covariate can add up to a substantial loss of data. A second general approach is to replace the missing values with reasonable estimates (imputed values) and then analyze the data. Several strategies to construct such estimates have been suggested. However, estimates of the variance of the estimated regression parameters from the artificially completed data set are invalid in general. This is because variance estimates have to be corrected for variations due to imputation. One solution to this problem is to assess this variance by computing repeated estimates; following multiple imputation method, see Rubin (1987) and Kalton and Kasprzyk (1986).

A third general approach is to parameterize the conditional probability relationship between X and Z through model $P_\eta(X|Z)$ and to maximize the likelihood

$$L(\boldsymbol{\beta}, \eta) = \prod_{i \in S_m} P_{\boldsymbol{\beta}}(Y_i|X_i) P_\eta(X_i|Z_i) \prod_{i \in S_{n-m}} P_{\boldsymbol{\beta},\eta}(Y_i|Z_i),$$

where $P_{\boldsymbol{\beta},\eta}(Y_i|Z_i) = \int P_{\boldsymbol{\beta}}(Y_i|X_i) P_\eta(X_i|Z_i) dX$, $\boldsymbol{\beta} = (\beta_0, \beta_1)$, S_m and S_{n-m} denote the primary and the validation sets, respectively. However, this parametric method is not generally used in applied work, in part, because misspecification of the nuisance function $P_\eta(X|Z)$ can lead to an inconsistent estimator of $\boldsymbol{\beta}$. Moreover, except for some special cases, implementation of the likelihood based approach is cumbersome; requiring either numerical integration to calculate $P_{\boldsymbol{\beta},\eta}(Y|Z)$ and its derivative, or other complicated algorithms such as expectation maximization (EM) and data augmentation algorithm. For more details, see Schafer (1987) and Tanner and Wong (1987).

The fourth general method for analyzing incomplete data with surrogate variable is to use a non-parametric kernel regression method on the validation data set S_{n-m} to estimate the probability $P_{\beta}(Y|Z)$ (see Carroll and Wand, 1991). They proposed an estimate of β as a solution to

$$\sum_{i \in S_m} S_{\beta}(Y_i|X_i) + \sum_{i \in S_{n-m}} \hat{H}_{\beta}(Y_i|Z_i) = \mathbf{0},$$

where $S_{\beta}(Y|X)$ was the score function of $(Y|X)$ and $\hat{H}_{\beta}(Y|Z)$ was a kernel regression estimate of $(Y|Z)$. A semi-parametric estimate of β based on this method was asymptotically normally distributed. Although this method is generally more robust than the others, it has the disadvantage of requiring a bandwidth selection. Pepe and Fleming (1991) considered a similar problem with discrete covariate Z. Stefanski and Carroll (1985) discussed the case in which all the X_i's were unobserved while $\{Z_i, i \in S_n\}$ were available.

Mak, Li, and Kuk (1986) assumed a model for $P_{\eta}(X|Z)$ and then proposed a bias-corrected estimator of the form $\hat{\boldsymbol{\beta}}_I - \boldsymbol{c}_0$, where $\hat{\boldsymbol{\beta}}_I$ was an estimator based on the imputation method and $\boldsymbol{c}_0$ was an estimate of bias obtained from a bootstrap method. However, in their bootstrap procedure, the information on Y and Z in the validation data set is not being used in the resampling process. Furthermore, it is non-robust with respect to a misspecification of the conditional density $P_{\eta}(X|Z)$. For these reasons, the resulting bootstrap procedure is questionable and hence, can lead to an inefficient estimator.

In the present chapter, we propose an alternative estimator under a logistic regression model with a surrogate variable using the empirical likelihood technique. This estimator is shown to be asymptotically normal and more efficient relative to the partial, imputation and bootstrap estimators.

Section 2.2 of this chapter presents the logistic regression model with surrogate covariate and the three methods of estimation. In Section 2.3, a brief introduction is given to the empirical likelihood and it is shown, explicitly, how empirical likelihood can be applied to the present problem. Section 2.4 derives the asymptotic variance of partial, imputation and empirical likelihood estimators. Some simulation results that compare these methods are given in Section 2.5.

2.2 The Model

Let Y denote a binary outcome variable, X be a covariate of interest and $P_{\beta}(Y|X) = F(\beta_0 + \beta_1 X)$ be the logistic regression model for the conditional distribution of Y given X. In the remainder of this chapter, we consider the case $\beta_0 = 0$ and without lost of generality we let $\beta = \beta_1$. The objective is to estimate the parameter β when some units of X are missing. The data sets available for analysis consist of m observations $\{(Y_i, X_i, Z_i), i \in S_m\}$ and $n - m$ observations $\{(Y_i, Z_i), i \in S_{n-m}\}$, where Z_i is the measurement on the surrogate variable Z for the i-th unit, $i \in S_n$ with $S_n = S_m \cup S_{n-m}$. We assume that the validation set, S_{n-m}, is a simple random sample from S_n.

2.2.1 Partial Case Method

The partial case method estimates the logistic parameter β by maximizing the likelihood function of m complete cases of X, ignoring Z. This likelihood function is written as

$$L_m(\beta|X_1,\ldots,X_m) = \prod_{i\in S_m} F(\beta x_i)^{y_i}[1-F(\beta x_i)]^{1-y_i}. \tag{2.2.1}$$

In practice, the partial case estimate of β, denoted by $\hat{\beta}_m$, and its estimated standard error can be computed by using some standard statistical packages.

2.2.2 Imputation Method

This method involves imputing missing values of X for units in S_{n-m} with predicted values obtained from a simple regression model $X = a + bZ + \varepsilon$, where ε denotes a random vector with mean 0 and variance σ^2. That is, $\{X_i, i \in S_{n-m}\}$ are imputed by

$$\hat{X}_i = \hat{a} + \hat{b}Z_i, \quad i = m+1, m+2, \ldots, n, \tag{2.2.2}$$

where $\hat{a}$ and $\hat{b}$ are the least square estimators based on $\{X_i, Z_i; i \in S_m\}$. It is common practice to treat these imputed values as if they are true values and then compute the variance estimate of β using standard likelihood theory. This procedure can lead to serious underestimation of the true variance of the estimate when the proportion of missing values is appreciable. As a result, the confidence interval based on the resulting estimate will have coverage probability smaller than its corresponding nominal level since the method ignores errors in the estimation of X from Z. To describe this method, consider the likelihood function,

$$L_I(\beta|X_1,\ldots,X_m,\hat{X}_{m+1},\ldots,\hat{X}_n) = \prod_{i\in S_n} F(\beta\tilde{x}_i)^{y_i}[1-F(\beta\tilde{x}_i)]^{1-y_i}, \tag{2.2.3}$$

where

$$\tilde{x}_i = \begin{cases} x_i & i \in S_m \\ \hat{x}_i & i \in S_{n-m}. \end{cases}$$

The estimate of β, denoted by $\hat{\beta}_I$, is then obtained by maximizing the likelihood equation (2.2.3). It can also be noted that the above estimator will be biased due to imputation of X_i's.

2.2.3 Bootstrap Method

Mak *et al.* (1986) proposed a bootstrap procedure to estimate the incurred bias $\hat{\beta}_I$ due to imputation. They suggested a bias-corrected estimator $\hat{\beta}_B = \hat{\beta}_I - c_0$, where c_0 was a correction for the bias induced by the bootstrap sampling. To describe their bootstrap sampling, let $\hat{G}$ and $\hat{H}$ be the empirical probability distributions with mass m^{-1} each at $\{X_i, i \in S_m\}$ and $\{\hat{\varepsilon}_i = X_i - \hat{a} - \hat{b}Z_i, i \in S_m\}$, respectively, where $\hat{a}$ and $\hat{b}$ are the least square estimators based on $\{X_i, Z_i; i \in S_m\}$.

1. Draw a sample $\{X_i^*, i \in S_m\}$ from $\hat{G}$ and generate

   $$Y_i^* = \begin{cases} 0 & \text{with probabilty} \quad 1 - F(\hat{\beta}_m X_i^*) \\ 1 & \text{with probabilty} \quad F(\hat{\beta}_m X_i^*), \end{cases}$$

   where $\hat{\beta}_m$ is the partial case estimator obtained from the logistic regression analysis based on $\{Y_i, X_i; i \in S_m\}$.

2. Draw a sample $\{\hat{\varepsilon}_i^*, i \in S_n\}$ from $\hat{H}$ and let $\{X_i^* = \hat{a} - \hat{b}Z_i^* + \hat{\varepsilon}_i^*, i \in S_n\}$.

3. Then the "bootstrap sample " will consist of $\{Y_i^*, X_i^*, Z_i^*; i \in S_m\}$ and $\{Y_i^*, Z_i^*; i \in S_{n-m}\}$.

4. Compute $\hat{\beta}_I$ using the imputation method outlined in Section 2.2.2.

5. Repeat steps (1)–(4) above k times to obtain $c_0 = k^{-1}\sum_{h=1}^{k}\hat{\beta}_I^{*h} - \hat{\beta}_m$, where $\hat{\beta}_I^{*h}$ is the value of the estimator $\hat{\beta}_I$ computed on the h-th bootstrap sample, $h = 1, \ldots, k$.

Note that in their bootstrap algorithm described above, $\{Y_i, X_i; i \in S_{n-m}\}$ are not being used in the resampling procedure. It is often the case that the number of units in S_{n-m} is much larger than the number of units in S_m; therefore, the procedure can lead to an unstable variance estimator.

2.3 The empirical Likelihood method

For the present problem, we employ the empirical likelihood method to use all the information from both S_m and S_{n-m} through the following constraint

$$\sum_{i \in S_n} S_\theta(Y_i|Z_i) = 0, \qquad (2.3.1)$$

where $S_\theta(Y|Z) = Z[Y - F(\theta Z)]$ is the score function obtained from the likelihood function

$$L(\theta|Y_1, \ldots, Y_n; Z_1, \ldots, Z_n) = \prod_{i \in S_n} F(\theta z_i)^{y_i}[1 - F(\theta z_i)]^{1-y_i}. \qquad (2.3.2)$$

Here, the goodness-of-fit of the logistic regression model of Y on Z is not relevant. The idea of fitting the above model is only to extract association between X and Y through the associated information between Z and Y when X and Z are correlated.

2.3.1 The Proposed Method

We apply empirical likelihood method for the model $P_\beta(Y|X) = F(\beta X)$ by maximizing the conditional likelihood

$$L(p_1, \ldots, p_m|S_n) = \prod_{i \in S_m} p_i, \tag{2.3.3}$$

with respect to p_i, $i = 1, \ldots, m$ subject to restrictions

$$\sum_{i \in S_m} p_i = 1, \quad p_i \geq 0, \quad i \in S_m \quad \text{and} \quad \sum_{i \in S_m} S_\theta(Y_i|Z_i)p_i = 0. \tag{2.3.4}$$

where $S_\theta(Y|Z) = Z[Y - F(\theta Z)]$. The last restriction follows from the fact that

$$\sum_{i \in S_n} S_\theta(Y_i|Z_i) = 0. \tag{2.3.5}$$

Then the maximum of $\log L(p_1, p_2, \ldots, p_m|S_n)$ may be found via Lagrange multipliers by letting

$$H = \sum_{i \in S_m} \log p_i + \lambda_1 \left(1 - \sum_{i \in S_m} p_i\right) - m\lambda_2 \sum_{i \in S_m} p_i S_{\hat{\theta}}(Y_i|Z_i), \tag{2.3.6}$$

where the λ's are Lagrange multipliers and $\hat{\theta}$ is the maximum likelihood estimator of θ obtained as the solution to the equation (2.3.5). Taking the derivatives with respect to p_i, we have

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - \lambda_1 - m\lambda_2 S_{\hat{\theta}}(Y_i|Z_i) = 0. \tag{2.3.7}$$

Hence,

$$\sum_{i \in S_m} p_i \frac{\partial H}{\partial p_i} = m - \lambda_1 = 0 \Rightarrow \lambda_1 = m. \tag{2.3.8}$$

Replacing $\lambda_1 = m$ in (2.3.7), we have

$$p_i = \left(\frac{1}{m}\right)\frac{1}{1+\lambda_2 S_{\hat{\theta}}(Y_i|Z_i)}, \quad i \in S_m. \tag{2.3.9}$$

Now, the restriction from the third part of (2.3.4) is

$$0 = \sum_{i\in S_m} p_i S_{\hat{\theta}}(Y_i|Z_i) = \left(\frac{1}{m}\right)\sum_{i\in S_m}\frac{S_{\hat{\theta}}(Y_i|Z_i)}{1+\lambda_2 S_{\hat{\theta}}(Y_i|Z_i)}. \tag{2.3.10}$$

from which λ_2 and hence the p_i's can be obtained . After obtaining the optimal p_i's, $i = 1\dots, m$ we obtain the empirical likelihood estimator of β from the estimating equation

$$S_m(\beta) = m\sum_{i\in S_m} S_\beta(Y_i|X_i)p_i = 0, \tag{2.3.11}$$

where $S_\beta(Y|X) = X[Y - F(\beta X)]$.

The solution $\hat{\beta}_E$ to equation (2.3.11) can be evaluated by implementing a root finding algorithm such as Brent's method (see Press, Flannery, Teukolsky, and Vetterling, 1993). In the next section, we consider the asymptotic variance of $\hat{\beta}_E$ along with the other estimators.

2.4 Asymptotic variances

This section is devoted to the derivation of the asymptotic variance of $\hat{\beta}_m$, $\hat{\beta}_I$ and $\hat{\beta}_E$. In addition, the asymptotic variance of the maximum likelihood estimator will also be given for the case when all the X_i's, $i \in S_n$ are observed. In the rest of this chapter, the asymptotic results are obtained by letting $m \to \infty$ and $n \to \infty$ such that $m/n \to k'$ where $k' \in (0, 1)$. Further, we assume that there exists a positive constant C such that $|X_i| \leq C$ and $|Z_i| \leq C$ for all i. The asymptotic distribution theory in a general case with density $f_Y(y;\beta)$ relies upon the following assumptions (see Cox and Hinkley, 1974 and Pepe, Reilly, and Fleming, 1994).

(a) The parameter space Ω has finite dimension, is closed and compact, and the true parameter value β is interior to Ω.

(b) The first three derivatives of the log likelihood $l(Y;\beta)$ with respect to β exist in the neighborhood, N_0, of the true parameter value almost surely. Further, in such a neighborhood, n^{-1} times the absolute value of the third derivative is bounded above by a function of Y whose expectation exists. The absolute value of the third derivative of the log likelihood $l(Y;\beta)$ with respect to β is bounded away from 0 in a neighborhood, N_0, almost surely.

(c) $-E\{(\partial^2 l(Y;\beta)/\partial^2\beta)\}$ is finite and positive in the neighborhood, N_0, of the true parameter β.

It can be noted that for the model considered in this chapter, the above conditions hold. The first and second derivatives with respect to β of the loglikelihood $l_m(\beta) = \log L_m(\beta|X_1, \ldots, X_m)$ defined in (2.2.1) are given by

$$\frac{\partial l_m(\beta)}{\partial\beta} = \sum_{i\in S_m} x_i[y_i - F(\beta x_i)], \quad \frac{\partial^2 l_m(\beta)}{\partial\beta^2} = -\sum_{i\in S_m} x_i^2 F(\beta x_i)[1 - F(\beta x_i)].$$

From standard likelihood theory, the asymptotic (unconditional) variance of $\hat{\beta}_m$, $Var(\hat{\beta}_m)$, is then given by

$$\begin{aligned} Var(\hat{\beta}_m) &= -\left\{E\left[\frac{\partial^2 l_m(\beta)}{\partial \beta^2}\right]\right\}^{-1} \\ &= \frac{1}{m}\left\{\frac{1}{n}\sum_{i\in S_n} x_i^2 F(\beta x_i)[1-F(\beta x_i)]\right\}^{-1}. \end{aligned} \tag{2.4.1}$$

If all the X_i's, $i \in S_n$ were observed, the maximum likelihood estimator of β could be obtained by maximizing the function $L(\beta|X_1, \dots, X_n)$ with respect to β. We denote the resulting estimator of β by $\hat{\beta}_C$. Then the asymptotic variance of $\hat{\beta}_C$ is given by

$$Var(\hat{\beta}_C) = \frac{1}{n}\left\{\frac{1}{n}\sum_{i\in S_n} x_i^2 F(\beta x_i)[1-F(\beta x_i)]\right\}^{-1}. \tag{2.4.2}$$

Turning to the asymptotic variance of $\hat{\beta}_I$, consider

$$Var(\hat{\beta}_I) = E\{Var(\hat{\beta}_I|S_n)\} + Var\{E(\hat{\beta}_I|S_n)\}. \tag{2.4.3}$$

Since for large m, $E(\hat{\beta}_I|S_n) \doteq \beta$ implies that $Var\{E(\hat{\beta}_I|S_n)\} \doteq 0$. Hence,

$$Var(\hat{\beta}_I) \doteq \frac{1}{n}\left\{\frac{1}{n}\left[\sum_{i\in S_m} x_i^2 F(\beta x_i)[1-F(\beta x_i)] + \sum_{i\in S_{n-m}} E\left\{\hat{x}_i^2 F(\beta \hat{x}_i)[1-F(\beta \hat{x}_i)]\right\}\right]\right\}^{-1}. \tag{2.4.4}$$

To establish the consistency of $\hat{\beta}_E$, we first obtain the consistency of $\tilde{\beta}_E$, which is the same as $\hat{\beta}_E$ with the condition that the p_i's are fixed known constants. We use the following result on estimating functions due to Foutz (1977).

Theorem 2.1. *There exists a unique consistent solution to an estimating equation $S_m(\beta)$ given in (2.3.11) in a neighborhood N_0 if*

(i) $|\partial S_m(\beta)/\partial\beta|$ *exists and is continuous in a neighborhood* N_0*;*

(ii) $m^{-1}\partial S_m(\beta)/\partial\beta$ *converges uniformly in probability to* $E\{m^{-1}\partial S_m(\beta)/\partial\beta\}$ *in* N_0*;*

(iii) with probability converging to 1, the quantity $\partial S_m(\beta)/\partial\beta$ *evaluated at the true parameter is negative as* $m \to \infty$*;*

(iv) $E\{S_m(\beta)\} \doteq 0$.

The next two theorems are along the lines of Theorems 3.1 and 3.2 in Pepe, Reilly, and Fleming (1994).

Theorem 2.2. *An estimator* $\tilde{\beta}_E$ *that satisfies equation (2.3.11) exists and is unique in a neighborhood* N_0*, with probability converging to 1 as* $m \to \infty$ *and* $n \to \infty$ *such that* $m/n \to k'$ *where* $k' \in (0, 1)$*. Furthermore,* $\tilde{\beta}_E$ *is consistent for the true parameter* β*.*

Proof. First, we assume that the p_i's are fixed and let $I_\beta(Y|X) = -\partial^2 \log L(\beta|X_1, \dots, X_m)/\partial\beta^2$. In this case, the score function is $S_m(\beta) = m\sum S_\beta(Y_i|X_i)p_i$, where $S_\beta(Y_i|X_i) = X_i[Y_i - F(\beta X_i)]$. Condition (i) of Foutz above follows from assumption (b).

Consider $m^{-1}\partial S_m(\beta)/\partial\beta) \doteq -m^{-1}\sum_{i\in S_m} x_i^2 F(\beta x_i)[1 - F(\beta x_i)]$ which is the average of independent and non-identically distributed random variables.

Then, the Kolmogorov strong law of large numbers for independent non-identically distributed random variables applies and yields $m^{-1}(\partial S_m(\beta)/\partial\beta) - \{-n^{-1}\sum_{i\in S_n} x_i^2 F(\beta x_i)[1 - F(\beta x_i)]\} \xrightarrow{p} 0$ as $m \to \infty$ (see Serfling, 1980, p.27). The pointwise convergence of $\partial S_m(\beta)/\partial\beta$ can be extended to uniform convergence on a neighborhood, N_0. This follows by noting that the assumption (b) is satisfied for our model, i.e., $\partial S_m(\beta)/\partial\beta$ has bounded derivative in a neighborhood, N_0, almost surely and by the application of the dominated convergence theorem to establish that $E\{I_\beta(Y_i|X_i)\}$ also has bounded derivatives. The pointwise convergence of $m^{-1}\partial S_m(\beta)/\partial\beta$ at the true parameter value together with assumption (c) implies condition (iii) of Foutz.

Finally, turning to condition (iv) we note that

$$\begin{aligned} E\{S_m(\beta)\} &= E\{m \sum_{i\in S_m} S_\beta(Y_i|X_i)p_i\} \\ &\doteq \sum_{i\in S_m} E\{S_\beta(Y_i|X_i)\} \\ &= 0. \end{aligned} \tag{2.4.5}$$

□

Hence the result of Theorem 2.2 follows for p_i's fixed.

We now give the asymptotic variance of $\tilde{\beta}_E$ in the following theorem.

Theorem 2.3. *As $m \to \infty$ and $n \to \infty$ such that $m/n \to k'$ where $k' \in (0,1)$, $m^{1/2}(\tilde{\beta}_E - \beta)$ converges in distribution to a normally distributed random variable* *such that for large m and n, $E(\tilde{\beta}_E) = 0$ and variance given by*

$$
\begin{aligned}
Var(\tilde{\beta}_E) = \frac{1}{n}\left(\frac{1}{n}\sum_{i\in S_n} x_i^2 F(\beta x_i)[1 - F(\beta x_i)]\right)^{-1} \\
+\left(\frac{1}{m}\right)\left(\frac{1}{n}\sum_{i\in S_n} x_i^2 F(\beta x_i)[1 - F(\beta x_i)]\right)^{-2} \\
\left[\frac{1}{n}\sum_{i\in S_n} x_i^2 F(\beta x_i)[1 - F(\beta x_i)] - \left(\frac{1}{n}\sum_{i\in S_n} x_i z_i F(\beta x_i)[1 - F(\beta x_i)]\right)^2\right. \\
\left.\times\left(\frac{1}{n}\sum_{i\in S_n} z_i^2 F(\theta z_i)[1 - F(\theta z_i)]\right)^{-1}\right]. \qquad (2.4.6)
\end{aligned}
$$

Proof. Consider a second order Taylor series expansion of $S_m(\tilde{\beta}_E)$ around β. Then we have

$$0 = S_m(\tilde{\beta}_E) = S_m(\beta) + \frac{\partial S_m(\beta)}{\partial \beta}(\tilde{\beta}_E - \beta) + o_p(m^{-1/2}),$$

so that

$$m^{1/2}(\tilde{\beta}_E - \beta) = [-m^{-1}\frac{\partial S_m(\beta)}{\partial \beta}]^{-1}\{m^{-1/2}S_m(\beta)\} + o_p(1).$$

Observe that $m^{-1}\partial S_m(\beta)/\partial\beta \doteq m^{-1}\sum_{i\in S_m} -I_\beta(Y_i|X_i)$ which is the mean of independent and non-identically distributed random variables. It was previously proven that $m^{-1}\partial S_m(\beta)/\partial\beta - \{-n^{-1}\sum_{i\in S_n} x_i^2\ F(\beta x_i)[1 - F(\beta x_i)]\} \xrightarrow{p} 0$, which is negative. Therefore, we look at the asymptotic distribution of $S_m(\beta)$. Consider

$$S_m(\beta) = m\sum_{i\in S_m} S_\beta(Y_i|X_i)p_i, \qquad (2.4.7)$$

which is the sum of independent and non-identically distributed random variables with mean 0. Asymptotic normality of $S_m(\beta)$ follows then from the Lindeberg-Feller central limit theorem by noting that $X_i^* = mX_i[Y_i - F(\beta X_i)]p_i$ with $EX_i^* =$

0 and for some $\nu > 2$, $B_m^{-\nu} \sum_{i=1}^m E|X_i^*|^\nu = o(1)$ where $B_m^2 = \sum_{i=1}^m E(X_i^*)^2$. Upon simplification, it is shown in the Lemma below that the variance of $S_m(\beta)$ is given by expression (2.4.9). It follows that the asymptotic distribution of $m^{1/2}(\tilde{\beta}_E - \beta)$ is normal with mean 0 and variance given by

$$Var(\sqrt{m}\tilde{\beta}_E) \doteq \left[-m^{-1}\frac{\partial S_m(\beta)}{\partial \beta}\right]^{-2} m^{-1}Var(S_m(\beta)), \tag{2.4.8}$$

which is the desired result. □

Lemma 2.1. *For large m and n, we have*

$$\begin{aligned} Var\{S_m(\beta)\} \doteq \left(\frac{m}{n}\right)^2 \sum_{i \in S_n} x_i^2 F(\beta x_i)[1 - F(\beta x_i)] + m \Bigg\{ \frac{1}{n} \sum_{i \in S_n} x_i^2 F(\beta x_i) \\ \times [1 - F(\beta x_i)] - \left(\frac{1}{n} \sum_{i \in S_n} x_i z_i F(\beta x_i)[1 - F(\beta x_i)]\right)^2 \\ \times \left(\frac{1}{n} \sum_{i \in S_n} z_i^2 F(\theta z_i)[1 - F(\theta z_i)]\right)^{-1} \Bigg\}. \end{aligned} \tag{2.4.9}$$

Proof. To prove the lemma, we use the standard formula,

$$Var\{S_m(\beta)\} = E[Var\{S_m(\beta)\}|S_n] + Var[E\{S_m(\beta)\}|S_n]. \tag{2.4.10}$$

First, consider

$$\begin{aligned} S_m(\beta) &= m \sum_{i \in S_m} x_i[y_i - F(\beta x_i)]p_i \\ &\doteq \frac{m}{m} \sum_{i \in S_m} x_i[y_i - F(\beta x_i)][1 - \lambda_2 S_{\hat{\theta}}(Y_i|Z_i)], \end{aligned} \tag{2.4.11}$$

and observe that by the strong law of large numbers $n^{-1}\sum_{i \in S_n} S_\theta(Y_i|Z_i) \xrightarrow{p} 0$, since $E[S_\theta(Y_i|Z_i)] = 0$ and then by noting that $S_{\hat{\theta}}(Y|Z) \xrightarrow{p} S_\theta(Y|Z)$, we have

$$E\{S_m(\beta)|S_n\} \doteq \frac{m}{n} \sum_{i \in S_n} x_i[y_i - F(\beta x_i)]. \tag{2.4.12}$$

It follows that

$$Var[E\{S_m(\beta)\}|S_n] \doteq \left(\frac{m}{n}\right)^2 \sum_{i\in S_n} x_i^2 F(\beta x_i)[1 - F(\beta x_i)]. \tag{2.4.13}$$

Turning to $E[Var\{S_m(\beta)\}|S_n]$, consider

$$\begin{aligned} S_m(\beta) &= m \sum_{i\in S_m} x_i[y_i - F(\beta x_i)]p_i \\ &\doteq m\left(\frac{1}{m}\sum_{i\in S_m} y_i^* - \frac{\hat{\lambda}_2}{m}\sum_{i\in S_m} S_{\hat{\theta}}(Y_i|Z_i)y_i^*\right) \\ &\doteq m\left(\bar{y}_m^* - \frac{\bar{x}^*}{m\hat{S}_{x^*}^2}\sum_{i\in S_m} x_i^* y_i^*\right) \\ &= m(\bar{y}_m^* - \hat{\beta}^*\bar{x}_m^*), \end{aligned} \tag{2.4.14}$$

where

$$y_i^* = x_i[y_i - F(\beta x_i)], \quad \bar{y}_m^* = m^{-1}\sum y_i^*, \tag{2.4.15}$$

$$x_i^* = z_i(y_i - F(\theta z_i)], \quad \bar{x}_m^* = m^{-1}\sum x_i^*, \tag{2.4.16}$$

$$\hat{\beta}^* = (m\hat{S}_{x^*}^2)^{-1}\sum x_i^* y_i^*, \quad \hat{S}_{x^*}^2 = m^{-1}\sum x_i^{*2}, \tag{2.4.17}$$

i.e. the empirical likelihood estimator is asymptotically equivalent to the regression estimator $\hat{\beta}_{lr}$, in the sense that $m^{1/2}(\tilde{\beta}_E - \hat{\beta}_{lr}) = o_p(1)$ (see Hartley and Rao, 1967), and $\hat{\lambda}_2$ satisfy (2.3.10). Along the lines of argument given in Chen and Qin (1993), it can be shown that

$$\begin{aligned} \hat{\lambda}_2 &= \left[\frac{1}{m}\sum_{i\in S_m} S_{\hat{\theta}}^2(Y_i|Z_i)\right]^{-1}\left[\frac{1}{m}\sum_{i\in S_m} S_{\hat{\theta}}(Y_i|Z_i)\right] + o_p(m^{-1/2}) \\ &= O_p(m^{-1/2}). \end{aligned} \tag{2.4.18}$$

Hence, the conditional variance of $S_m(\beta)|S_n$ is given by

$$\begin{aligned} Var\{S_m(\beta)|S_n\} &\doteq m^2\left(\frac{1}{m}\right)(1-\hat{\rho}^2_{y^*x^*})\hat{S}^2_{y^*} \\ &= m^2\left(\frac{1}{m}\right)\left(\hat{S}^2_{y^*}-\frac{\hat{S}^2_{y^*x^*}}{\hat{S}^2_{x^*}}\right), \qquad (2.4.19) \end{aligned}$$

where $\hat{\rho}_{y^*x^*}$ and $\hat{S}_{y^*x^*}$ are the correlation coefficient and covariance between y^* and x^*, respectively and $\hat{S}^2_{y^*}$ is the variance of y^*. Taking expectation of (2.4.19), we have

$$\begin{aligned} E[Var\{S_m(\beta)|S_n\}] \doteq m &\left\{\frac{1}{n}\sum_{i\in S_n} x_i^2F(\beta x_i)[1-F(\beta x_i)]\right. \\ &-\left(\frac{1}{n}\sum_{i\in S_n} x_iz_iF(\beta x_i)[1-F(\beta x_i)]\right)^2 \\ &\left.\times\left(\frac{1}{n}\sum_{i\in S_n} z_i^2F(\theta z_i)[1-F(\theta z_i)]\right)^{-1}\right\}. \qquad (2.4.20) \end{aligned}$$

Combining (2.4.13) and (2.4.20), we get

$$\begin{aligned} Var\{S_m(\beta)\} \doteq \left(\frac{m}{n}\right)^2 \sum_{i\in S_n} x_i^2F(\beta x_i)[1-F(\beta x_i)] + m&\left\{\frac{1}{n}\sum_{i\in S_n} x_i^2F(\beta x_i)\right. \\ \times[1-F(\beta x_i)] - \left(\frac{1}{n}\sum_{i\in S_n} x_iz_iF(\beta x_i)[1-F(\beta x_i)]\right)^2& \\ \times\left(\frac{1}{n}\sum_{i\in S_n} z_i^2F(\theta z_i)[1-F(\theta z_i)]\right)^{-1}&\left.\vphantom{\frac{1}{n}}\right\}. \qquad (2.4.21) \end{aligned}$$

□

We note that the first component of the variance $Var(\tilde{\beta}_E)$ is the expected information based on $L(\beta|X_1,\ldots,X_n)$, the likelihood for observed data if

$\{X_i, i \in S_n\}$ were known. The second term is, therefore, the penalty induced for not knowing $\{X_i, i \in S_{n-m}\}$.

All of the above results regarding $\tilde{\beta}_E$ assume that p_i's are fixed. Define $\hat{\beta}_E$ the same as $\tilde{\beta}_E$ except the p_i's are replaced by their respective estimates

$$\hat{p}_i = \left(\frac{1}{m}\right)\frac{1}{1+\hat{\lambda}_2 S_{\hat{\theta}}(Y_i|Z_i)}, \quad i \in S_m. \tag{2.4.22}$$

Now, consider

$$m^{1/2}(\hat{\beta}_E - \beta) = m^{1/2}(\tilde{\beta}_E - \beta) + m^{1/2}(\hat{\beta}_E - \tilde{\beta}_E). \tag{2.4.23}$$

By noting that the second term of the right hand side of (2.4.22) converges to 0 in probability, we conclude that $m^{1/2}(\hat{\beta}_E - \beta)$ has the same asymptotic distribution as $m^{1/2}(\tilde{\beta}_E - \beta)$.

2.5 A Simulation study

We conduct a simulation study to investigate the properties of the estimators studied in this chapter. In particular, the absolute bias and relative efficiencies are considered for (a) the partial case estimator ($\hat{\beta}_m$), (b) the imputed estimator ($\hat{\beta}_I$), (c) bootstrap estimator ($\hat{\beta}_B$), and (d) the empirical likelihood estimator ($\hat{\beta}_E$). The values of (X, Z) are generated according to the following three different models:

(I) Linear: $X_i = 1 + 2Z_i + \varepsilon_i$ with $Z_i \overset{ind}{\sim}$ Unif(0, 2) and $\varepsilon_i \sim \sqrt{Z_i}N(0, 0.25)$ for $i \in S_n$.

(II) Quadratic: $X_i = 2.0 - 2.0Z_i + 2.0Z_i^2 + \varepsilon_i$ with Z_i and ε_i are defined as in Model I for $i \in S_n$.

(III) Linear-Quadratic: $X_i = 1 + 2Z_i + \varepsilon_i$ for $i \in S_m$ and $X_i = 1 + 2Z_i^2 + \varepsilon_i$ for $i \in S_{n-m}$ with Z_i and ε_i are defined as in Model I.

Other parameters covered in this study are $\beta = 1$ and $n = 5,000$. The outcome Y is generated from a Bernoulli random variable conditional on X, whereby $Y = 1$ with probability $(1+\exp\{-\beta x\})^{-1}$ and $Y = 0$ otherwise. Furthermore, for each combination of the above parameter values, $N = 200$ independent samples are generated for $n = 5,000$ according to the above three models. A simple random sampling of size $m = 100, 200$ and 500 are then taken without replacement from these populations of n elements. This sampling is repeated for $S = 50$ times. To run the bootstrap procedure with 200 simulations and 100 bootstrap samples with $S = 50$ on a Sun SPARC station 20 model 712, with dual 75 MHZ super SPARC CPU's and 192 MB of RAM, 600 hours of CPU time are required. Hence, to minimize the computer time, the bootstrap method is implemented only for the Model III. The Mak's bootstrap estimator is computed based on $B = 100$ bootstrap samples. The absolute bias for each estimator is computed using the following formula:

$$\text{Absolute bias}(\hat{\beta}_{s,n}) = \frac{1}{NS}\sum_{n=1}^{N}\sum_{s=1}^{S}|\hat{\beta}_t^{s,n} - \beta|, \tag{2.5.1}$$

$$\tag{2.5.2}$$

where $\hat{\beta}_t^{s,n}$ is the estimator based on the method $t = m, I, B$ and E. The relative efficiencies of $\hat{\beta}_m$, $\hat{\beta}_B$ and $\hat{\beta}_E$ are defined as the ratios of the mean square error of $\hat{\beta}_I$ to their respective mean square errors., i.e.,

$$\text{Relative Efficiency}(\hat{\beta}_{s,n}) = \frac{\text{MSE}(\hat{\beta}_I^{s,n})}{\text{MSE}(\hat{\beta}_t^{s,n})}, \text{ where} \tag{2.5.3}$$

$$\text{MSE}(\hat{\beta}_t^{s,n}) = \frac{1}{NS}\sum_{n=1}^{N}\sum_{s=1}^{S}(\hat{\beta}_t^{s,n} - \beta)^2, \tag{2.5.4}$$

The results are reported in Tables 2.1–2.3. The simulation results show the advantages of the empirical likelihood approach over its competitors on both grounds: biasedness and efficiency. The absolute bias values under the Model I are in the range of 0.068–0.167 for $\hat{\beta}_m$, 0.348–0.381 for $\hat{\beta}_I$ and 0.062–0.149 for $\hat{\beta}_E$. While the efficiencies relative to $\hat{\beta}_I$ are in the range of 3.003–17.720 for $\hat{\beta}_m$ and 3.973–21.857 for $\hat{\beta}_E$. Similarly, the absolute bias and efficiencies values under Model II have the similar pattern as in Model I. The absolute bias values under the Model III are in the range of 0.079–0.155 for $\hat{\beta}_m$, 0.140–0.160 for $\hat{\beta}_I$, 0.055–0.094 for $\hat{\beta}_B$ and 0.036–0.076 for $\hat{\beta}_E$. While the efficiencies relative to $\hat{\beta}_I$ are in the range of 0.599–2.039 for $\hat{\beta}_m$, 1.514–4.146 for $\hat{\beta}_B$ and 2.753–10.250 for $\hat{\beta}_E$. There is a negligible absolute bias in $\hat{\beta}_E$ with greater relative efficiency over all other estimators.

The empirical likelihood estimator performs well even when the relationship between X and Z is not linear for $i \in S_{n-m}$, whereas, the performance of the bootstrap estimator in this case is rather poor. These results are encouraging and imply that the proposed method is robust to misspecification of the relationship between X and Z. It is interesting to note that the bootstrap estimator still performs better than the imputed estimator.

In summary, the method we have proposed is a useful alternative to the standard procedure and it is not as computationally intensive as the bootstrap method. The computations can be carried out with an existing maximization subroutine such as Brent's method.

Table 2.1: Absolute bias and efficiencies of three estimators relative to $\hat{\beta}_I$ for $n = 5,000$, $S = 50$ and $N = 200$ under linear model $X_i = 1.0 + 2.0Z_i + \varepsilon_i$ for $i \in S_n$.

	Absolute bias			Rel. Efficiency	
m	$\hat{\beta}_m$	$\hat{\beta}_I$	$\hat{\beta}_E$	$\hat{\beta}_m$	$\hat{\beta}_E$
100	0.167	0.381	0.149	3.003	3.973
200	0.118	0.348	0.102	4.838	6.586
500	0.068	0.355	0.062	17.720	21.857

Table 2.2: Absolute bias and efficiencies of three estimators relative to $\hat{\beta}_I$ for $n = 5,000$, $S = 50$ and $N = 200$ under quadratic model $X_i = 2.0 - 2.0Z_i + 2.0Z_i^2 + \varepsilon_i$ for $i \in S_n$.

	Absolute bias			Rel. Efficiency	
m	$\hat{\beta}_m$	$\hat{\beta}_I$	$\hat{\beta}_E$	$\hat{\beta}_m$	$\hat{\beta}_E$
100	0.195	0.357	0.171	1.521	1.915
200	0.112	0.371	0.095	6.745	9.351
500	0.074	0.317	0.063	11.121	15.143

Table 2.3: Absolute bias and efficiencies of three estimators relative to $\hat{\beta}_I$ for $n = 5,000$, $S = 50$, $B = 100$ and $N = 200$ under linear model $X_i = 1 + 2Z_i + \varepsilon_i$ for $i \in S_m$ and quadratic model $X_i = 1 + 2Z_i^2 + \varepsilon_i$ for $i \in S_{n-m}$.

	Absolute bias				Rel. Efficiency		
m	$\hat{\beta}_m$	$\hat{\beta}_I$	$\hat{\beta}_B$	$\hat{\beta}_E$	$\hat{\beta}_m$	$\hat{\beta}_B$	$\hat{\beta}_E$
100	0.155	0.160	0.094	0.076	0.599	1.514	2.753
200	0.122	0.153	0.071	0.049	0.996	2.693	6.513
500	0.079	0.140	0.055	0.036	2.039	4.146	10.25

Chapter 3

Variance estimation under two-phase sampling

3.1 Introduction

In survey sampling, we are often interested in estimating the mean value of a characteristic Y of a particular population when the information on the auxiliary variable X, correlated with Y, is already available or can be easily observed. For such situations, the estimation for the mean value of Y through ratio and regression techniques has been discussed in the literature for two different cases (see Cochran, 1977) (1) Single-phase case: When the population mean of the characteristics X is already known and the information on Y is observed for units in a sample of size n; (2) Two-phase case: When the population mean of the characteristics X is not known and it is estimated by taking a large random sample of size n' and observing X, then drawing a subsample of size n observing Y.

Two-phase sampling is generally employed when it is economically feasible to take a large preliminary sample in which an auxiliary variable X, correlated with a characteristic of interest Y, is measured alone. The initial sample gives an estimate $\bar{x}_{n'}$ of the population mean $\bar{X}$, while the subsample in which Y is measured is employed to estimate the population mean $\bar{Y}$ through ratio or regression estimation using $\bar{x}_{n'}$. For example, in a survey that estimates the total wheat yield in a given locality in Canada, one might use a large sample of n' farms to estimate the total area under wheat cultivation and a subsample of n farms to determine the actual yield.

Chen and Qin (1993) employed the empirical likelihood method to use summary information on the auxiliary variable at the estimation stage. Benhin and Prasad (1997) extended the empirical likelihood to double sampling when two auxiliary variables were available.

Turning to variance estimation under the ratio method, Rao and Sitter (1995) proposed a new linearization variance estimator for a ratio estimator that made better use of the sample data than the standard textbook formula. They also obtained a jackknife variance estimator and concluded through a simulation study that their conditional and unconditional variances had better properties than the standard formula (see Sukhatme and Sukhatme,1970). Subsequently, Sitter (1997) extended this method to regression estimation along the same lines as ratio estimation. He showed under a model proposed by Dorfman (1994) that the resulting variance estimators were design-unbiased and approximately model-unbiased. For more information on variance estimation under two-phase sampling under model-based approach, see Dorfman (1994).

This chapter considers a new variance estimator for ratio estimation based on the empirical likelihood approach under simple random sampling without replacement both in single and two-phase sampling. We will use this approach to choose the probability weights under constraints formulated from the information on the auxiliary variable.

In Section 3.2 and Section 3.3, we review variance estimators available for ratio estimation in single and double sampling. The proposed variance estimator is derived under each case using empirical likelihood. We extend the empirical likelihood method to regression estimation in Section 3.4. A simulation study to examine the unconditional and conditional repeated sampling properties of the proposed variance estimator in two phase sampling is presented in Section 3.5.

3.2 Variance estimator of the ratio estimator under single-phase sampling

Suppose that a population consists of N distinct units with values (y_i, x_i), where $x_i > 0$ $(i = 1, \ldots, N)$. Denote the population means of Y and X, respectively, by $\bar{Y}$ and $\bar{X}$. To estimate $\bar{Y}$ under simple random sampling of size n , it is customary to use the ratio estimator $\bar{y}_{rs} = (\bar{y}/\bar{x})\bar{X}$, where $\bar{y}$ and $\bar{x}$ are the sample means of y and x. The variance of $\bar{y}_{rs}$ is approximated by (Cochran, 1977, p. 155) and given by

$$V(\bar{y}_{rs}) \doteq \left(\frac{1}{n} - \frac{1}{N}\right) S_D^2, \tag{3.2.1}$$

where $S_D^2 = (N-1)^{-1}\sum_{i=1}^{N} D_i^2$, $D_i = Y_i - RX_i$ with $R = \bar{Y}/\bar{X}$. Two commonly used estimators of $V(\bar{y}_{rs})$ are

$$v_0(\bar{y}_{rs}) = \left(\frac{1}{n} - \frac{1}{N}\right) s_d^2 \quad \text{and} \quad v_1(\bar{y}_{rs}) = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{\bar{X}}{\bar{x}}\right)^2 s_d^2, \tag{3.2.2}$$

where

$$s_d^2 = \frac{1}{n-1}\sum_{i\in s} d_i^2 = s_y^2 - 2rs_{xy} + r^2 s_x^2, \qquad (3.2.3)$$

with $r = \bar{y}/\bar{x}$, $d_i = y_i - rx_i$. Although the original motivation for $v_1(r) = v_1(\bar{y}_{rs})/\bar{X}^2$ as a variance estimator of the ratio R is the unavailability of $\bar{X}$, it is not clear that $v_1(\bar{y}_{rs})$ is indeed worse than $v_0(\bar{y}_{rs})$ (see Cochran, 1977 and Rao and Rao, 1971). Chen and Qin (1993) applied the empirical likelihood approach in conjunction with summary information on the auxiliary variable in improving the customary estimator under simple random sampling. They showed that the empirical likelihood estimator was asymptotically equivalent to the linear regression estimator when the population mean of the auxiliary variable was known (see Hartley and Rao, 1968). To use the empirical likelihood method as described in the previous chapter, we maximize the empirical likelihood

$$L(F) = \prod_{i=1}^{n} p_i, \qquad (3.2.4)$$

where $p_i = P(Y = y_i)$. The p_i's are subject to $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$. With these weights an empirical likelihood estimator for S_D^2 is given by

$$s_d^2(el) = \sum_{i=1}^{n} p_i d_i^2. \qquad (3.2.5)$$

The resulting empirical likelihood variance estimator for $\bar{y}_{rs}$ is

$$v_3(\bar{y}_{rs}) = \left(\frac{1}{n} - \frac{1}{N}\right) s_d^2(el). \qquad (3.2.6)$$

3.3 The two-phase sampling procedure

Assume that a simple random sample s' of size n' is selected without replacement from a population of N units and x_i is observed for $i \in s'$. A simple random subsample s of size n is then selected without replacement from s' and y_i is observed for $i \in s$. Several estimates of $\bar{Y} = \sum_{i=1}^{N} Y_i/N$ can be formed. The simplest is the usual biased ratio estimate with the population mean $\bar{X}$ replaced by its estimates $\bar{x}_{n'}$, given by $\bar{y}_{rt} = (\bar{y}_n/\bar{x}_n)\bar{x}_{n'} = r\bar{x}_{n'}$, where $\bar{y}_n$ and $\bar{x}_n$ are the means for s and $\bar{x}_{n'}$ for s'.

3.3.1 The ratio estimator and some preliminary results

The estimator $\bar{y}_{rt}$ is design-consistent for $\bar{Y}$, i.e., $p\lim_{\substack{\pi \\ n\to\infty}}(\bar{y}_{rt} - \bar{Y}) = 0$, where π denotes the probability space generated by the sampling scheme. The variance of $\bar{y}_{rt}$ is approximated by a standard formula and is given by

$$V(\bar{y}_{rt}) \doteq \left(\frac{1}{n} - \frac{1}{n'}\right) S_D^2 + \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2, \tag{3.3.1}$$

where

$$S_D^2 = \frac{1}{N-1}\sum_{i=1}^{N} D_i^2 = S_y^2 - 2RS_{xy} + R^2 S_x^2, \quad S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2,$$

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2, \quad S_{xy}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y}),$$

with $D_i = y_i - Rx_i$ and $R = \bar{Y}/\bar{X}$. Note that if $n' = n$, $(\bar{y}_n/\bar{x}_n)\bar{x}_{n'} = \bar{y}_n$ and so (3.3.1) reduces to $(1/n - 1/N)S_y^2$, which is the variance of $\bar{y}_n$ under simple random sampling in single phase sampling. It is observed that if $n' = N$, the estimator is the ratio estimator under single-phase sampling, and the variance reduces to the approximate formula for its variance. It follows that the estimate $\bar{y}_{rt}$ based on two-phase sampling is more efficient than the estimate $\bar{y}_n$ based on simple random sampling when no auxiliary variable is used, if

$$R^2S_x^2 - 2RS_{xy} < 0,$$

i.e., if

$$\rho_{xy}\frac{C_y}{C_x} > \frac{1}{2},$$

where ρ_{xy} is the population correlation coefficient between x and y, and C_x and C_y are population coefficients of variation of x and y, respectively. A design-consistent estimator of the variance estimator of $\bar{y}_{rt}$ is given by

$$v_0(\bar{y}_{rt}) = \left(\frac{1}{n} - \frac{1}{n'}\right)s_d^2 + \left(\frac{1}{n'} - \frac{1}{N}\right)s_y^2, \tag{3.3.2}$$

where

$$s_d^2 = \frac{1}{n-1}\sum_{i\in s} d_i^2, \quad s_y^2 = \frac{1}{n-1}\sum_{i\in s}(y_i - \bar{y}_n)^2, \tag{3.3.3}$$

with $d_i = y_i - rx_i$. The second term in $v_0(\bar{y}_{rt})$ is obtained by using the sample variance s_y^2 to estimate the population variance S_y^2.

3.3.2 Linearization variance estimator

Rao and Sitter (1995) proposed a linearization variance estimator of $\bar{y}_{rt}$ that made better use of the sample data than the standard one, v_0. They first expressed S_y^2 as

$$\begin{aligned} S_y^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(y_i - R\bar{X})^2 \\ &= \frac{1}{N-1}\sum_{i=1}^{N}(y_i - Rx_i + Rx_i - R\bar{X})^2 \\ &= \frac{1}{N-1}\sum_{i=1}^{N}\left\{(y_i - Rx_i)^2 + 2R(y_i - Rx_i)(x_i - \bar{X}) + R^2S_x^2\right\} \\ &= S_D^2 + 2RS_{Dx} + R^2S_x^2, \qquad (3.3.4) \end{aligned}$$

where S_D^2 and S_x^2 were the corresponding population variances of D_i and x_i, S_{Dx} was the population covariance between D_i and x_i. Then the sample variance of s_y^2 can be written as

$$s_y^2 = s_d^2 + 2rs_{dx} + r^2s_x^2, \qquad (3.3.5)$$

where s_d^2, s_{dx} and s_x^2 are the sample analogues of S_D^2, S_{Dx} and S_x^2 based on subsample s. It follows from (3.3.4) and (3.3.5) that an alternative estimator of S_y^2 that makes more use of the sample data is obtained by using

$$s_x'^2 = \frac{1}{n'-1}\sum_{i\in s'}(x_i - \bar{x}_{n'})^2$$

in place of s_x^2. The linearization variance estimator of $\bar{y}_{rt}$ is

$$v_1(\bar{y}_{rt}) = \left(\frac{1}{n} - \frac{1}{N}\right)s_d^2 + 2\left(\frac{1}{n'} - \frac{1}{N}\right)rs_{dx} + \left(\frac{1}{n'} - \frac{1}{N}\right)r^2s_x'^2. \qquad (3.3.6)$$

This variance estimator is also design-consistent. We can rewrite (3.3.1) using (3.3.5) as

$$v_0(\bar{y}_{rt}) = \left(\frac{1}{n} - \frac{1}{N}\right)s_d^2 + 2\left(\frac{1}{n'} - \frac{1}{N}\right)rs_{dx} + \left(\frac{1}{n'} - \frac{1}{N}\right)r^2s_x^2. \qquad (3.3.7)$$

3.3.3 Jackknife variance estimator

Another approach to variance estimation is to use a jackknife technique. Rao and Sitter (1995) proposed a jackknife method which entailed recalculating $\bar{y}_{rt}$ with the jth element removed for each $j \in s'$ and then using the variance of these n' jackknife values, $\bar{y}_{rt}(j)$. Clearly, deleting unit j will affect $\bar{x}_n$ and $\bar{y}_n$ only for $j \in s$ but not for $j \in s' - s$, while it will affect $\bar{x}_{n'}$ for all $j \in s'$. Thus, they defined

$$\bar{y}_{rt}(j) = \{\bar{y}_n(j)/\bar{x}_n(j)\}\bar{x}_{n'}(j)$$

for all $j \in s'$, where

$$\bar{x}_n(j) = \begin{cases} \frac{n\bar{x}_n - x_j}{n-1} & \text{if} \quad j \in s \\ \bar{x}_n & \text{if} \quad j \in s' - s, \end{cases} \qquad \bar{y}_n(j) = \begin{cases} \frac{n\bar{y}_n - y_j}{n-1} & \text{if} \quad j \in s \\ \bar{y}_n & \text{if} \quad j \in s' - s, \end{cases}$$

and $\bar{x}_{n'}(j) = (n'\bar{x}_{n'} - x_j)/(n'-1)$ for all $j \in s'$. Now the usual jackknife method to $\bar{y}_{rt}(j)$ will yield the following variance estimator:

$$v_J(\bar{y}_{rt}) = \frac{n'-1}{n'} \sum_{j \in s'} \{\bar{y}_{rt}(j) - \bar{y}_{rt}\}^2. \tag{3.3.8}$$

This jackknife estimator ignores the finite population corrections $1 - n/N$ and $1 - n'/N$.

For a nonlinear parameter $\theta = g(\bar{Y})$, a jackknife variance estimator is obtained by replacing $\bar{y}_{rt}(j)$ and $\bar{y}_{rt}$ in (3.3.8) by $\hat{\theta}_{rt}(j) = g(\bar{Y}_{rt}(j))$ and $\hat{\theta}_{rt} = g(\bar{y}_{rt})$. A linearized version of v_J, for large n, is obtained by noting that

$$\bar{y}_{rt}(j) - \bar{y}_{rt} = \begin{cases} -r\left(\frac{x_j - \bar{x}_{n'}}{n'-1}\right) - \frac{\bar{x}_{n'}(j)}{\bar{x}_n(j)}\left(\frac{y_j - rx_j}{n-1}\right) & \text{if} \quad j \in s \\ \\ -r\left(\frac{x_j - \bar{x}_{n'}}{n'-1}\right) & \text{if} \quad j \in s' - s, \end{cases} \tag{3.3.9}$$

and assuming $\bar{x}_{n'}(j)/\bar{x}_n(j) \doteq \bar{x}_{n'}/\bar{x}_n$ in (3.3.9). From (3.3.8) and (3.3.9), we get

$$v_J(\bar{y}_{rt}) \doteq \left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)^2 \frac{s_d^2}{n} + 2\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right) \frac{r s_{dx}}{n'} + \frac{r^2 s_x'^2}{n'}. \tag{3.3.10}$$

Ignoring the finite population corrections and comparing (3.3.9) and (3.3.10), it now follows that v_J is also design-consistent for $V(\bar{y}_{rt})$ since $\bar{x}_{n'}/\bar{x}_n \doteq 1$ for large n. It also follows from (3.3.9) and (3.3.10) that another design-consistent linearization variance estimator, when the finite population corrections are not ignorable, is given by

$$v_2(\bar{y}_{rt}) \doteq \left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)^2 \left(\frac{1}{n} - \frac{1}{N}\right) s_d^2 + 2\left(\frac{1}{n'} - \frac{1}{N}\right)\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right) r s_{dx} + \left(\frac{1}{n'} - \frac{1}{N}\right) r^2 s_x'^2. \tag{3.3.11}$$

Rao and Sitter (1995) noted that if the finite population corrections could be ignored, v_J should perform well conditionally given, $\bar{x}_{n'}/\bar{x}_n$, since it was asymptotically equivalent to v_2.

In the next section, we propose two alternative variance estimators. One of them is a modification of v_0 while the other one is suggested by the empirical likelihood principle. Both utilize the information collected in the first phase as supplementary information in order to improve the precision of variance estimator of population characteristics.

3.3.4 The empirical likelihood for the double sampling

Since no auxiliary information is available beyond the initial sample s', we maximize the empirical conditional likelihood given by

$$L(s|s') = \prod_{i \in s} p_i \tag{3.3.12}$$

subject to

$$p_i \geq 0, \quad \sum_{i \in s} p_i = 1 \quad \text{and} \quad \sum_{i \in s} p_i w_i = 0, \tag{3.3.13}$$

with $w_i = x_i - \bar{x}_{n'}$. Then the empirical likelihood-based estimator for the sample variance s_y^2 and sample covariance s_{xy} are obtained by replacing n^{-1} with the p_i's in the plug-in estimator, i.e.

$$s_y^2(el) = \sum_{i \in s} p_i (y_i - \bar{y})^2, \tag{3.3.14}$$

$$s_{xy}(el) = \sum_{i \in s} p_i (x_i - \bar{x})(y_i - \bar{y}). \tag{3.3.15}$$

We use arguments of Rao and Sitter (1995) to obtain the variance estimator of $\bar{y}_{rt}$. First we observe that $\sum_{i=1}^{N} D_i = 0$ and that S_D^2 is expressed as

$$\begin{aligned} S_D^2 &= \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - RX_i)^2 \\ &= \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y} - R(X_i - \bar{X}))^2 \\ &= \frac{1}{N-1} \sum_{i=1}^{N} \{(Y_i - \bar{Y})^2 - 2R(X_i - \bar{X})(Y_i - \bar{Y}) + R^2(X_i - \bar{X})^2\} \\ &= S_y^2 - 2RS_{xy} + R^2 S_x^2, \end{aligned} \tag{3.3.16}$$

where S_{xy} is the population covariance between x_i and y_i. Thus $v_0(\bar{y}_{rt})$ in (3.3.2) can be re-expressed as

$$v_0(\bar{y}_{rt}) = \left(\frac{1}{n} - \frac{1}{N}\right) s_y^2 - 2r \left(\frac{1}{n} - \frac{1}{n'}\right) s_{xy} + r^2 \left(\frac{1}{n} - \frac{1}{n'}\right) s_x^2. \tag{3.3.17}$$

The resulting variance estimator based on empirical likelihood is given as

$$v_4(\bar{y}_{rt}) \doteq \left(\frac{1}{n} - \frac{1}{N}\right) s_y^2(el) - 2r \left(\frac{1}{n} - \frac{1}{n'}\right) s_{xy}(el) + r^2 \left(\frac{1}{n} - \frac{1}{n'}\right) s_x'^2. \tag{3.3.18}$$

Intuitively, one would expect the variance estimator based on the empirical likelihood to be more efficient than the Rao-Sitter estimator, since it makes use of extra information, i.e., the knowledge of the mean of a subsample of x.

An alternative variance estimator of $V(\bar{y}_{rt})$ can also be obtained. We note that when the y_i's are exactly proportional to x_i's for $i = 1, \ldots, N$, i.e., $y_i = kx_i$, with k as a constant, then the variance $V(\bar{y}_{rt})$ reduces to $(1/n' - 1/N)k^2 S_x^2$, which could be estimated by $k^2(1/n' - 1/N)s_x'^2$. Putting $y_i = kx_i$ in (3.3.2), we get $v_0(\bar{y}_{rt}) = k^2(1/n' - 1/N)s_x^2$, which is less efficient than $k^2(1/n' - 1/N)s_x'^2$. In view of this, we propose a modified estimator of v_0 given by

$$v_3(\bar{y}_{rt}) = \left(\frac{1}{n} - \frac{1}{n'}\right) s_d^2 + \left(\frac{1}{n'} - \frac{1}{N}\right) s_y^2 \frac{s_x'^2}{s_x^2}. \tag{3.3.19}$$

Note that $v_3(\bar{y}_{rt})$ reduces to $k^2(1/n' - 1/N)s_x'^2$ when $y_i = kx_i$.

3.4 The regression estimator

In this section, we will consider the extension of the ratio method of estimation in two-phase sampling to the case of linear regression estimation under the empirical likelihood framework. To this end, consider the two-phase sampling scheme described in Section 3.2. The simple linear regression estimator for two-phase sampling defined by

$$\bar{y}_{lr} = \bar{y}_n + b(\bar{x}_{n'} - \bar{x}_n), \tag{3.4.1}$$

where $b = s_{xy}/s_x^2$ is the least square regression coefficient of y_i on x_i based on s. This estimator is design consistent for $\bar{Y}$. A design consistent linearization variance estimator of $\bar{y}_{lr}$ is given by the standard formula

$$v_0(\bar{y}_{lr}) = \left(\frac{1}{n} - \frac{1}{n'}\right) s^2_{d'}, + \left(\frac{1}{n'} - \frac{1}{N}\right) s^2_y \tag{3.4.2}$$

where $s^2_{d'} = (n-1)^{-1} \sum_{i \in s} d'^2_i$ and s^2_y are the sample variances of $d'_i = y_i - \bar{y} - b(x_i - \bar{x}_n)$ and y_i. Alternatively, $v_0(\bar{y}_{lr})$ can be expressed as

$$v_0(\bar{y}_{lr}) = \left(\frac{1}{n} - \frac{1}{N}\right) s^2_y - 2b\left(\frac{1}{n} - \frac{1}{n'}\right) s_{xy} + b^2 \left(\frac{1}{n} - \frac{1}{n'}\right) s^2_x. \tag{3.4.3}$$

Sitter (1997) proposed three variance estimators for regression estimation along the same lines as the ratio estimation. The linearization and jackknife variance estimators in this case are given, respectively, by

$$v_1(\bar{y}_{lr}) \;= \left(\tfrac{1}{n} - \tfrac{1}{N}\right) s^2_{d'} + \left(\tfrac{1}{n'} - \tfrac{1}{N}\right) b^2 s'^2_x, \tag{3.4.4}$$

and

$$v_J(\bar{y}_{lr}) \;\doteq \tfrac{s^2_{d'}}{n} + \tfrac{b^2 s'^2_x}{n'} + \left\{\tfrac{\bar{x}_{n'} - \bar{x}_n}{(n-1)s^2_x}\right\}^2 \textstyle\sum_{j \in s} \tfrac{d'^2_j (x_j - \bar{x}_n)^2}{(1-k_j)^2} + R, \tag{3.4.5}$$

where $k_j = 1/n + (x_j - \bar{x}_n)^2/\{(n-1)s^2_x\}$, and

$$R = \frac{2}{n}\left\{\frac{1}{n}\sum_{i \in s} \frac{d'^2_j a_j}{(1-k_j)} + \frac{b}{n'-1}\sum_{j \in s} \frac{d'_j a_j (x_j - \bar{x}_{n'})}{(1-k_j)}\right\}, \tag{3.4.6}$$

with $a_j = \{n(x_j - \bar{x}_n)(\bar{x}_{n'} - \bar{x}_n)\}/\{(n-1)s^2_x\}$.

Also, noting that the first two terms on the right hand side of (3.4.5), and comparing these to (3.4.4), a linearized version of $v_J(\bar{y}_{lr})$ when the finite population corrections are not ignorable is given by

$$\begin{aligned} v_2(\bar{y}_{lr}) = {} & \left(\frac{1}{n} - \frac{1}{N}\right) s^2_{d'} + \left(\frac{1}{n} - \frac{1}{N'}\right) \frac{b^2 s'^2_x}{n'} \\ & + \left\{\frac{\bar{x}_{n'} - \bar{x}_n}{(n-1)s^2_x}\right\}^2 \sum_{j \in s} \frac{d'^2_j (x_j - \bar{x}_n)^2}{(1-k_j)^2} + R. \end{aligned} \tag{3.4.7}$$

In a similar motivation as in Section 3.3.4, a variance estimator for $\bar{y}_{lr}$ based on the empirical likelihood approach is

$$v_3(\bar{y}_{lr}) = \left(\frac{1}{n} - \frac{1}{N}\right) s_y^2(el) - 2b\left(\frac{1}{n} - \frac{1}{n'}\right) s_{xy}(el) + b^2\left(\frac{1}{n} - \frac{1}{n'}\right) s_x'^2. \quad (3.4.8)$$

where $s_y^2(el)$ and $s_{xy}(el)$ are defined, respectively, by (3.3.14) and (3.3.15).

3.5 A Simulation study

We study the finite sample properties of various variance estimators through a simulation study. We adopt the model and parameter settings used by Rao and Sitter (1995). The model we consider is

$$y_i = \beta x_i + x_i^{1/2} \varepsilon_i,$$

where $\varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$, $x_i \overset{ind}{\sim}$ gamma(a, b) and ε_i and x_i are independent of each other. Thus the mean, the variance and coefficient of variation of x are given by $\mu_x = ab$, $\sigma^2 = ab^2$ and $C_x = \sigma_x/\mu_x = a^{-1/2}$, respectively. Furthermore, the mean and variance of y are $\mu_y = \beta\mu_x$ and $\sigma_y^2 = \beta^2\sigma_x^2 + \mu_x\sigma^2$, and the correlation between x and y is $\rho = \beta\sigma_x/\sigma_y$.

We confine our simulation study for $n = 20$, $n' = 100$, and $n = 80$, $n' = 400$. We generate $R = 10,000$ independent two-phase random samples according to the above model with $\beta = 1.0$ and $\mu_x = 100$ and σ and σ_x chosen to match specified values of ρ and C_x. Here, we ignore the finite population corrections since the two-phase samples are generated from an infinite population. The Monte Carlo estimator of true mean squared error of $\bar{y}_{rt}$, is computed using

$$MSE(\bar{y}_{rt}) = \frac{1}{R}\sum_{t=1}^{R}(\bar{y}_{rt}^{(t)} - \mu_y)^2, \quad (3.5.1)$$

where $\bar{y}_{rt}^{(t)}$ denotes the value of $\bar{y}_{rt}$ for the t-th Monte Carlo run. The Monte Carlo estimate of mean squared error of a specified estimator say v is computed using

$$MSE(v) = \frac{1}{R}\sum_{t=1}^{R}(v^{(t)} - MSE(\bar{y}_{rt}))^2. \tag{3.5.2}$$

Table 3.1 gives the values of $MSE(v)/MSE(v_0)$ for $v = v_1, \ldots, v_4$ and v_j for different values of ρ and C_x where for convenience, $v_t = v_t(\bar{y}_{rt})$, $t = 0, \ldots, 4$ and $v_J = v_J(\bar{y}_{rt})$. It is clear from Table 3.1 that v_4 is substantially more efficient than other variance estimators. On the other hand, v_3 is more efficient than v_1, v_2 and v_J only for $C_x = 1.4, 1.0, 0.5, 0.33$ and $\rho = 0.8$ and substantially more efficient than v_0 for all values of ρ and C_x. Note that v_J is more efficient than v_0 only for large $n = 80$ as the factor $\bar{x}_{n'}/\bar{x}_n \doteq \bar{x}_{n'}(j)/\bar{x}_n(j) \doteq 1$ becomes more stable.

We also investigate the conditional properties of each variance estimator along the lines of Rao and Sitter (1995). The 10,000 simulated samples are first ordered on the values of $\bar{x}_{n'}/\bar{x}_n$ and then grouped into 20 successive groups each containing $G = 1,000$ samples. For each group, the simulated conditional mean squared error of $\bar{y}_{rt}$ and conditional mean of $v_t, t = 0, \ldots, 4$ and v_J are calculated, respectively,

$$MSE_c = \frac{1}{G}\sum_{g=1}^{G}\{\bar{y}_{rt}^{(g)} - \mu_y\}^2 \quad \text{and} \quad E_c v_t = \frac{1}{G}\sum_{g=1}^{G} v_t^{(g)}. \tag{3.5.3}$$

For each of the 20 groups, the values of $E_c v_t$ for $t = 0, \ldots, 4$, Ev_J and MSE_c are plotted against the group averages of $\bar{x}_{n'}/\bar{x}_n$ for 12 selected values of ρ and

C_x. Figures 3.1–3.12 with $n' = 100$ and $n = 20$ show these results. The case $n' = 400$ and $n = 20$ produce similar plots and therefore were omitted. It is clear from these plots that v_1, v_2, v_3, v_4 and v_J perform well in tracking the conditional MSE when $\bar{x}_{n'}/\bar{x}_n$ is between 0.9 and 1.4 with v_J and v_4 slightly better, i.e. they exhibit a similar pattern to the conditional MSE. However, v_0 is able track the conditional MSE only when $\bar{x}_{n'}/\bar{x}_n$ is near 1. This means that with a balanced design, v_0 does not deviate much from the conditional MSE.

It is noticed that v_1, v_2, v_3, v_4 and v_J perform poorly in tracking the conditional MSE when $\bar{x}_{n'}/\bar{x}_n \leq 0.9$ and also when $\bar{x}_{n'}/\bar{x}_n \geq 1.4$. Whereas, v_0 leads to significant overestimation of conditional MSE when $\bar{x}_{n'}/\bar{x}_n \leq 0.9$ and lead to significant underestimation when $\bar{x}_{n'}/\bar{x}_n \geq 1.2$. Thus, all things considered, v_1, v_2, v_3, v_4 and v_J behave more closely to the conditional MSE than do v_0.

The simulation study suggests that the proposed variance estimator v_4 provides more stable standard errors for ratio estimation. It has a competitive conditional performance, having smallest unconditional MSE. The commonly used estimator v_0 fails on both grounds.

Table 3.1: Mean square error of v_1, v_2, v_3, v_4 relative to v_0.

	$n = 20,\ n' = 100$				$n = 80,\ n' = 400$			
	C_x				C_x			
ρ	1.4	1.0	0.5	0.33	1.4	1.0	0.5	0.33
	$MSE(v_1)/MSE(v_0)$							
0.9	0.51	0.54	0.63	0.67	0.52	0.55	0.63	0.66
0.8	0.73	0.77	0.84	0.88	0.72	0.76	0.84	0.88
0.7	0.82	0.86	0.93	0.94	0.85	0.86	0.92	0.94
	$MSE(v_2)/MSE(v_0)$							
0.9	0.55	0.57	0.64	0.68	0.53	0.55	0.63	0.65
0.8	0.87	0.86	0.86	0.87	0.74	0.78	0.83	0.87
0.7	0.98	0.94	0.98	0.94	0.89	0.88	0.93	0.94
	$MSE(v_J)/MSE(v_0)$							
0.9	0.89	0.77	0.73	0.74	0.61	0.58	0.65	0.67
0.8	1.62	1.24	1.01	1.00	0.85	0.86	0.87	0.89
0.7	1.86	1.35	1.18	1.07	1.03	0.98	0.95	0.96
	$MSE(v_3)/MSE(v_0)$							
0.9	0.52	0.63	0.79	0.90	0.57	0.63	0.81	0.88
0.8	0.65	0.71	0.82	0.85	0.66	0.71	0.84	0.93
0.7	0.67	0.78	0.88	0.92	0.73	0.75	0.90	0.95
	$MSE(v_4)/MSE(v_0)$							
0.9	0.42	0.48	0.59	0.64	0.46	0.49	0.60	0.63
0.8	0.60	0.66	0.75	0.78	0.61	0.66	0.78	0.85
0.7	0.64	0.74	0.85	0.88	0.71	0.74	0.87	0.91

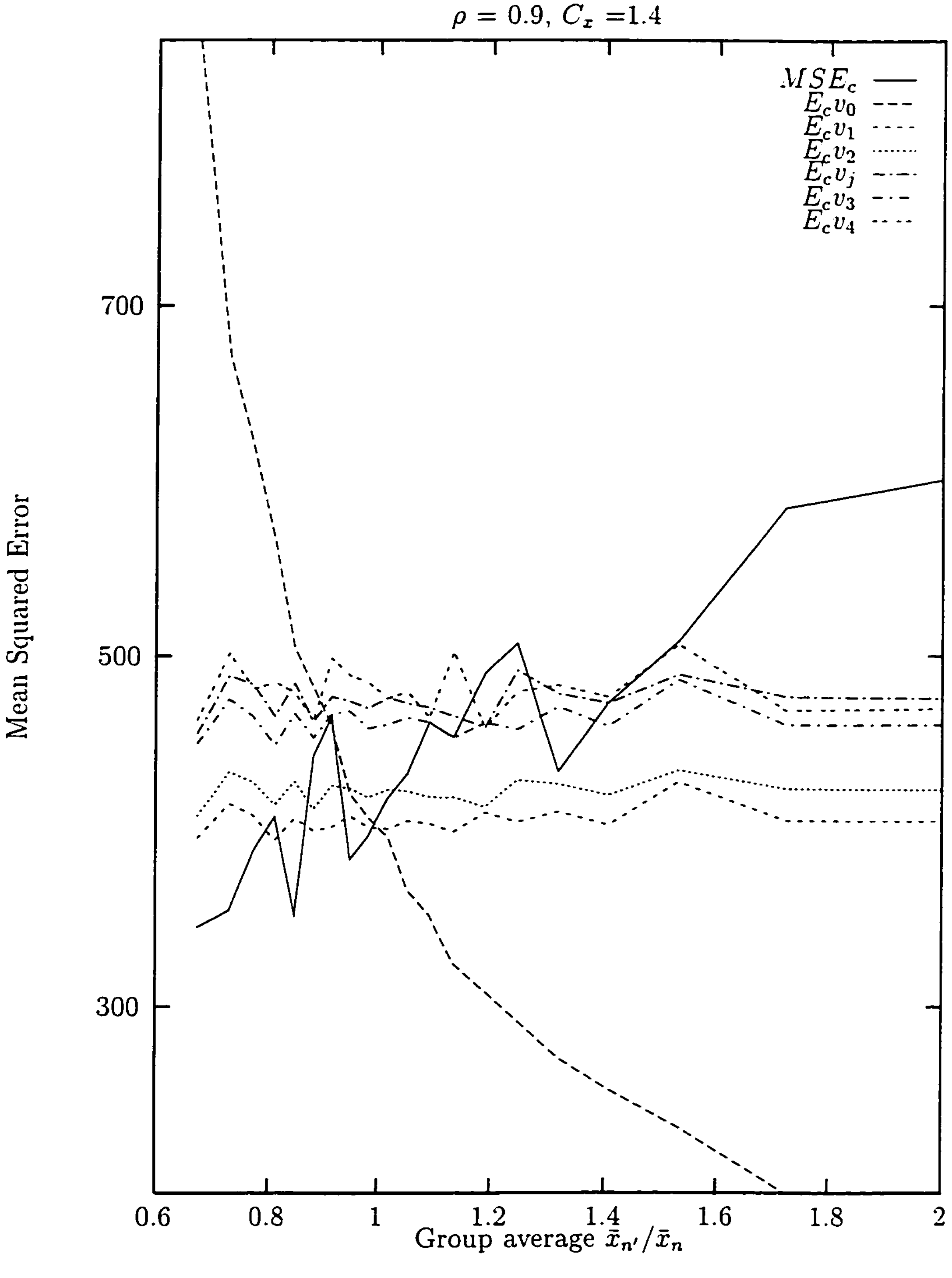


Figure 3.1: Conditional means $E_c v_0$, $E_c v_1$, $E_c v_2$, $E_c v_3$, $E_c v_4$, $E_c v_j$ and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and n = 20.

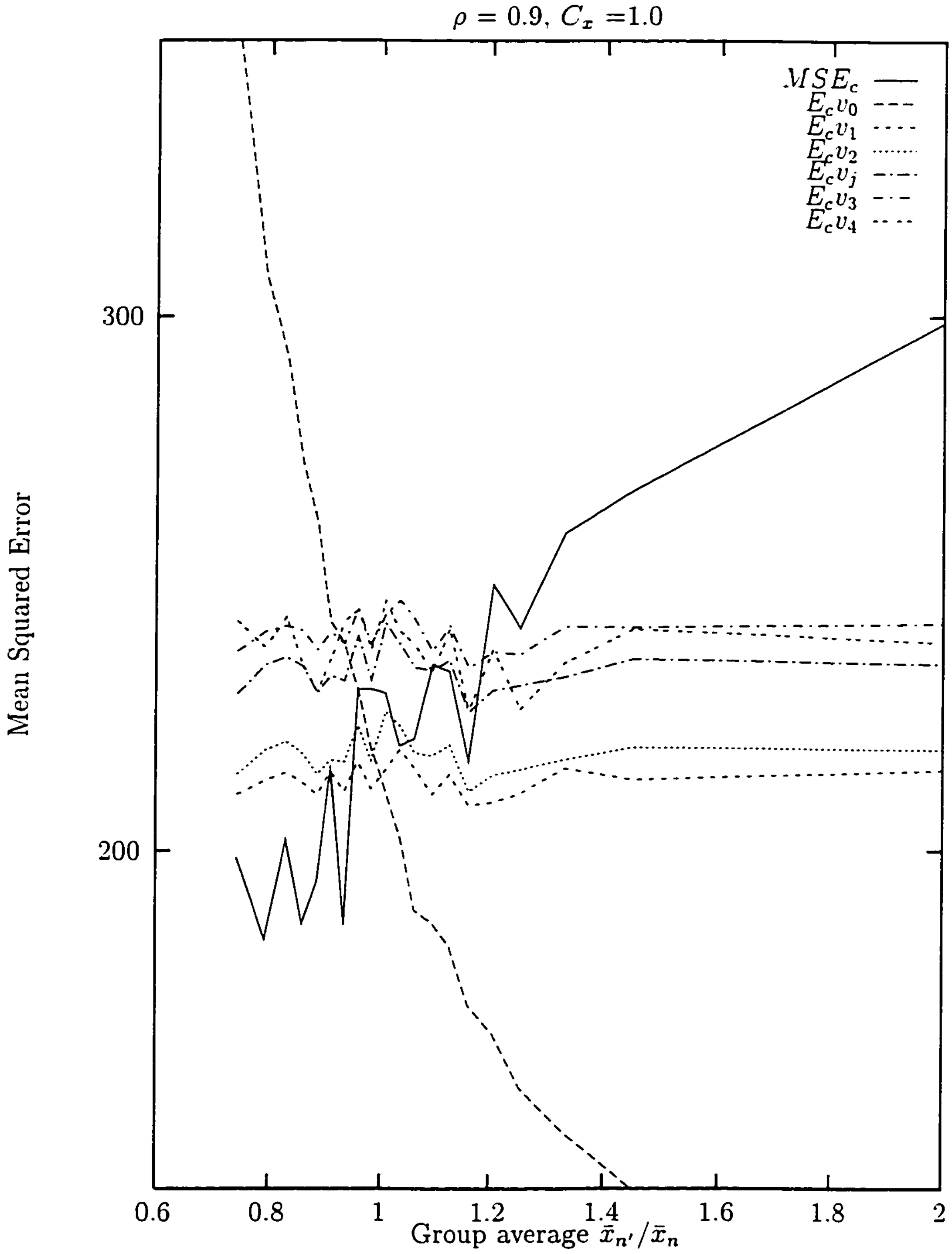


Figure 3.2: Conditional means $E_c v_0$, $E_c v_1$, $E_c v_2$, $E_c v_3$, $E_c v_4$, $E_c v_j$ and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and $n = 20$.

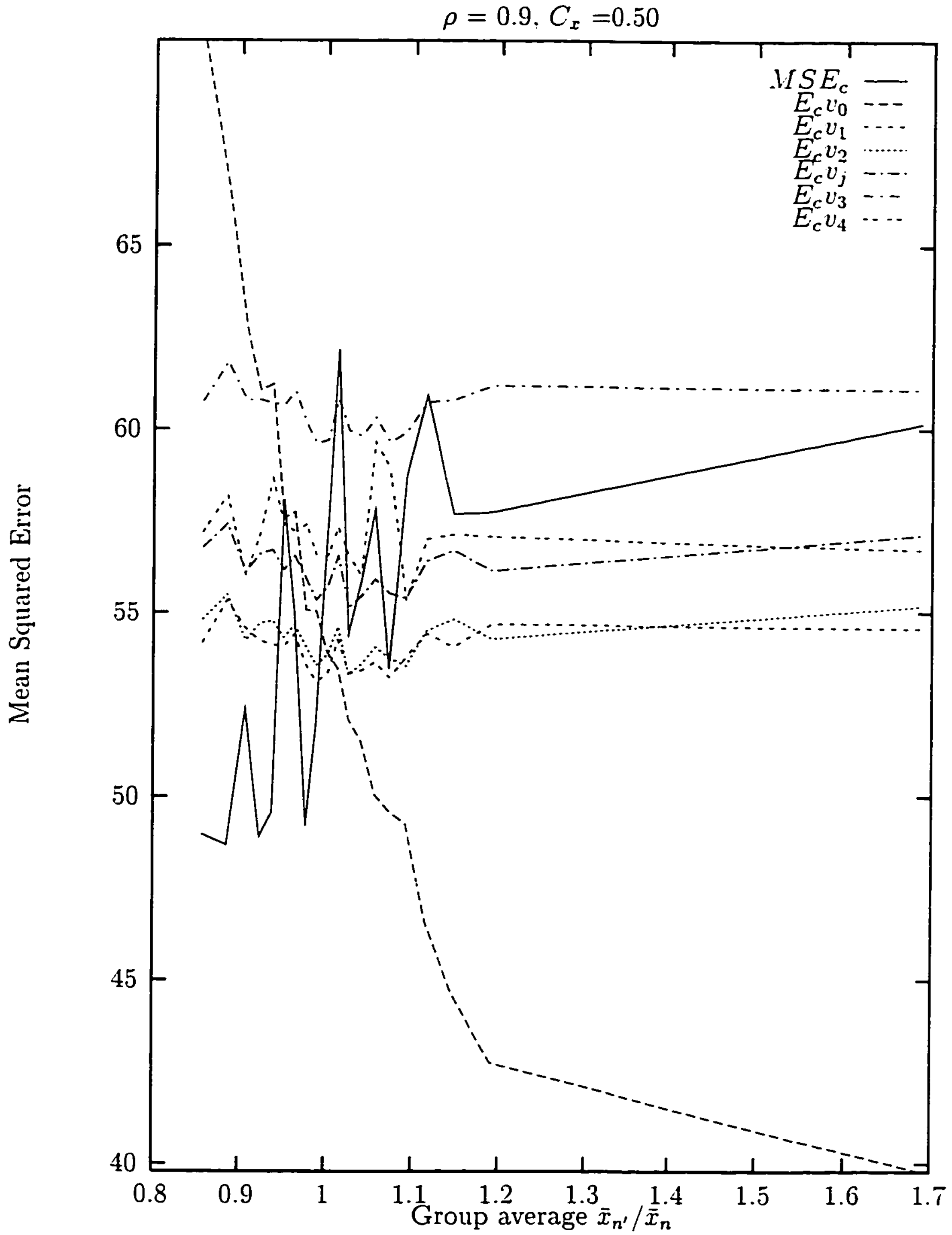


Figure 3.3: Conditional means $E_c v_0$, $E_c v_1$, $E_c v_2$, $E_c v_3$, $E_c v_4$, $E_c v_j$ and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$, versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and $n = 20$.

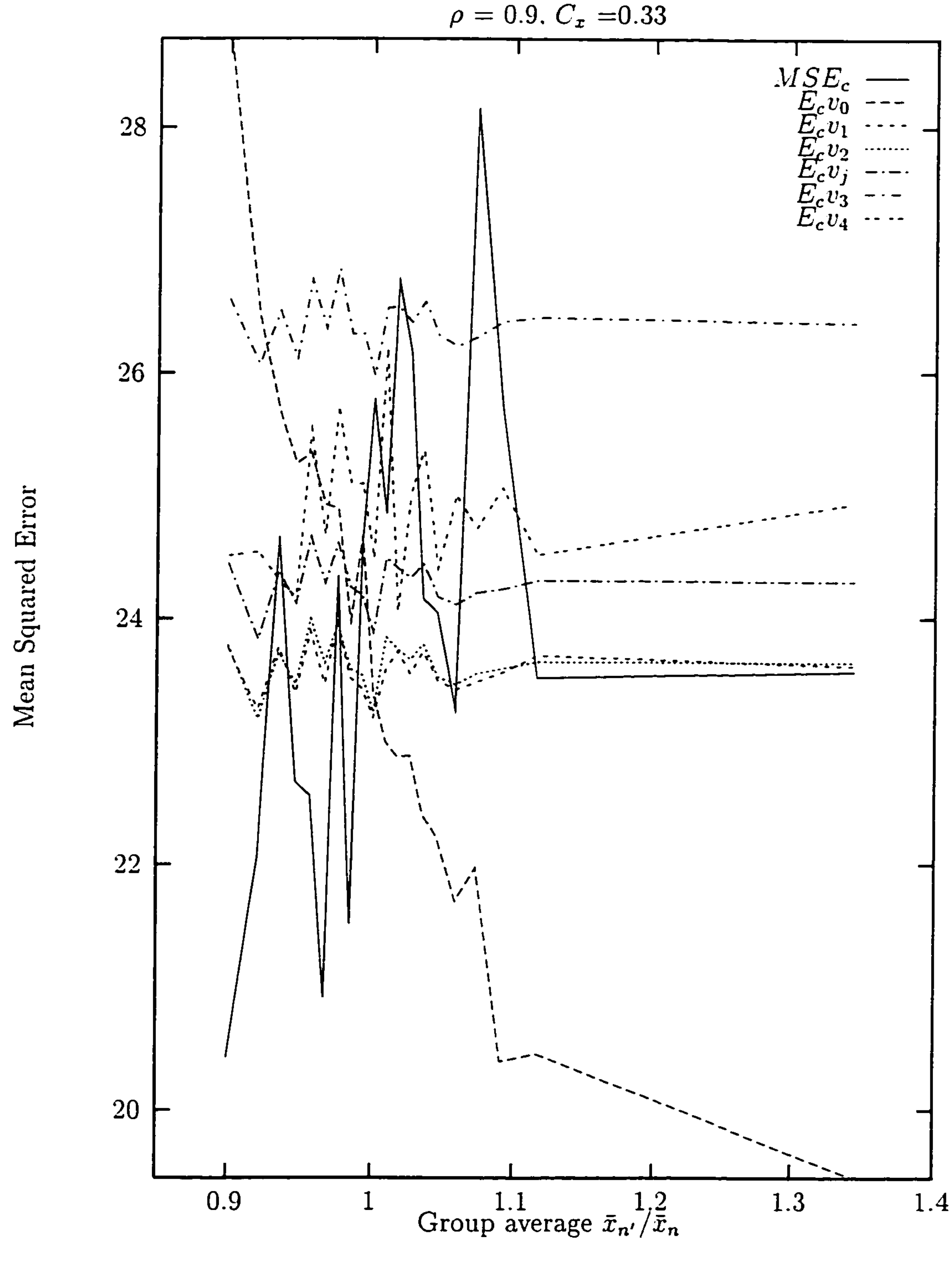


Figure 3.4: Conditional means E_cv_0, E_cv_1, E_cv_2, E_cv_3, E_cv_4, E_cv_j and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and n = 20.

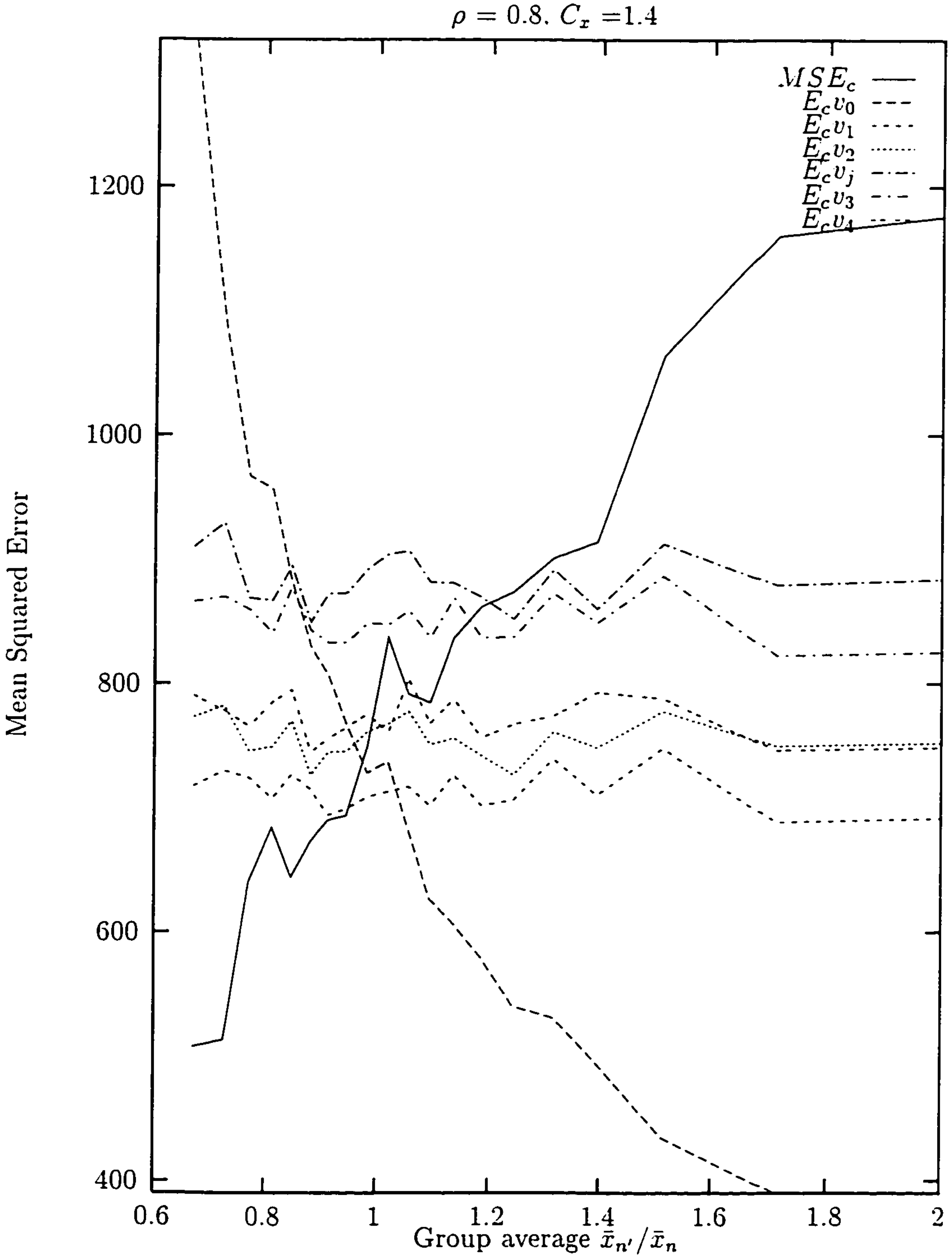


Figure 3.5: Conditional means $E_c v_0$, $E_c v_1$, $E_c v_2$, $E_c v_3$, $E_c v_4$, $E_c v_j$ and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and $n = 20$.

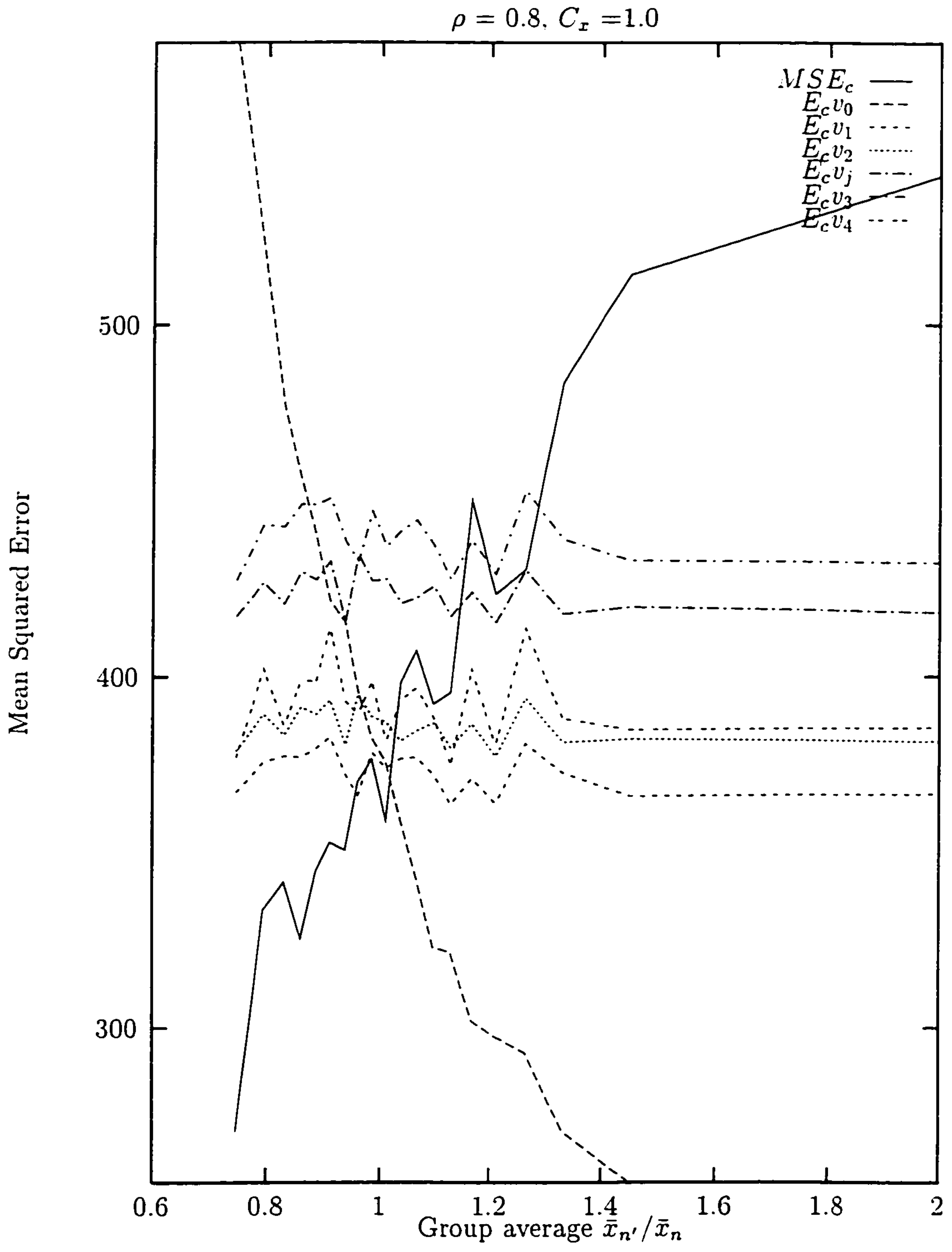


Figure 3.6: Conditional means $E_c v_0$, $E_c v_1$, $E_c v_2$, $E_c v_3$, $E_c v_4$, $E_c v_j$ and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and n = 20.

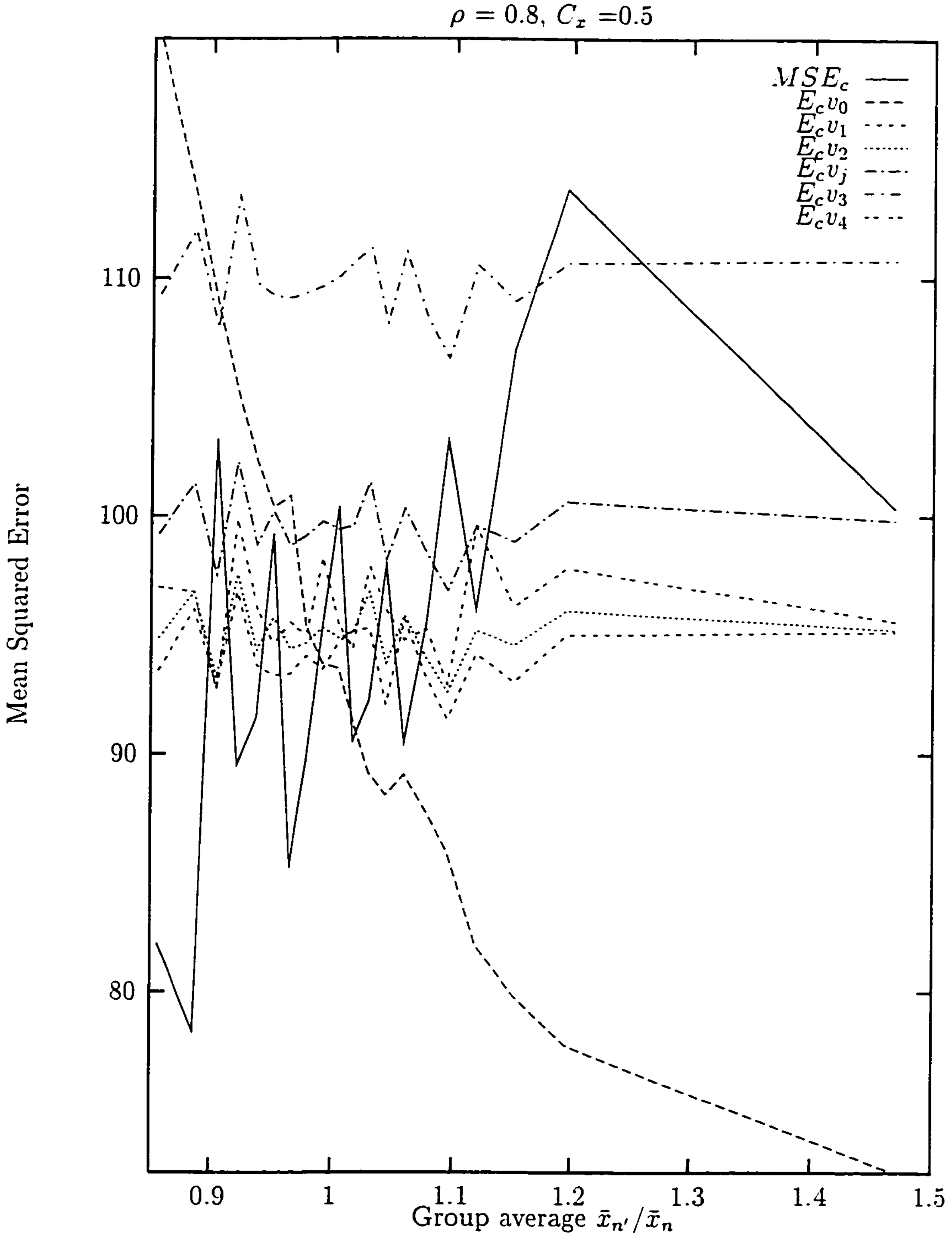


Figure 3.7: Conditional means $E_c v_0$, $E_c v_1$, $E_c v_2$, $E_c v_3$, $E_c v_4$, $E_c v_j$ and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and n = 20.

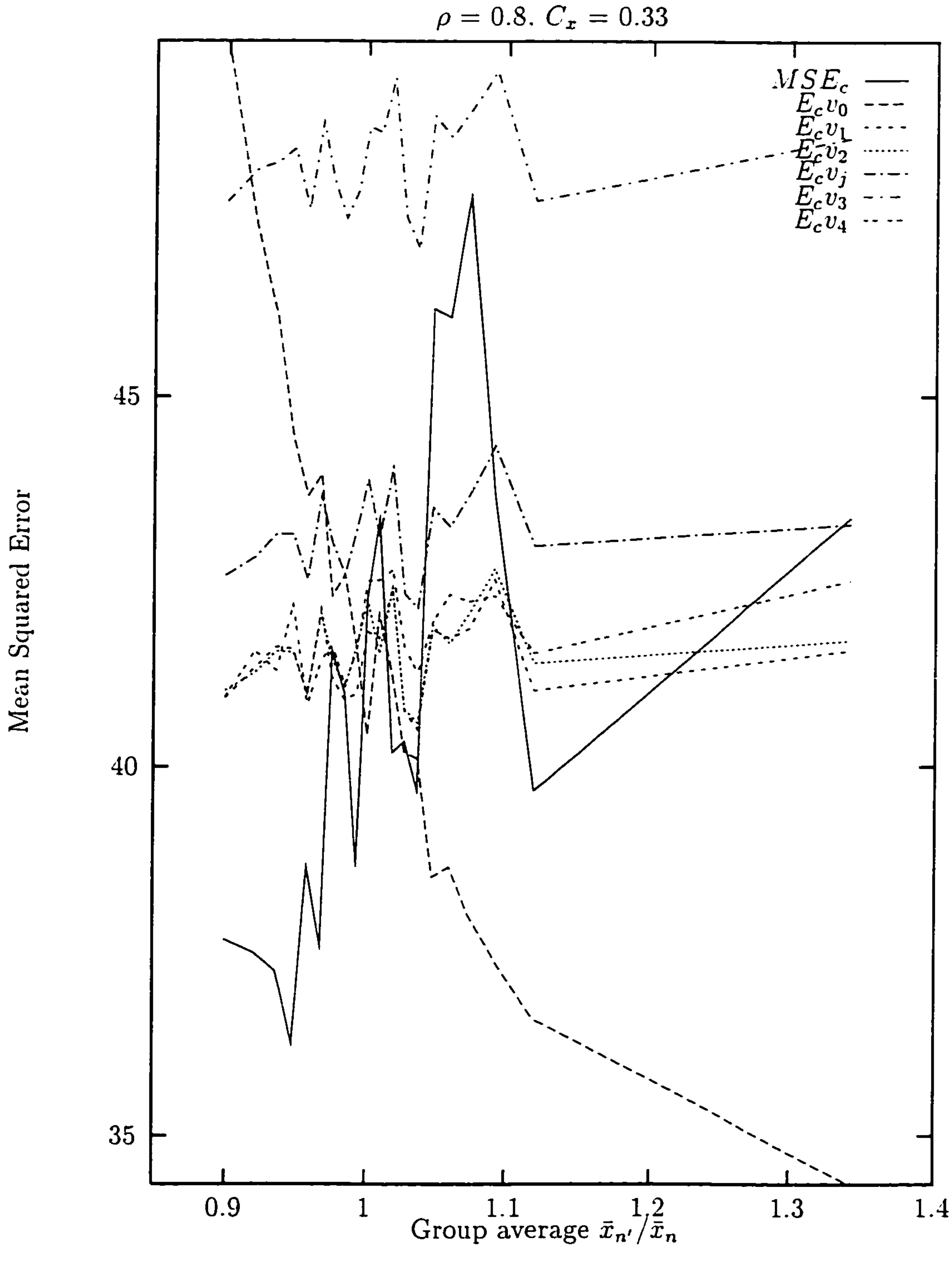


Figure 3.8: Conditional means $E_c v_0$, $E_c v_1$, $E_c v_2$, $E_c v_3$, $E_c v_4$, $E_c v_j$ and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and $n = 20$.

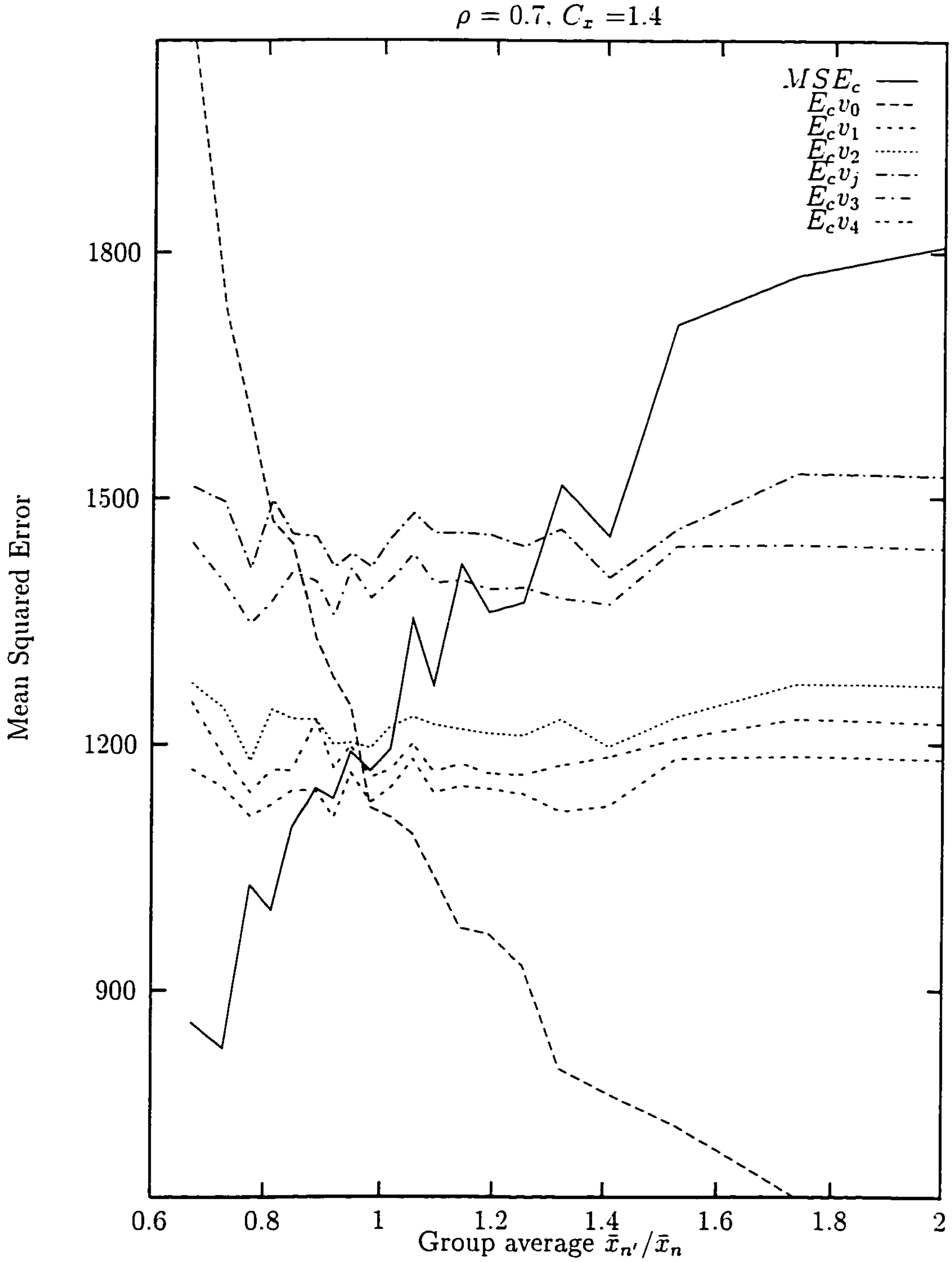


Figure 3.9: Conditional means $E_c v_0$, $E_c v_1$, $E_c v_2$, $E_c v_3$, $E_c v_4$, $E_c v_j$ and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and $n = 20$.

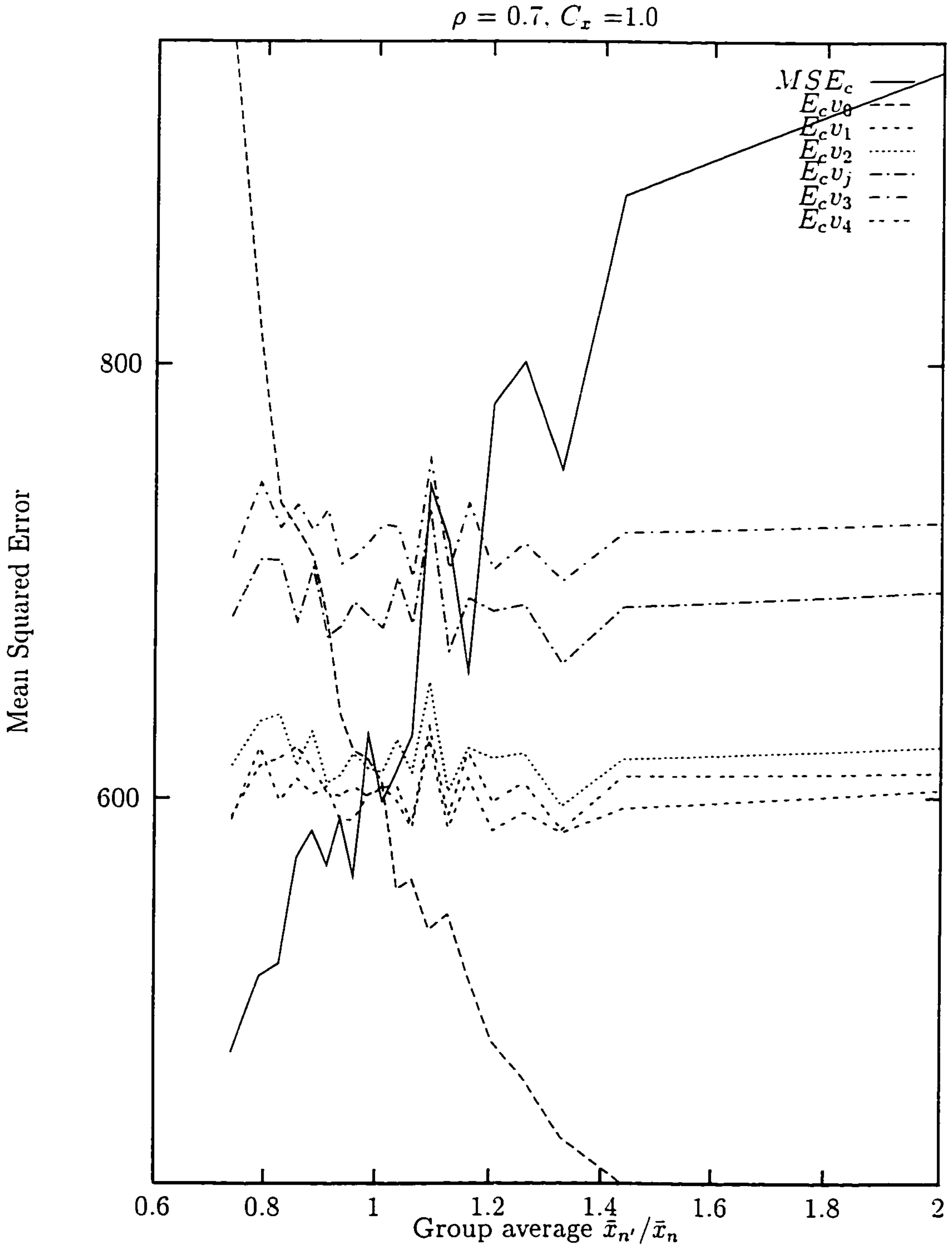


Figure 3.10: Conditional means E_cv_0, E_cv_1, E_cv_2, E_cv_3, E_cv_4, E_cv_j and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and n = 20.

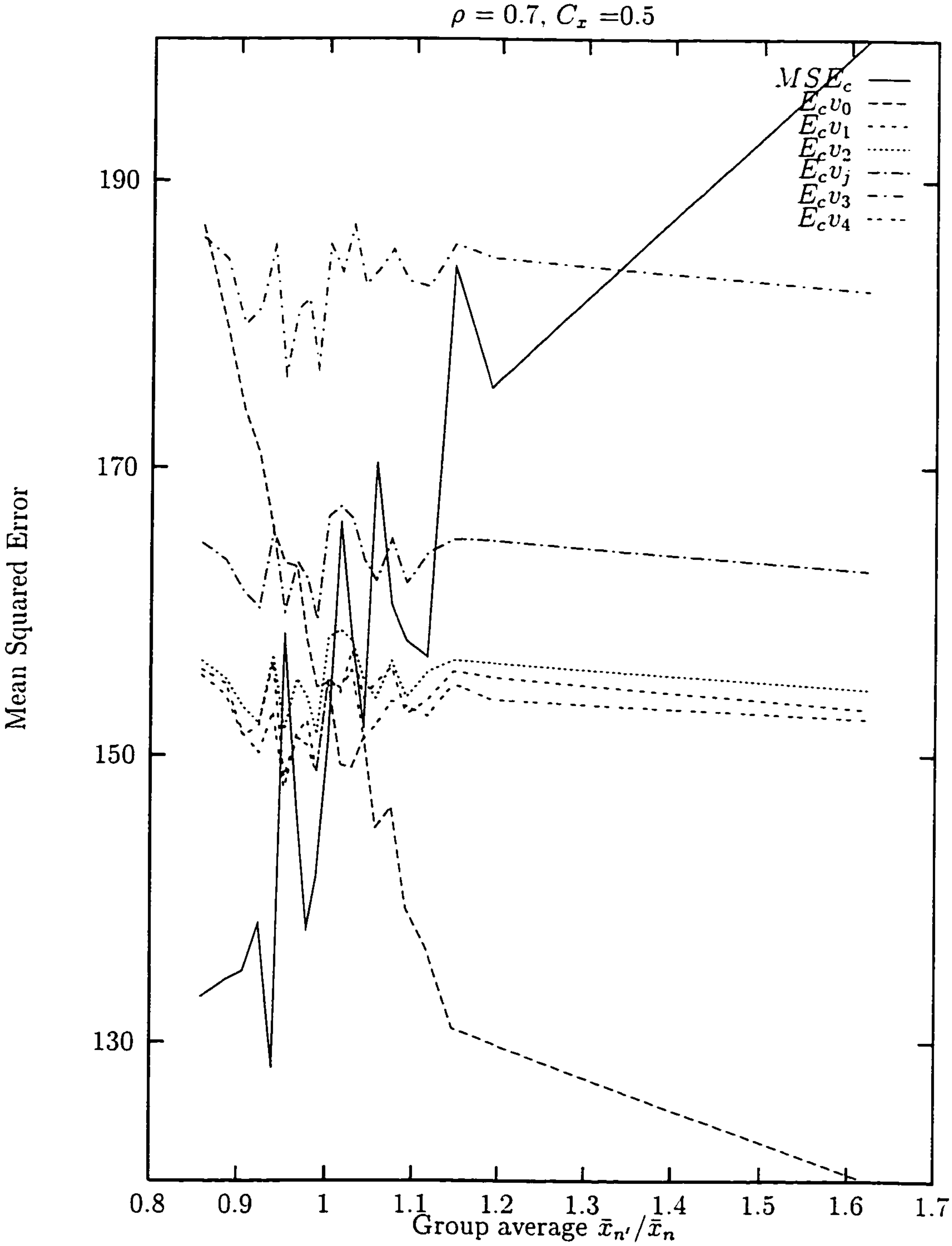


Figure 3.11: Conditional means E_cv_0, E_cv_1, E_cv_2, E_cv_3, E_cv_4, E_cv_j and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and n = 20.

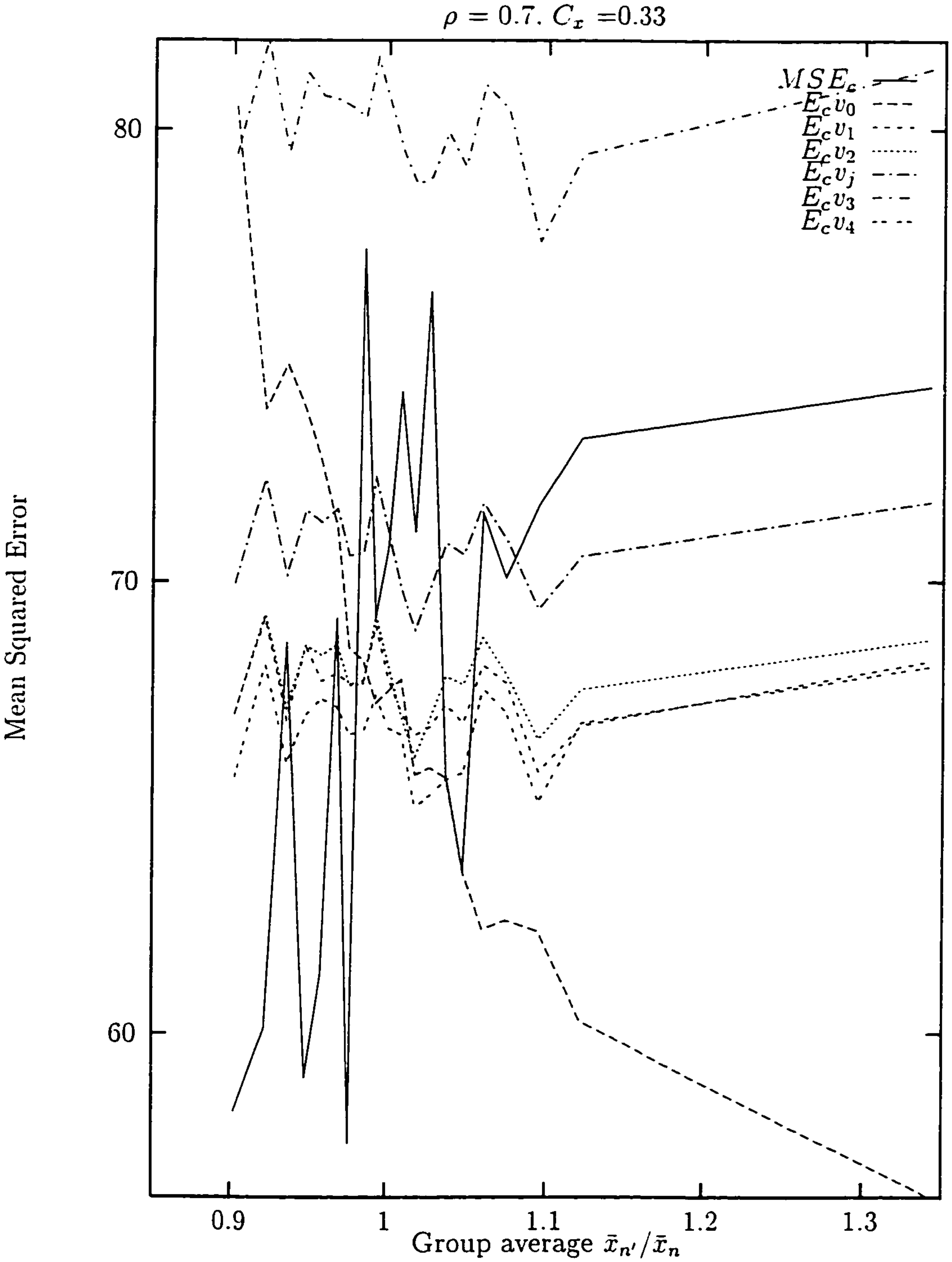


Figure 3.12: Conditional means E_cv_0, E_cv_1, E_cv_2, E_cv_3, E_cv_4, E_cv_j and conditional mean squared error (MSE_c) of $\bar{y}_{rt}$ versus group average $\bar{x}_{n'}/\bar{x}_n$ with n'=100 and n = 20.

Chapter 4

Bootstrap method for measurement error model

4.1 Introduction

It has been generally recognized that true measurements of a characteristic are often difficult to observe but they are observed with measurement errors. In view of this, considerable effort has been expended on the development of methods of analyzing data which are contaminated with measurement errors. In this chapter, we consider a model that is related to measurement errors. This model can be viewed as a generalization of the simple regression model, which takes into account random measurement errors on both the dependent and independent variables.

More precisely, we assume that two random variables U_1 and U_2 are observed subject to measurement error, both are related by $U_2 = \alpha + \beta U_1$, where α and β are unknown parameters. Further, we assume that actual observed values are $X = U_1 + \delta$ and $Y = U_2 + \varepsilon$. The pairs $(\delta_i, \varepsilon_i)$ are independently distributed for different i's, and δ_i and ε_i may or may not be independent of each other although they are independent of (U_{1i}, U_{2i}). When U_1 and U_2 are assumed to be unknown constants, the model is known as a functional model, whereas if the U_{1i}'s are independent random variables with the same distribution the model is known as a structural model. Fuller (1987) provided more details on these models.

It is well-known (see Kendall and Stuart, 1979 and Riersøl, 1950) that if δ and ε are normally distributed, then α and β are unidentifiable if and only if U_1 and U_2 are constants or U_1 and U_2 are normally distributed. Hence, assuming normal errors and without further information, α and β cannot be estimated in the functional model or in the structural model. However, if the error variance ratio $\lambda^2 = \sigma_\varepsilon^2/\sigma_\delta^2$ is known then both α and β can be estimated consistently when $\sigma_{U_1}^2$, σ_δ^2 and σ_ε^2 are finite. The above assumption is often satisfied if X and Y represent similar characteristics measured in the same units, in which case $\lambda^2 = 1$. In other instances, information from another independent sample, such as a preliminary study, often provides a suitable value for λ^2. In the functional case where the U_{1i}'s are true unknown values, Solari (1969) pointed out that the solution of maximum likelihood equations was a saddle point. Birch (1964) and Barnett (1967) obtained the maximum likelihood solution when both σ_δ^2 and σ_ε^2 were known. For a comprehensive coverage of this work and related topics on this subject, see for example Fuller (1987), Gleser (1981), and Chan and Mak (1983) on a multivariate model. Lindley and El-Sayyad (1968) and Zellner (1971) considered a Bayesian approach to these models.

Much of the interest has been focused on the estimation and testing procedure of β. Little investigation has been devoted to bootstrap procedure for estimating the standard error and confidence interval for β. The sampling distribution of the regression estimator $\hat{\beta}$ is skewed (see Anderson and Sawa, 1982). As a result, the large sample normal approximation as well as the likelihood ratio chi-square approximation performs poorly for small samples. In contrast, the bootstrap sampling distribution incorporates the skewness of the true sampling distributions. This feature is referred to as the second-order correctness of the bootstrap. Babu and Bai (1992) obtained a two-term Edgeworth expansion for $\hat{\beta}$ for a linear functional error-in-variables model. They showed that by using these expansions the bootstrap approximation of the sampling distribution was superior to the classical normal approximation. Linder and Babu (1994) proposed a bootstrap procedure based on the residuals for functional measurement error model with known error variance ratio and symmetric errors. However, their method is cumbersome, in part, because it involves calculations of correction factors so that the first two moments of the bootstrap estimator $\hat{\beta}^*$ match with the usual estimates of the first two moments of $\hat{\beta}$. Moreover, implementation of this approach requires a different correction factor for each parameter. Kelly (1984) considered the structural model with known error variance ratio from the influence function of β and obtained an estimate of the variance of $\hat{\beta}$. She noted that this method performed poorly, as the influence function estimate of the variance was biased.

In this chapter, we investigate the classical and a weighted bootstrap methods for the structural relationship model with known variance ratio for an arbitrary error distribution. Wu (1986) first proposed the weighted bootstrap in the context of the classical regression problem. In this procedure i.i.d. $\{t_i, i = 1, 2, ..., n\}$

observations are drawn from an external population having mean 0 and variance 1, independent of the original data. For the second order accuracy of the bootstrap estimator based on this method, Liu (1988) suggested another restriction on the external population, namely that third central moment of t_i must also be equal 1. This chapter is divided into three main sections. In Section 4.2, the method of moments is used to estimate the parameters and some preliminary results on asymptotic properties of the estimators are also given. Section 4.3 reviews the Linder and Babu method and describes the proposed bootstrap methods along with their asymptotic properties. In Section 4.4, the results of a simulation study are given.

4.2 The Structural Model

Consider the structural equations model for n random vectors $\mathbf{Z}_i = (X_i, Y_i)^T$. It is assumed that for each $i = 1 \ldots n$, we have

$$\mathbf{Z}_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} U_{1i} \\ U_{2i} \end{pmatrix} + \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} = \mathbf{U}_i + \boldsymbol{\xi}_i, \tag{4.2.1}$$

$$U_{2i} = \alpha + \beta U_{1i}, \tag{4.2.2}$$

where the $\mathbf{U}_i$'s are independently distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Gamma}_{\mathbf{U}}$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}_{\mathbf{U}} = \begin{pmatrix} \sigma^2_{U_1} & \beta\sigma^2_{U_1} \\ & \beta^2\sigma^2_{U_1} \end{pmatrix}. \tag{4.2.3}$$

The random variables $\boldsymbol{\xi}_i$'s are i.i.d. with mean vector $\mathbf{0}$ and covariance matrix

$$\boldsymbol{\Gamma}_{\boldsymbol{\xi}} = \begin{pmatrix} \sigma^2_\delta & 0 \\ & \sigma^2_\varepsilon \end{pmatrix}. \tag{4.2.4}$$

such that

$$\lambda^2 = \sigma_\varepsilon^2/\sigma_\delta^2 \text{ is known.} \tag{4.2.5}$$

Further, we assume that for each i,

$$\mathbf{U}_i \text{ and } \boldsymbol{\xi}_i \text{ are independent.} \tag{4.2.6}$$

Let F denote the common distribution of the $\mathbf{Z}_i$'s. By (4.2.1)–(4.2.6), the mean vector $\boldsymbol{\mu}_F$ and covariance matrix $\boldsymbol{\Gamma}_F$ are, respectively, given by

$$\boldsymbol{\mu}(F) = \begin{pmatrix} \mu_X(F) \\ \mu_Y(F) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \alpha + \beta\mu_1 \end{pmatrix} \tag{4.2.7}$$

and

$$\boldsymbol{\Gamma}(F) = \begin{pmatrix} \sigma_{XX}(F) & \sigma_{XY}(F) \\ & \sigma_{YY}(F) \end{pmatrix} = \begin{pmatrix} \sigma_{U_1}^2 + \sigma_\delta^2 & \beta\sigma_{U_1}^2 \\ & \beta^2\sigma_{U_1}^2 + \sigma_\varepsilon^2 \end{pmatrix}. \tag{4.2.8}$$

If we substitute for U_1 and U_2 from (4.2.1) into (4.2.2), we obtain

$$Y_i = \alpha + \beta X_i + \varepsilon_i - \beta\delta_i, \quad i = 1, \ldots, n. \tag{4.2.9}$$

This is not a classical regression model since here, X is a random variable which is correlated with the error term $(\varepsilon - \beta\delta)$. From (4.2.3)–(4.2.8), we have

$$Cov_F(X, \varepsilon - \beta\delta) = -\beta\sigma_\delta^2, \tag{4.2.10}$$

which is 0 only if $\sigma_\delta^2 = 0$, the case corresponding to the simple regression situation, or in the trivial case $\beta = 0$. Thus, the existence of errors in both U_1 and U_2 poses a problem quite distinct from that of conventional regression model.

4.2.1 The method of moments estimators

The parameter vector $\boldsymbol{\theta} = (\alpha, \beta)^T$ can be written as a functional of the unknown distribution function F (see Kelly, 1984) by letting

$$\alpha = \alpha(F) = \mu_Y(F) - \beta(F)\mu_X(F), \tag{4.2.11}$$

$$\begin{aligned} \beta &= \beta(F) \\ &= \frac{1}{2\sigma_{XY}(F)}[\sigma_{YY}(F) - \lambda^2\sigma_{XX}(F) + \{[\sigma_{XY}(F) - \lambda^2\sigma_{XX}(F)]^2 + 4\lambda^2\sigma_{XY}^2(F)\}^{1/2}], \end{aligned} \tag{4.2.12}$$

where by definition,

$$\begin{aligned} \mu_Y(F) &= \int\int ydF(x,y), \\ \sigma_{XX}(F) &= \int\int x^2dF(x,y) - \left[\int\int xdF(x,y)\right], \end{aligned}$$

and the other quantities are defined in a similar fashion. Note that $\beta(F)$ may be rewritten as

$$\beta(F) = h(F) + [h^2(F) + \lambda^2]^{1/2}, \tag{4.2.13}$$

with

$$h = h(F) = \frac{1}{2\sigma_{XY}(F)}\{\sigma_{YY}(F) - \lambda^2\sigma_{XX}(F)\}. \tag{4.2.14}$$

Here and in what follows, for any sequences $\{H_i\}$ and $\{R_i\}$, we use the notation

$$\bar{H} = n^{-1}\sum_{i=1}^{n} H_i, \quad S_{HR} = n^{-1}\sum_{i=1}^{n}(H_i - \bar{H})(R_i - \bar{R}). \tag{4.2.15}$$

Let F_n denote the sample distribution function corresponding to F. Denote the sample mean and covariance matrix, respectively, by

$$\boldsymbol{\mu}(F_n) = \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix} \text{and} \quad \boldsymbol{\Gamma}(F_n) = \begin{pmatrix} S_{XX} & S_{XY} \\ & S_{YY} \end{pmatrix}. \tag{4.2.16}$$

Under the model defined by (4.2.1)–(4.2.6), the method of moments estimator of $\boldsymbol{\theta}(F) = (\alpha(F), \beta(F))^T$ is given by

$$\boldsymbol{\theta}(F_n) = (\alpha(F_n), \beta(F_n))^T, \tag{4.2.17}$$

where

$$\hat{\alpha} = \alpha(F_n) = \bar{Y} - \beta(F_n)\bar{X}, \tag{4.2.18}$$

$$\hat{\beta} = \beta(F_n) = h(F_n) + (h^2(F_n) + \lambda^2)^{1/2}, \tag{4.2.19}$$

with

$$\hat{h} = h(F_n) = \frac{1}{2S_{XY}}\{S_{YY} - \lambda^2 S_{XX}\}. \tag{4.2.20}$$

By the law of large numbers, $\hat{\alpha}$ and $\hat{\beta}$ are consistent estimators for α and β, respectively, for all distribution functions F with finite second moments. When F is bivariate normal, $\boldsymbol{\theta}(F_n)$ is the maximum likelihood estimator of $\boldsymbol{\theta}(F)$ (see Kendall and Stuart, 1979).

Theorem 4.1. *Let the model defined by (4.2.1)-(4.2.6) hold with known error variance ratio $\lambda^2 > 0$ and suppose that X and Y have finite sixth moment. Then*

(i)

$$E_F(\hat{\beta} - \beta) = -\frac{\beta}{2n\mu_{11}^2\sqrt{h^2 + \lambda^2}}\left\{\mu_{13} - \lambda^2\mu_{31} - 2h\mu_{22}\right\} + O(n^{-2}). \tag{4.2.21}$$

(ii)

$$E_F(\hat{\beta}-\beta)^2 = \frac{\beta^2}{4n\mu_{11}^2(h^2+\lambda^2)}\left\{\mu_{04}+\lambda^4\mu_{40}+2\mu_{22}(2h^2-\lambda^2)\right.$$
$$\left.-4h(\mu_{13}-\lambda^2\mu_{31})\right\}+O(n^{-2}). \qquad (4.2.22)$$

(iii) If the joint distribution of X and Y are symmetric, i.e., $\mu_{30}=\mu_{03}=\mu_{12}=\mu_{21}=0$, *then,*

$$E_F(\hat{\beta}-\beta)^3 = \frac{1}{n^2}\left(\frac{\beta}{2\mu_{11}\sqrt{h^2+\lambda^2}}\right)^3\left\{\mu_{06}-\lambda^6\mu_{60}-3\lambda^2(\mu_{24}-\lambda^2\mu_{42})\right.$$
$$\left.-6h[\mu_{15}+\lambda^4\mu_{51}-2\lambda^2\mu_{33}]+12h^2(\mu_{24}-\lambda^2\mu_{42})-8h^3\mu_{33}\right\}$$
$$+O(n^{-3}) \qquad (4.2.23)$$

where $\mu_{uv}=E(X-\mu_X)^u(Y-\mu_Y)^v$.

Proof. See Appendix A. □

The expression in (4.2.22) can also be derived from the influence function for $\hat{\beta}$; the details are given in Kelly (1984). To get an idea of the magnitude of the bias of $\hat{\beta}$, we consider that the population follows a standard bivariate normal distribution, so that

$$\phi(x,y)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right]; \qquad (4.2.24)$$

and

$$\mu_{21}=\mu_{12}=\mu_{30}=\mu_{03}=0, \quad \mu_{20}=\mu_{02}=1, \quad \mu_{31}=\mu_{13}=3\rho,$$
$$\mu_{22}=1+2\rho^2, \quad \mu_{40}=\mu_{04}=3, \quad \mu_{24}=\mu_{42}=3+12\rho^2 \text{ and } \mu_{33}=9\rho+6\rho^3.$$

This yields

$$E_F(\hat{\beta}-\beta)=\frac{\beta h}{\sqrt{h^2+\lambda^2}}\left(\frac{1}{n}-\frac{4}{n^2}\right)\left(\frac{1-\rho^2}{\rho^2}\right)+O(n^{-3}). \tag{4.2.25}$$

Denoting the first and second order approximation of the relative bias of $\hat{\beta}$ by $B_1(\hat{\beta})$ and $B_2(\hat{\beta})$, respectively, we have

$$B_1(\hat{\beta})=\frac{E_F(\hat{\beta})-\beta}{\beta}\doteq\frac{h}{n\sqrt{h^2+\lambda^2}}\left(\frac{1-\rho^2}{\rho^2}\right). \tag{4.2.26}$$

To a second approximation, the relative bias of $\hat{\beta}$ can, therefore, be expressed as

$$B_2(\hat{\beta})\doteq B_1(\hat{\beta})\left(1-\frac{4}{n}\right). \tag{4.2.27}$$

Equation (4.2.27) shows that the contribution of the second and third order terms to the relative bias of $\hat{\beta}$ is $4/n$ times the value of the latter to a first approximation. Unless n is small, the contribution can be considered negligible.

Comparing (4.2.21) and (4.2.22), we see that both the bias and the variance of $\hat{\beta}$ are of order n^{-1}. Hence, for n sufficiently large, the bias is negligible as compared to the standard error which is of the order $n^{-1/2}$.

The exact sampling behavior of the estimator $\hat{\beta}$ defined in (4.2.19) cannot be obtained easily. Therefore, it seems necessary to use large sample theory to develop an approximation of the distribution of $\hat{\beta}$. We now give the asymptotic normal distribution of the estimators for the slope $\hat{\beta}$ and intercept $\hat{\alpha}$, under the general structural linear relationship (4.2.1)–(4.2.6).

Theorem 4.2. *Let the model defined by (4.2.1)–(4.2.6) hold with known error variance ratio $\lambda^2>0$ and X and Y have finite sixth moment. Then, as $n\to\infty$,* *1.* $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$, *for* $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta})^T$, *where*

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_\delta^2(\beta^2 + \lambda^2) + \mu_1^2\Sigma_{22} & -\mu_1\Sigma_{22} \\ -\mu_1\Sigma_{22} & \Sigma_{22} \end{bmatrix},$$

and

$$\Sigma_{22} = \frac{\beta^2}{4\mu_{11}^2(h^2 + \lambda^2)}\left\{\mu_{04} + \lambda^4\mu_{40} + 2\mu_{22}(2h^2 - \lambda^2) - 4h(\mu_{13} - \lambda^2\mu_{31})\right\}.$$

2. Furthermore, $\hat{\boldsymbol{\Sigma}}$ *converges in probability to* $\boldsymbol{\Sigma}$, *where*

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \hat{\sigma}_\delta^2(\hat{\beta}^2 + \lambda^2) + \hat{\mu}_1^2\hat{\Sigma}_{22} & -\hat{\mu}_1\hat{\Sigma}_{22} \\ -\hat{\mu}_1\hat{\Sigma}_{22} & \hat{\Sigma}_{22} \end{bmatrix},$$

$$\hat{\sigma}_\delta^2 = \sum_{i=1}^{n} e_i^2/(n(\hat{\beta}^2 + \lambda^2)), \quad e_i = Y_i - \hat{\alpha} - \hat{\beta}X_i, \quad \hat{\mu}_1 = n^{-1}\sum_{i=1}^{n}\hat{U}_{1i},$$

$$\hat{\Sigma}_{22} = \frac{\hat{\beta}^2}{4\hat{\mu}_{11}^2(\hat{h}^2 + \lambda^2)}\left\{\hat{\mu}_{04} + \lambda^4\hat{\mu}_{40} + 2\hat{\mu}_{22}(2\hat{h}^2 - \lambda^2) - 4\hat{h}(\hat{\mu}_{13} - \lambda^2\hat{\mu}_{31})\right\},$$

with $\hat{U}_{1i}$ *is defined in (4.3.1) and* $\hat{\mu}_{rs}$ *is a plug-in sample moment estimate, and is given by*

$\hat{\mu}_{rs} = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^r(Y_i - \bar{Y})^s$.

Proof. We note that S_{YY}, S_{XX} and S_{XY} are asymptotically unbiased and from Kendall and Stuart (1979), we have the following results

$$\begin{aligned} Var_F[S_{YY}] &= n^{-1}(\mu_{04} - \mu_{02}^2), & Var_F[S_{XX}] &= n^{-1}(\mu_{40} - \mu_{20}^2), \\ Var_F[S_{XY}] &= n^{-1}(\mu_{22} - \mu_{11}^2), & Cov_F[S_{YY}, S_{XY}] &= n^{-1}(\mu_{13} - \mu_{02}\mu_{11}), \\ Cov_F[S_{XX}, S_{XY}] &= n^{-1}(\mu_{31} - \mu_{20}\mu_{11}), & Cov_F[S_{YY}, S_{XX}] &= n^{-1}(\mu_{22} - \mu_{02}\mu_{20}). \end{aligned}$$

Because the sample moments are converging in probability to their respective population moments, we can expand $\hat{\beta}$ using a Taylor series expansion about β to obtain

$$\hat{\beta} = \beta + \beta(h^2 + \lambda^2)^{-1/2}(\hat{h} - h) + O_p(n^{-1}), \tag{4.2.28}$$

or equivalently,

$$n^{1/2}(\hat{\beta} - \beta) = n^{1/2}\beta(\hat{h} - h)(h^2 + \lambda^2)^{-1/2} + O_p(n^{-1/2}), \tag{4.2.29}$$

which implies that the limiting distribution of $n^{1/2}(\hat{\beta} - \beta)$ is the same as that of $n^{1/2}\beta(\hat{h} - h)(h^2 + \lambda^2)^{-1/2}$. The asymptotic variance of $\hat{h}$ is given by

$$\begin{aligned} Var_F(\hat{h}) = h^2 \Bigg\{ & \frac{1}{E_F^2[S_{YY} - \lambda^2 S_{XX}]} \big(Var_F[S_{YY}] - 2\lambda^2 Cov_F[S_{YY}, S_{XX}] \\ & + \lambda^4 Var_F[S_{XX}]\big) + \frac{Var[S_{XY}]}{E_F^2[S_{XY}]} - \frac{2}{E_F[S_{XY}]E_F[S_{YY} - \lambda^2 S_{XX}]} \\ & \big(Cov_F[S_{YY}, S_{XY}] - \lambda^2 Cov_F[S_{XX}, S_{XY}]\big)\Bigg\} \\ = & \frac{1}{4n\mu_{11}^4}\{\mu_{11}^2[\mu_{04} - \mu_{02}^2 - 2\lambda^2(\mu_{22} - \mu_{02}\mu_{20}) + \lambda^4(\mu_{40} - \mu_{20}^2)] \\ & + 4h^2\mu_{11}^2(\mu_{22} - \mu_{11}^2) - 2\mu_{11}(\mu_{02} - \lambda^2\mu_{20}) \\ & \times [\mu_{13} - \mu_{02}\mu_{11} - \lambda^2(\mu_{31} - \mu_{20}\mu_{11})]\} + O(n^{-2}) \\ = & \frac{1}{4n\mu_{11}^2}\left\{\mu_{04} + \lambda^4\mu_{40} + 2\mu_{22}(2h^2 - \lambda^2) - 4h(\mu_{13} - \lambda^2\mu_{31})\right\} + O(n^{-2}). \end{aligned}$$

Hence the asymptotic variance of $\hat{\beta}$ is given by

$$Var_F(\hat{\beta}) = \frac{\beta^2}{h^2 + \lambda^2} Var_F(\hat{h}) + O(n^{-2}).$$

Turning to $Var_F(\hat{\alpha})$, consider

$$\begin{aligned} \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} &= \alpha + \beta U_{1i} + \bar{\varepsilon} - \hat{\beta}(\bar{U}_{1i} + \bar{\delta}) \\ &= \alpha - (\hat{\beta} - \beta)\mu_1 + \bar{v} + O_p(n^{-1}), \end{aligned}$$

where $\bar{v} = \bar{\varepsilon} - \beta\bar{\delta}$ and hence the distribution of $n^{1/2}(\hat{\alpha} - \alpha)$ has the same distribution as that of $n^{1/2}[\bar{v} - (\hat{\beta} - \beta)\mu_1)]$. This yields

$$\begin{aligned} Var_F(\hat{\alpha}) &= Var_F(\bar{v} - (\hat{\beta} - \beta)\mu_1) \\ &= \sigma_\delta^2(\beta^2 + \lambda^2) + \mu_1^2 Var_F(\hat{\beta}), \end{aligned} \tag{4.2.30}$$

and

$$\begin{aligned} Cov_F(\hat{\alpha}, \hat{\beta}) &= Cov_F[\bar{v} - (\hat{\beta} - \beta)\mu_1), (\hat{\beta} - \beta)\mu_1)] \\ &= -\mu_1 Var_F(\hat{\beta}). \end{aligned} \tag{4.2.31}$$

To show the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ and the consistency of $\hat{\boldsymbol{\Sigma}}$, we note that $\hat{\boldsymbol{\theta}}$ is a continuous differentiable function of the U-statistics $(\bar{X}, \bar{Y}, S_{XX}, S_{YY}, S_{XY})$. Thus by Theorems 8 and 9 of Arvesen (1969), the desired results follow. □

Because $\hat{\Sigma}$ is a consistent estimator of Σ, it follows that

$$t = n^{1/2}\hat{\Sigma}_{22}^{-1/2}(\hat{\beta} - \beta) \tag{4.2.32}$$

is approximately distributed as a $N(0, 1)$ random variable. In practice it seems reasonable to approximate the distribution of (4.2.32) with the distribution of Student's t with $n - 2$ degree of freedom. Instead of using $n^{-1}\hat{\Sigma}_{22}$ to estimate $Var_F(\hat{\beta})$, one could use a jackknife procedure. The next section describes a jackknife procedure.

4.2.2 Jackknife Variance Estimation

In this section, we give the jackknife variance estimator for $\hat{\boldsymbol{\theta}}$ (see Kelly, 1984). Let $\hat{\boldsymbol{\theta}}_{-i}$ be an estimator of $\boldsymbol{\theta}$ with the i-th observation (X_i, Y_i) omitted and define the pseudo values

$$\hat{\boldsymbol{\theta}}_i = n\hat{\boldsymbol{\theta}} - (n-1)\hat{\boldsymbol{\theta}}_{-i}. \qquad (4.2.33)$$

The jackknife estimator of $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{\theta}}_{(\cdot)} \quad = n^{-1}\sum \hat{\boldsymbol{\theta}}_i. \qquad (4.2.34)$$

The jackknife estimator of the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$\hat{\Sigma}_J = \frac{n-1}{n}\sum[\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{(\cdot)}][\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{(\cdot)}]^T. \qquad (4.2.35)$$

Since $\hat{\alpha}$ and $\hat{\beta}$ are continuously differentiable functions of the U-statistics $(\bar{X}, \bar{Y}, S_{XX}, S_{YY}, S_{XY})$ and when $E_F(X^4) < \infty$ and $E_F(Y^4) < \infty$, by Theorem 9 of Arvessen (1969), we have

$$n\hat{\boldsymbol{\Sigma}}_J - \boldsymbol{\Sigma} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty. \qquad (4.2.36)$$

4.3 The Linder and Babu Method

Linder and Babu (1994) proposed a bootstrap method where resampling was done by taking a sample with replacement from the residuals and then repeating this a number of times to match the usual variance estimates. However, in such resampling method, one needs to modify the residuals and the usual bootstrap variance estimator.

Let $\hat{U}_{1i}$ and $\hat{U}_{2i}$ denote the fitted values U_{1i} and U_{2i}, respectively. We require, for every i, that $\lambda^2 = (Y_i - \hat{U}_{2i})^2/(X_i - \hat{U}_{1i})^2$, which in turn requires the redefinition of the fitted values,

$$\begin{aligned} \hat{U}_{1i} &= X_i + r_i/(\lambda + |\hat{\beta}|), \\ \hat{U}_{2i} &= \hat{\alpha} + \hat{\beta}\hat{U}_{1i} = Y_i - \lambda e_i/(\lambda + |\hat{\beta}|), \end{aligned} \tag{4.3.1}$$

where $e_i = Y_i - \hat{\alpha} - \hat{\beta}X_i$.

The residuals $(X_i - \hat{U}_{1i}, Y_i - \hat{U}_{2i})$ underestimate the true error, i.e. the mean squared of the residuals are asymptotically negatively biased for variances $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\varepsilon^2$, respectively. This results from the fact that

$$s^2 = \hat{\sigma}_\delta^2 = \sum_1^n e_i^2/(n(\lambda^2 + \hat{\beta}^2)) \tag{4.3.2}$$

is a consistent estimator for σ_δ^2, see Kendall and Stuart (1979). Hence, to estimate the error variances consistently, the residuals are adjusted by multiplying with the correction factor $d_n = (\lambda + |\hat{\beta}|)/(\lambda^2 + \hat{\beta}^2)^{1/2}$, resulting in a set of "pseudo" residuals

$$\begin{aligned} r_i &= -e_i/(\lambda^2 + \hat{\beta}^2)^{1/2}, \\ s_i &= \lambda e_i/(\lambda^2 + \hat{\beta}^2)^{1/2} = -\lambda r_i. \end{aligned} \tag{4.3.3}$$

We now describe the Linder and Babu bootstrap algorithm

1. Given data set $(X_i, Y_i), i = 1, \ldots, n$ and the estimator $\hat{\beta}$ as defined in (4.2.19) for slope, compute the fitted values $(\hat{U}_{1i}, \hat{U}_{2i}), i = 1, \ldots, n$ using (4.3.1).

2. Resample $\hat{\delta}_i^*$ with replacement from the set $c_n\{r_1, \ldots, r_n\}$, where r_i are given (4.3.3) and $c_n = \hat{\beta}S_{\hat{U}_1\hat{U}_1}/S_{XY}$.

3. Independently of (2), resample $\hat{\varepsilon}_i^*$ with replacement from the set $c_n\{s_1, \ldots, s_n\}$, where s_i are given by (4.3.3) and c_n is the same as in step (2).

4. Obtain bootstrap estimate $\hat{\beta}^*$ which is the analogue of $\hat{\beta}$ as in (4.2.19) from the replicate data (X_i^*, Y_i^*).

5. Repeat steps (1)–(5) above B times, where B is large, typically between 100 and 1,000.

6. Calculate estimates of $Var_F(\hat{\beta})$ by

$$var_*(\hat{\beta}^*) = (B-1)^{-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \hat{\beta})^2.$$

7. In the b-th run, calculate the bootstrap-t, defined as

$$t_b^* = (\hat{\beta}_b^* - \hat{\beta})/\sqrt{\hat{\Phi}^*(\hat{\beta}^*)},$$

where $\hat{\beta}_b^*$ is the analogue of $\hat{\beta}$ computed in b-th bootstrap sample and $\hat{\Phi}^*(\hat{\beta}^*)$ is the bootstrap estimate of the variance of $\hat{\beta}^*$ given by

$$\hat{\Phi}^*(\hat{\beta}^*) = var_*(\hat{\beta}^*) + n^{-1}\hat{\Psi}$$

where $\hat{\Psi} = 4\lambda^6\hat{\beta}^4 s^4 kur(\hat{\beta}^2 + \lambda^2)^{-4}\hat{v}^{-4}$ and

$$kur = \frac{n^{-1}SD^4 - 6\lambda^2\hat{\beta}^2 s^4}{(\hat{\beta}^4 + \lambda^4)s^4} - 3$$

with $SD^4 = \sum_{i=1}^{n} e_i^4$, $\hat{v} = \hat{\mu}_{11}/\hat{\beta}$ and s^2 given by (4.3.2).

Remark: Linder and Babu (1994) noted that under this bootstrap procedure the usual variance estimator given in Step 6 above was not a proper estimator of $Var_F(\hat{\beta})$. Hence they suggested to use $\hat{\Phi}^*(\hat{\beta}^*) = var_*(\hat{\beta}^*) + n^{-1}\hat{\Psi}$. The computation of $\hat{\Psi}$ involves the fourth moment of residuals. To avoid this "after" correction, we propose two new bootstrap procedures where the usual bootstrap variance estimator is a valid estimator of $Var_F(\hat{\beta})$.

4.3.1 The Proposed Bootstrap Procedure

The bootstrap procedure proposed here for the structural model differs from the classical method in that it does not resample the data (X_i, Y_i) directly. Instead it starts with an estimating function for the parameter and independently resamples residuals in that function. We assume the conditions in Theorem 4.1 hold and that $\hat{h}$ has a first order continuous derivative around $h = h(\mu_{02}, \mu_{20}, \mu_{11})$. By a multivariate Taylor series expansion of $\hat{h}$, we have

$$\begin{aligned}\hat{h} &= h + \frac{1}{2}\mu_{11}^{-1}\{(\hat{\mu}_{02} - \mu_{02}) - \lambda^2(\hat{\mu}_{20} - \mu_{20}) - 2h(\hat{\mu}_{11} - \mu_{11})\} + O_p(n^{-1}) \\ &= h + \frac{1}{2}\mu_{11}^{-1}\{\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2h\hat{\mu}_{11}\} + O_p(n^{-1}). \qquad (4.3.4)\end{aligned}$$

We re-express (4.3.4) as

$$\begin{aligned}\hat{h} &= h + \frac{1}{2n}\mu_{11}^{-1}\sum_{i=1}^{n}\{(Y_i - \bar{Y})^2 - \lambda^2(X_i - \bar{X})^2 - 2h(X_i - \bar{X})(Y_i - \bar{Y})\} + O_p(n^{-1}) \\ &= h + \frac{1}{2n}\mu_{11}^{-1}\sum_{i=1}^{n}\{(\varepsilon_i - \bar{\varepsilon})^2 - \lambda^2(\delta_i - \bar{\delta})^2 - 2h(\varepsilon_i - \bar{\varepsilon})(\delta_i - \bar{\delta}) \\ &\quad + 2\gamma U_{1i}(\varepsilon_i - \bar{\varepsilon}) - 2\gamma\beta U_{1i}(\delta_i - \bar{\delta})\} + O_p(n^{-1}) \\ &= h + \frac{1}{2n}\mu_{11}^{-1}\sum A_i + O_p(n^{-1}), \qquad (4.3.5)\end{aligned}$$

where $A_i = (Y_i - \bar{Y})^2 - \lambda^2(X_i - \bar{X})^2 - 2h(X_i - \bar{X})(Y_i - \bar{Y})$ and $\gamma = \beta - h$.

Define

$$\begin{aligned}\hat{A}_i &= (Y_i - \bar{Y})^2 - \lambda^2(X_i - \bar{X})^2 - 2\hat{h}(X_i - \bar{X})(Y_i - \bar{Y}) \\ &= (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})^2 - \lambda^2(\hat{\delta}_i - \bar{\hat{\delta}})^2 - 2\hat{h}(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})(\hat{\delta}_i - \bar{\hat{\delta}}) \\ &\quad + 2\hat{\gamma}\hat{U}_{1i}(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}) - 2\hat{\gamma}\hat{\beta}\hat{U}_{1i}(\hat{\delta}_i - \bar{\hat{\delta}}).\end{aligned} \tag{4.3.6}$$

where $\hat{\gamma} = \hat{\beta} - \hat{h}$. Equation (4.3.5) suggests that our bootstrap estimator for $\hat{h}$ is

$$\hat{h}^* = \hat{h} + \frac{1}{2n}\hat{\mu}^*_{11}\hat{\mu}_{11}^{-2}\sum_{i=1}^{n}\hat{A}^*_i, \tag{4.3.7}$$

and using (4.2.12) that for $\hat{\beta}$ is

$$\hat{\beta}^* = \hat{h}^* + (\hat{h}^{2*} + \lambda^2)^{1/2}, \tag{4.3.8}$$

where $\hat{\mu}^*_{11} = n^{-1}\sum_{i=1}^{n}(X^*_i - \bar{X})(Y^*_i - \bar{Y})$. Note that $\sum_{i=1}^{n}\hat{A}_i = 0$. We now describe the resampling algorithm for the slope parameter in the structural linear relationship (4.2.1)–(4.2.6).

1. Given a data set (X_i, Y_i), $i = 1, \ldots, n$ and the estimator (4.2.19) for the slope of the model, compute the fitted values $(\hat{U}_1, \hat{U}_2)$ using (4.3.1).

2. Resample $\hat{\delta}^*_i$, with replacement from the set $\{r_1, \ldots, r_n\}$, where r_i are given in (4.3.3).

3. Independently of step (2), resample $\hat{\varepsilon}^*_i$ at random with replacement from the set $\{s_1, \ldots, s_n\}$, where s_i are given in (4.3.3).

4. Compute $\hat{A}_i^* = (\hat{\varepsilon}_i^* - \bar{\hat{\varepsilon}})^2 - \lambda^2(\hat{\delta}_i^* - \bar{\hat{\delta}})^2 - 2\hat{h}(\hat{\varepsilon}_i^* - \bar{\hat{\varepsilon}})(\hat{\delta}_i^* - \bar{\hat{\delta}}) + 2\hat{\gamma}\hat{U}_{1i}(\hat{\varepsilon}_i^* - \bar{\hat{\varepsilon}}) - 2\hat{\gamma}\hat{\beta}\hat{U}_{1i}(\hat{\delta}_i^* - \bar{\hat{\delta}})$.

5. Compute

$$\hat{h}^* = \hat{h} + \frac{1}{2n}\hat{\mu}_{11}^*\hat{\mu}_{11}^{-2}\sum_{i=1}^{n}\hat{A}_i^*, \tag{4.3.9}$$

$$\hat{\beta}^* = \hat{h}^* + (\hat{h}^{2*} + \lambda^2)^{1/2}, \tag{4.3.10}$$

where $\hat{\mu}_{11}^* = n^{-1}\sum_{i=1}^{n}(X_i^* - \bar{X})(Y_i^* - \bar{Y})$.

6. Repeat steps (1)–(5) above a large number of times, B, to obtain $\hat{\beta}_1^*, \ldots, \hat{\beta}_B^*$.

7. Calculate estimate of $Var_F(\hat{\beta})$ with

$$var_*(\hat{\beta}^*) = E_*(\hat{\beta}^* - E_*\hat{\beta}^*)^2, \tag{4.3.11}$$

where E_* denotes expectation with respect to bootstrap sampling which can be approximated by

$$var_*(\hat{\beta}^*) = (B-1)^{-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \hat{\beta}_{(\cdot)})^2, \tag{4.3.12}$$

where $\hat{\beta}_{(\cdot)}^* = \sum_{b=1}^{B}\hat{\beta}_b^*/B$.

8. In the b-th run, calculate the bootstrap-t:

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\sqrt{var_J^*(\hat{\beta}_b^*)}}, \quad b = 1, \ldots, B \tag{4.3.13}$$

where $var_J^*(\hat{\beta}_b^*)$ is the jackknife variance estimator of $\hat{\beta}_b^*$ given by

$$var_J^*(\hat{\beta}_b^*) = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\beta}_b^{*(i)} - \hat{\beta})^2 \tag{4.3.14}$$

with $\hat{\beta}_b^{*(i)}$ being the estimator of the same functional form as $\hat{\beta}_b^*$, but computed from the reduced sample of size $n-1$ obtained by omitting the i-th observation.

The asymptotic distribution of the bootstrap estimator proposed above is given in the following two theorems.

Theorem 4.3. *In the structural relationship (4.2.1)-(4.2.5), we assume $E(X^6 + Y^6) < \infty$. Then, $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ converges in distribution to a normal random variable with zero mean and variance $\hat{\Sigma}_{22}$, where $\hat{\beta}^*$ is the bootstrap estimator of $\hat{\beta}$ resulting from the proposed resampling procedure in Section 4.3.1. and $\hat{\Sigma}_{22}$ as defined in (4.2.28).*

Proof. Let $\bar{\hat{A}}^* = n^{-1}\sum_{i=1}^{n}\hat{A}_i^*$ where $\hat{A}_i^*$ is defined in step (4) in the proposed bootstrap procedure above and write

$$\hat{h}^* = \hat{h} + \frac{\hat{\mu}_{11}^*}{\hat{\mu}_{11}}\frac{\bar{\hat{A}}^*}{2\hat{\mu}_{11}}. \tag{4.3.15}$$

We observe that $\bar{\hat{A}}^*$ is the mean of an i.i.d. sample with population mean $\bar{\hat{A}}$, where $\bar{\hat{A}} = n^{-1}\sum_{i=1}^{n}\hat{A}_i$. Then, by the central limit theorem, we have

$$\sqrt{n}(\bar{\hat{A}}^* - \bar{\hat{A}}) \xrightarrow{d} N(0, \hat{\sigma}_A^2), \tag{4.3.16}$$

with

$$\begin{aligned}\hat{\sigma}_A^2 &= \frac{1}{n}\sum_{i=1}^{n}(\hat{A}_i - \bar{\hat{A}})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\hat{A}_i^2 - \bar{\hat{A}}^2 \\ &= \hat{\mu}_{04} + \lambda^4\hat{\mu}_{40} + 2\hat{\mu}_{22}(2\hat{h}^2 - \lambda^2) - 4\hat{h}(\hat{\mu}_{13} - \lambda^2\hat{\mu}_{31}) + o_p(n^{-1}) \quad (4.3.17)\end{aligned}$$

By the law of large numbers $\hat{\mu}_{11}^* \xrightarrow{p} \hat{\mu}_{11}$ and by Slutsky's theorem, we have

$$\sqrt{n}(\hat{h}^* - \hat{h}) \overset{d}{=} \frac{1}{2\hat{\mu}_{11}}\sqrt{n}(\bar{\hat{A}}^* - \bar{\hat{A}}) \tag{4.3.18}$$

which implies that

$$\sqrt{n}(\hat{h}^* - \hat{h}) \xrightarrow{d} N(0, \frac{1}{4\hat{\mu}_{11}^2}\hat{\sigma}_A^2). \tag{4.3.19}$$

We conclude that $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ converges to $N(0, \hat{\Sigma}_{22})$ in distribution by the delta method (see Bishop, Fienberg, and Holland, 1975). □

Theorem 4.4. *Assume the conditions of Theorem 4.3 hold. Then, under the proposed sampling procedure described in Section 4.3.1,*

(i) $E_F[E_*(\hat{\beta}^* - \hat{\beta})] = -E_F(\hat{\beta} - \beta) + O(n^{-2})$.

In addition, if the joint distribution of X and Y is symmetric, we have

(ii) $E_F[E_*(\hat{\beta}^* - \hat{\beta})^2] = E_F(\hat{\beta} - \beta)^2 + O(n^{-2})$ *and*

(iii) $E_F[E_*(\hat{\beta}^* - \hat{\beta})^3] = E_F(\hat{\beta} - \beta)^3 + O(n^{-3})$,

where E_* represents expectation with respect to the distribution induced by bootstrap sampling described in Section 4.3.1.

Proof. See Appendix B. □

4.3.2 The Weighted Bootstrap

Wu (1986) proposed a weighted bootstrap method in the context of classical regression Generally, the method entails first taking i.i.d. samples $\{t_i; i = 1, 2, ..., n\}$ from an external population having mean 0 and variance 1 and then generating bootstrap data by setting

$$y_i^* = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} + t_i e_i, \quad i = 1, 2, \ldots, n, \tag{4.3.20}$$

where $\boldsymbol{x}_i$ is a $p \times 1$ deterministic vector, $\hat{\boldsymbol{\beta}}$ is the $p \times 1$ vector of least squares estimators of $\boldsymbol{\beta}$ and $e_i = y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$. Liu (1988) suggested that another restriction needed to be imposed on t_i namely, $E(t_i^3) = 1$, to modify Wu's bootstrap procedure to share the usual second order asymptotic properties of the classical bootstrap. We begin by using the idea of the weighted bootstrap to construct a bootstrap procedure in the context of measurement error model. We describe below the weighted resampling algorithm for the slope parameter in the structural linear relationship (4.2.1)–(4.2.6).

1. Generate $D_i, i = 1, \ldots, n$; i.i.d. random variables with gamma distribution having density $g_D(x) = [p^q/(q-1)!]x^{q-1}e^{-px}I_{\{x>0\}}$, where $p = 2$ and $q = 4$.

2. Compute the bootstrap data, for $i = 1, \ldots, n$

$$\hat{\mu}_{02}^{*(i)} = \hat{\mu}_{02} + t_i[(Y_i - \bar{Y})^2 - \hat{\mu}_{02}], \tag{4.3.21}$$

$$\hat{\mu}_{20}^{*(i)} = \hat{\mu}_{20} + t_i[(X_i - \bar{X})^2 - \hat{\mu}_{20}], \tag{4.3.22}$$

$$\hat{\mu}_{11}^{*(i)} = \hat{\mu}_{11} + t_i[(X_i - \bar{X})(Y_i - \bar{Y}) - \hat{\mu}_{11}], \tag{4.3.23}$$

where $t_i = D_i - E(D_i)$.

3. Obtain bootstrap estimates $\hat{\mu}_{20}^* = n^{-1}\sum_{i=1}^n \hat{\mu}_{20}^{*(i)}$, $\hat{\mu}_{02}^* = n^{-1}\sum_{i=1}^n \hat{\mu}_{02}^{*(i)}$ and $\hat{\mu}_{11}^* = n^{-1}\sum_{i=1}^n \hat{\mu}_{11}^{*(i)}$, computed from the bootstrap data $\hat{\mu}_{02}^{*(i)}$ $\hat{\mu}_{20}^{*(i)}$ and $\hat{\mu}_{11}^{*(i)}$, respectively.

4. Obtain bootstrap estimates $\hat{h}^*$ and $\hat{\beta}^*$ along the lines of $\hat{h}$ and $\hat{\beta}$.

5. Repeat steps (1)–(5) above a large number of times, B, to obtain $\hat{\beta}_1^*, \ldots, \hat{\beta}_B^*$.

6. Calculate estimate of standard error of $\hat{\beta}^*$ with

$$var(\hat{\beta}) = E_t(\hat{\beta}^* - E_t\hat{\beta}^*)^2, \tag{4.3.24}$$

where E_t denotes expectation with respect to the weighted bootstrap sampling which can be approximated by

$$var_t(\hat{\beta}^*) = (B-1)^{-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \hat{\beta}_{(\cdot)})^2, \tag{4.3.25}$$

where $\hat{\beta}_{(\cdot)}^* = \sum_{b=1}^B \hat{\beta}_b^*/B$.

7. In the b-th run, calculate the bootstrap-t, defined as

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\sqrt{var_t(\hat{\beta}_b^*)}}, \quad b = 1, \ldots, B, \tag{4.3.26}$$

where $var_t(\hat{\beta}_b^*)$ is the bootstrap variance applied to b-th bootstrap sample and given by

$$var_t(\hat{\beta}_b^*) = \frac{\hat{\beta}^2}{4n\hat{\mu}_{11}^2(\hat{h}^2 + \lambda^2)}\sum_{i=1}^n t_i^2\hat{A}_i^2. \tag{4.3.27}$$

Asymptotic properties of the weighted bootstrap estimators are given in the next two theorems.

Theorem 4.5. *In the structural relationship (4.2.1)-(4.2.5), we assume $E(X^6 + Y^6) < \infty$. Then, $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ converges in distribution to a normal random variable with zero mean and variance $\hat{\Sigma}_{22}$, where $\hat{\beta}^*$ is the bootstrap estimator of $\hat{\beta}$ resulting from the proposed resampling procedure in Section 4.3.2 and $\hat{\Sigma}_{22}$ is as defined in (4.2.28).*

Proof. Consider

$$\begin{aligned}\hat{h}^* &= \hat{h} + \frac{1}{2\hat{\mu}_{11}^*}(\hat{\mu}_{02}^* - \lambda^2\hat{\mu}_{20}^*) - \frac{1}{2\hat{\mu}_{11}}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20})\\ &= \frac{1}{2\hat{\mu}_{11}^*}(\hat{\mu}_{02}^* - \lambda^2\hat{\mu}_{20}^* - 2\hat{h}\hat{\mu}_{11}^*) - \frac{1}{2\hat{\mu}_{11}}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2\hat{h}\hat{\mu}_{11}),\end{aligned}$$

or equivalently, we have

$$\begin{aligned}\hat{h}^* - \hat{h} &= \frac{1}{2\hat{\mu}_{11}^*}(\hat{\mu}_{02}^* - \lambda^2\hat{\mu}_{20}^* - 2\hat{h}\hat{\mu}_{11}^*) - \frac{1}{2\hat{\mu}_{11}}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2\hat{h}\hat{\mu}_{11})\\ &= \frac{1}{2\hat{\mu}_{11}}\{(\hat{\mu}_{02}^* - \hat{\mu}_{02}) - \lambda^2(\hat{\mu}_{20}^* - \hat{\mu}_{20}) - 2\hat{h}(\hat{\mu}_{11}^* - \hat{\mu}_{11})\}\frac{\hat{\mu}_{11}}{\hat{\mu}_{11}^*}\\ &+ \frac{1}{2\hat{\mu}_{11}^*}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2\hat{h}\hat{\mu}_{11}) - \frac{1}{2\hat{\mu}_{11}}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2\hat{h}\hat{\mu}_{11}).\end{aligned}$$

Since $\hat{\mu}_{11}^*$ is the mean of i.i.d samples, by the law of large numbers $\hat{\mu}_{11}^* \xrightarrow{p} \hat{\mu}_{11}$ and by Slutsky's theorem, we have

$$\sqrt{n}(\hat{h}^* - \hat{h}) \overset{d}{=} \frac{\sqrt{n}}{2\hat{\mu}_{11}}\{(\hat{\mu}_{02}^* - \hat{\mu}_{02}) - \lambda^2(\hat{\mu}_{20}^* - \hat{\mu}_{20}) - 2\hat{h}(\hat{\mu}_{11}^* - \hat{\mu}_{11})\} \quad (4.3.28)$$

and

$$E_t(\hat{\mu}^*_{02} - \hat{\mu}_{02})^2 = n^{-1}(\hat{\mu}_{04} - \hat{\mu}^2_{20}), \tag{4.3.29}$$

$$E_t(\hat{\mu}^*_{20} - \hat{\mu}_{20})^2 = n^{-1}(\hat{\mu}_{40} - \hat{\mu}^2_{02}), \tag{4.3.30}$$

$$E_t(\hat{\mu}^*_{11} - \hat{\mu}_{11}) = n^{-1}(\hat{\mu}_{22} - \hat{\mu}^2_{11}), \tag{4.3.31}$$

$$E_t(\hat{\mu}^*_{02} - \hat{\mu}_{02})(\hat{\mu}^*_{20} - \hat{\mu}_{20}) = n^{-1}(\hat{\mu}_{22} - \hat{\mu}_{02}\hat{\mu}_{20}), \tag{4.3.32}$$

$$E_t(\hat{\mu}^*_{02} - \hat{\mu}_{02})(\hat{\mu}^*_{11} - \hat{\mu}_{11}) = n^{-1}(\hat{\mu}_{13} - \hat{\mu}_{02}\hat{\mu}_{11}), \tag{4.3.33}$$

$$E_t(\hat{\mu}^*_{20} - \hat{\mu}_{20})(\hat{\mu}^*_{11} - \hat{\mu}_{11}) = n^{-1}(\hat{\mu}_{31} - \hat{\mu}_{20}\hat{\mu}_{11}). \tag{4.3.34}$$

We have

$$\sqrt{n}(\hat{h}^* - \hat{h}) \xrightarrow{d} N(0, \frac{1}{4\hat{\mu}^2_{11}}\hat{\sigma}^2_A), \tag{4.3.35}$$

with $\hat{\sigma}^2_A$ is given by (4.3.17). We conclude that $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ converges in distribution to $N(0, \hat{\Sigma}_{22})$, where $\hat{\Sigma}_{22}$ is given by (4.2.28) by the delta method (see Bishop, Fienberg, and Holland, 1975). □

Theorem 4.6. *Assume the conditions of Theorem 4.5 hold. Then, under the proposed weighted sampling procedure described in Section 4.3.2.*

(i) $E_F[E_t(\hat{\beta}^* - \hat{\beta})] = E_F(\hat{\beta} - \beta) + O(n^{-2})$,

(ii) $E_F[E_t(\hat{\beta}^* - \hat{\beta})^2] = E_F(\hat{\beta} - \beta)^2 + O(n^{-2})$,

(iii) $E_F[E_t(\hat{\beta}^* - \hat{\beta})^3] = E_F(\hat{\beta} - \beta)^3 + O(n^{-3})$,

where E_t represents the expectation with respect to the distribution induced by bootstrap sampling described in Section 4.3.2.

Proof. See Appendix C. □

4.4 A Simulation Study

This section describes a simulation study which compares the performance of the proposed resampling methods to various other methods. Along the lines of simulation study done by Linder and Babu (1994), a total of 12 different cases are used to generate data sets from (4.2.1) and (4.2.2) with $n = 20, 30$, $\alpha = 1, \beta = 2$ and $\lambda^2 = 1$. The U_{1i}'s are generated according to the following four "design" distributions:

1. Uniform(1.5,8.5),

2. Normal(5,4),

3. (5-$\sqrt{2}$) + $\sqrt{2}W$, where W is chi-square(1),

4. $(2/3)N(4, 2.5) + (1/3)N(7, 26)$ (mixture normal).

For each design, independent pairs of errors $(\delta_i, \varepsilon_i)$ are generated according to the following distributions:

1. $N(0, 0.48)$,

2. Double exponential(0.49) i.e. with density $f(\delta) = 1.02 \exp(-|\delta|/0.49)$,

3. Contaminated normal: $0.1N(1.8, 0.84) + 0.9N(-0.2, 0.04)$

4. "Moderate" heteroscedastic normal: $(\delta_i, \varepsilon_i) \sim \{0.4N(0,1); i = 1, \ldots, 5\}$, $\{0.6N(0,1); i = 6, \ldots, 10\}$, $\{0.8N(0,1); i = 11, \ldots, 15\}$, $\{N(0,1); i = 16, ..., 20\}$ for $n = 20$ and $(\delta_i, \varepsilon_i) \sim \{0.4N(0,1); i = 1, \ldots, 6\}$,

$\{0.6N(0,1); i = 7, \dots, 12\}$, $\{0.8N(0,1); i = 13, \dots, 19\}$, $\{N(0.1); i = 20, \dots, 25\}$, $\{N(0,1); i = 26, \dots, 30\}$ for $n = 30$.

5. "Heavy" heteroscedastic normal: $(\delta_i, \varepsilon_i) \sim \{N(0,1); i = 1.\dots.10\}$, $\{2.0N(0,1); i = 11, \dots, 20\}$ for $n = 20$ and $(\delta_i, \varepsilon_i) \sim \{N(0,1); i = 1, \dots, 15\}$, $\{2.0N(0,1); i = 16, \dots, 30\}$ for $n = 30$.

For every case, Monte Carlo expectations are computed based on $N = 10,000$ simulations. Within each simulation, Monte Carlo expectations with respect to bootstrap are computed based on $B = 1,000$ bootstraps. The absolute bias of $\hat{\beta}$ and the confidence intervals for β are calculated as follows:

1. **Normal Approx.** : Absolute bias $= N^{-1} \sum_{n=1}^{N} |\hat{\beta}_n - \beta|$. Large sample confidence interval for the linear structural error model:

$$\hat{\beta} \pm z_{1-\alpha/2} \sqrt{var_F(\hat{\beta})}$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ percentile of the standard normal distribution and $var(\hat{\beta}) = n^{-1}\hat{\Sigma}_{22}$ with $\hat{\Sigma}_{22}$ given by (4.2.28).

For the next three bootstrap methods, the absolute bias of $\hat{\beta}$ is computed as $(NB)^{-1} \sum_{n=1}^{N} \sum_{b=1}^{B} |\hat{\beta}_{nb} - \beta|$.

2. **LINDER & BABU:** The $100(1 - \alpha)$ % confidence interval for β is given by

$$\left[\hat{\beta} - t^{*(LB)}_{1-\alpha/2} \sqrt{var_F(\hat{\beta})}, \hat{\beta} - t^{*(LB)}_{\alpha/2} \sqrt{var_F(\hat{\beta})}\right],$$

where $t^{*(LB)}_{\alpha/2}$ and $t^{*(LB)}_{1-\alpha/2}$ are the percentiles of the histogram of the Studentized values

$$t_b^{*(LB)} = \frac{(\hat{\beta}_b^* - \hat{\beta})}{\sqrt{\hat{\Phi}^*(\hat{\beta}_b^*)}}, \quad b = 1, \dots, B.$$

3. **PROPOSED METHOD 1:** The $100(1-\alpha)$ % confidence interval for β is given by

$$\left[\hat{\beta} - t^{*(1)}_{1-\alpha/2}\sqrt{var_F(\hat{\beta})}, \hat{\beta} - t^{*(1)}_{\alpha/2}\sqrt{var_F(\hat{\beta})}\right],$$

where $t^{*(1)}_{\alpha/2}$ and $t^{*(1)}_{1-\alpha/2}$ are the percentiles of the histogram of the Studentized values

$$t_b^{*(1)} = \frac{(\hat{\beta}_b^* - \hat{\beta})}{\sqrt{var_J^*(\hat{\beta}_b^*)}}, \quad b = 1, \dots, B,$$

and $var_J^*(\hat{\beta}_b^*)$ is the jackknife variance estimator of $\hat{\beta}^*$ in the b-th bootstrap sample and given by (4.3.27).

4. **PROPOSED METHOD 2:** The $100(1-\alpha)$ % confidence interval for $\hat{\beta}$ is given by

$$\left[\hat{\beta} - t^{*(2)}_{1-\alpha/2}\sqrt{var_F(\hat{\beta})}, \hat{\beta} - t^{*(2)}_{\alpha/2}\sqrt{var_F(\hat{\beta})}\right],$$

where $t^{*(2)}_{\alpha/2}$ and $t^{*(2)}_{1-\alpha/2}$ are the percentiles of the histogram of the Studentized values

$$t_b^{*(2)} = \frac{(\hat{\beta}_b^* - \hat{\beta})}{\sqrt{var_t(\hat{\beta}_b^*)}}, \quad b = 1, \dots, B,$$

and $var_t(\hat{\beta}_b^*)$ is the bootstrap variance estimator of $\hat{\beta}^*$ in the b-th bootstrap sample and given by (4.3.27).

For each of $N = 10,000$ simulations, we compute 90%, 95% and 99% confidence limits for β and their lower and upper tail frequencies (in percents). The tail frequencies represent tail probabilities and hence, the coverage probabilities of the confidence intervals. We also compute the confidence lengths (median). The following summaries are reported in Tables 4.1–4.14 and are calculated according to the above methods.

LOW: Error rate in the lower tail defined by

Normal Approx. : $\sum_{i=1}^{N} I_{L_{iN}}(\beta)/N$, where $L_{iN} = (\hat{\beta}_i - z_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)}, \infty)$. The three bootstrap methods are given by $\sum_{i=1}^{N} I_{L_i}(\beta)/N$, where $L_i = (\hat{\beta}_i - t^*_{1-\alpha/2}\sqrt{var_F(\hat{\beta}_i)}, \infty)$ and $I_A(\cdot)$ in an indicator function defined by

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{otherwise.} \end{cases}$$

UP: Error rate in the upper tail defined by

Normal Approx. : $\sum_{i=1}^{N} I_{U_{iN}}(\beta)/N$, where $U_{iN} = (-\infty, \hat{\beta}_i + z_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)})$. The three bootstrap methods are given by $\sum_{i=1}^{N} I_{U_i}(\beta)/N$, where $U_i = (-\infty, \hat{\beta}_i - t^*_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)})$.

CP: The coverage probability defined by

Normal Approx. : $\sum_{i=1}^{N} I_{C_{iN}}(\beta)/N$, where $C_{iN} = (\hat{\beta}_i - z_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)}, \hat{\beta}_i + z_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)})$. The three bootstrap methods are given by $\sum_{i=1}^{N} I_{C_i}(\beta)/N$ where $C_i = (\beta \in (\hat{\beta}_i - t^*_{1-\alpha/2}\sqrt{var_F(\hat{\beta}_i)}, \hat{\beta}_i - t^*_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)})$ with $t^*_{1-\alpha/2} = t^{*(LB)}_{1-\alpha/2}, t^{*(1)}_{1-\alpha/2}$ and $t^{*(2)}_{1-\alpha/2}$ and $t^*_{\alpha/2} = t^{*(LB)}_{\alpha/2}, t^{*(1)}_{\alpha/2}$ and $t^{*(2)}_{\alpha/2}$.

LGT: The median length of confidence intervals.

4.4.1 Summary of findings and conclusions

Table 4.15 reports the summary of the findings from Tables 4.1–4.14. For $n = 20$ the coverage probabilities for 90% are in the range of 81.23–87.55% for Normal Approx., 86–89.94% for Linder and Babu's method, 89.93-91.78% for proposed method 1 and 89.65-93.56% for proposed method 2. For 95%, they are in the range of 87.33–92.64% for Normal Approx., 92.86–95.16% for Linder and Babu's method, 94.98–97.32% for proposed method 1 and 93.60–97.03% for proposed method 2. For 99% coverage probability, they are in the range of 94.65–97.66% for Normal Approx., 98.22–99.07% for Linder and Babu's method, 98.88–99.33% for proposed method 1 and 97.72–99.54% for proposed method 2. Similarly, the coverage probabilities for $n = 30$ are in the same range.

The obvious conclusion to be drawn from Tables 4.11–4.14 is that the traditional large sample intervals are not corrected for the skewness of the distribution of $\hat{\beta}$. Its coverage probabilities are understated by their respective nominal rates. Our bootstrap procedures perform well throughout even in comparison with the normal theory estimates in normal situations, i.e., they have better coverage accuracy than the normal approximation. The tail errors rates show that all bootstrap methods result in heavier upper tail indicating a skewed distribution of $\hat{\beta}$ with a long tail to the left. This suggests that the use of bootstrap histograms to construct confidence interval is more appropriate. The weighted bootstrap tends to have inflated coverage probabilities and have long lengths, the reason being that the jackknife is not resistant to extreme values and perhaps data should be trimmed before jackknifing. It is not surprising that Linder and Babu's method does well here since they applied a correction factor in the bootstrap which makes their methods less appealing than ours, see remark on page 78. Another disadvantage of this method is that it is substantially more computer intensive than our proposed methods .

For the case of heteroscedastic errors, we present simulation results only for the case of uniform design with normal error distribution. These results appear to be robust against heteroscedascity and are enough to suggest that our methods are ahead compared to others in attaining their respective nominal coverage levels, as Tables 4.13–4.14 show.

Table 4.1: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is uniform with normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	4.43	8.76	86.81	0.539	0.139
		0.95	2.39	5.76	91.85	0.642	0.139
		0.99	0.56	2.51	96.93	0.844	0.139
	30	0.90	4.76	7.59	87.65	0.441	0.112
		0.95	2.41	4.58	93.01	0.526	0.112
		0.99	0.55	2.04	97.41	0.691	0.112
LINDER & BABU	20	0.90	3.29	7.58	89.13	0.583	0.196
		0.95	1.70	3.75	94.55	0.718	0.196
		0.99	0.48	0.82	98.70	1.025	0.196
	30	0.90	3.03	8.24	88.73	0.455	0.163
		0.95	1.54	4.30	94.16	0.554	0.163
		0.99	0.39	1.00	98.61	0.769	0.163
PROPOSED METHOD 1	20	0.90	1.34	7.56	91.10	0.643	0.196
		0.95	0.55	3.82	95.63	0.794	0.196
		0.99	0.11	0.87	99.02	1.122	0.196
	30	0.90	1.29	8.64	90.07	0.492	0.163
		0.95	0.46	4.52	95.02	0.601	0.163
		0.99	0.07	1.20	98.73	0.833	0.163
PROPOSED METHOD 2	20	0.90	2.67	7.25	90.08	0.605	0.132
		0.95	1.17	4.79	94.04	0.714	0.132
		0.99	0.26	1.98	97.76	0.946	0.132
	30	0.90	3.19	6.92	89.89	0.474	0.107
		0.95	1.49	4.38	94.13	0.556	0.107
		0.99	0.30	1.92	97.78	0.723	0.107

Table 4.2: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is uniform with double exponential error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	5.00	8.57	86.43	0.526	0.143
		0.95	2.46	5.43	92.11	0.627	0.143
		0.99	0.60	2.08	97.32	0.824	0.143
	30	0.90	4.71	7.39	87.90	0.435	0.115
		0.95	2.31	4.48	93.21	0.518	0.115
		0.99	0.46	1.53	98.01	0.681	0.115
LINDER & BABU	20	0.90	3.81	7.29	88.90	0.563	0.199
		0.95	1.87	3.42	94.71	0.686	0.199
		0.99	0.45	0.64	98.91	0.966	0.199
	30	0.90	3.17	7.86	88.97	0.446	0.165
		0.95	1.59	4.14	94.27	0.541	0.165
		0.99	0.41	0.69	98.90	0.740	0.165
PROPOSED METHOD 1	20	0.90	1.44	7.11	91.45	0.623	0.199
		0.95	0.60	3.23	96.17	0.761	0.199
		0.99	0.11	0.56	99.33	1.065	0.199
	30	0.90	1.25	8.00	90.75	0.486	0.165
		0.95	0.49	4.26	95.25	0.589	0.165
		0.99	0.11	0.78	99.11	0.807	0.165
PROPOSED METHOD 2	20	0.90	2.65	6.85	90.50	0.590	0.132
		0.95	1.20	4.18	94.62	0.694	0.132
		0.99	0.33	1.49	98.18	0.920	0.132
	30	0.90	3.08	6.51	90.41	0.469	0.108
		0.95	1.51	4.07	94.42	0.550	0.108
		0.99	0.22	1.29	98.49	0.714	0.108

Table 4.3: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is uniform with contaminated normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	6.72	6.99	86.29	0.069	0.018
		0.95	3.82	4.17	92.01	0.082	0.018
		0.99	1.36	1.50	97.14	0.108	0.018
	30	0.90	6.74	6.49	86.77	0.058	0.015
		0.95	3.61	3.74	92.65	0.069	0.015
		0.99	1.05	1.28	97.67	0.091	0.015
LINDER & BABU	20	0.90	4.90	5.16	89.94	0.077	0.022
		0.95	2.49	2.35	95.16	0.094	0.022
		0.99	0.60	0.60	98.80	0.133	0.022
	30	0.90	5.15	5.29	89.56	0.063	0.019
		0.95	2.45	2.69	94.86	0.076	0.019
		0.99	0.65	0.59	98.76	0.105	0.019
PROPOSED METHOD 1	20	0.90	3.85	4.69	91.46	0.082	0.075
		0.95	1.95	2.10	95.95	0.100	0.075
		0.99	0.45	0.46	99.09	0.142	0.075
	30	0.90	4.33	4.91	90.76	0.065	0.070
		0.95	2.05	2.45	95.50	0.079	0.070
		0.99	0.43	0.55	99.02	0.109	0.070
PROPOSED METHOD 2	20	0.90	4.97	5.38	89.65	0.076	0.016
		0.95	3.05	3.35	93.60	0.089	0.016
		0.99	1.14	1.14	97.72	0.117	0.016
	30	0.90	5.26	5.59	89.15	0.062	0.014
		0.95	3.07	3.23	93.70	0.073	0.014
		0.99	0.93	1.17	97.90	0.095	0.014

Table 4.4: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is normal with normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	5.30	9.83	84.87	0.543	0.148
		0.95	2.96	6.74	90.30	0.647	0.148
		0.99	0.83	3.15	96.02	0.850	0.148
	30	0.90	5.45	8.50	86.05	0.440	0.115
		0.95	2.89	5.57	91.54	0.525	0.115
		0.99	0.79	2.20	97.01	0.689	0.115
LINDER & BABU	20	0.90	3.41	7.63	88.96	0.628	0.207
		0.95	1.87	3.97	94.16	0.768	0.207
		0.99	0.54	0.76	98.70	1.080	0.207
	30	0.90	3.20	8.69	88.11	0.470	0.168
		0.95	1.78	4.40	93.82	0.575	0.168
		0.99	0.53	0.92	98.55	0.808	0.168
PROPOSED METHOD 1	20	0.90	1.43	7.75	90.82	0.674	0.207
		0.95	0.75	3.99	95.26	0.836	0.207
		0.99	0.12	0.81	99.07	1.192	0.207
	30	0.90	1.57	8.94	89.49	0.508	0.168
		0.95	0.63	4.75	94.62	0.623	0.168
		0.99	0.10	1.06	98.84	0.872	0.168
PROPOSED METHOD 2	20	0.90	2.59	7.21	90.20	0.659	0.141
		0.95	1.20	4.50	94.30	0.791	0.141
		0.99	0.22	1.52	98.26	1.110	0.141
	30	0.90	3.22	6.93	89.85	0.503	0.110
		0.95	1.40	4.29	94.31	0.597	0.110
		0.99	0.34	1.42	98.24	0.807	0.110

Table 4.5: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is normal with double exponential error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	4.94	9.36	85.70	0.549	0.155
		0.95	2.45	6.09	91.46	0.655	0.155
		0.99	0.71	2.52	96.77	0.860	0.155
	30	0.90	5.12	7.73	87.15	0.439	0.116
		0.95	2.68	4.80	92.52	0.524	0.116
		0.99	0.59	1.61	97.80	0.688	0.116
LINDER & BABU	20	0.90	3.39	7.23	89.38	0.608	0.220
		0.95	1.65	3.62	94.73	0.750	0.220
		0.99	0.40	0.60	99.00	1.074	0.220
	30	0.90	3.35	8.03	88.62	0.462	0.172
		0.95	1.69	3.96	94.35	0.561	0.172
		0.99	0.44	0.72	98.84	0.778	0.172
PROPOSED METHOD 1	20	0.90	1.38	7.01	91.61	0.603	0.220
		0.95	0.65	3.50	95.85	0.733	0.220
		0.99	0.15	0.63	99.22	1.053	0.220
	30	0.90	1.39	8.22	90.39	0.502	0.172
		0.95	0.55	4.17	95.28	0.610	0.172
		0.99	0.10	0.71	99.19	0.844	0.172
PROPOSED METHOD 2	20	0.90	2.05	6.21	91.74	0.673	0.153
		0.95	0.98	3.64	95.38	0.810	0.153
		0.99	0.15	1.17	98.68	1.145	0.153
	30	0.90	2.69	5.81	91.50	0.501	0.112
		0.95	1.14	3.35	95.51	0.597	0.112
		0.99	0.23	0.99	98.78	0.802	0.112

Table 4.6: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is normal with contaminated normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	6.97	7.83	85.20	0.072	0.019
APPROX.		0.95	4.26	4.96	90.78	0.086	0.019
		0.99	1.47	2.01	96.52	0.113	0.019
	30	0.90	6.59	7.55	85.86	0.058	0.015
		0.95	3.89	4.54	91.57	0.070	0.015
		0.99	1.21	1.66	97.13	0.091	0.015
LINDER	20	0.90	4.71	5.42	89.87	0.084	0.018
& BABU		0.95	2.42	2.80	94.78	0.104	0.018
		0.99	0.58	0.51	98.91	0.149	0.018
	30	0.90	4.70	5.82	89.48	0.065	0.019
		0.95	2.25	2.87	94.88	0.080	0.019
		0.99	0.66	0.72	98.62	0.111	0.019
PROPOSED	20	0.90	3.74	4.75	91.51	0.090	0.050
METHOD 1		0.95	1.76	2.39	95.85	0.111	0.050
		0.99	0.34	0.46	99.20	0.161	0.050
	30	0.90	4.14	5.39	90.47	0.068	0.037
		0.95	1.93	2.60	95.47	0.083	0.037
		0.99	0.50	0.68	98.82	0.117	0.037
PROPOSED	20	0.90	4.60	5.32	90.08	0.085	0.018
METHOD 2		0.95	2.77	3.34	93.89	0.102	0.018
		0.99	0.72	1.05	98.23	0.139	0.018
	30	0.90	4.72	5.51	89.77	0.059	0.014
		0.95	2.63	3.09	94.28	0.071	0.014
		0.99	0.86	0.96	98.18	0.100	0.014

Table 4.7: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is chi-square with normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	7.68	9.89	82.43	0.364	0.106
		0.95	4.87	7.00	88.13	0.434	0.106
		0.99	1.58	3.43	94.99	0.571	0.106
	30	0.90	7.10	9.62	83.28	0.319	0.090
		0.95	4.17	6.34	89.49	0.380	0.090
		0.99	1.40	2.91	95.69	0.499	0.090
LINDER & BABU	20	0.90	3.91	6.51	89.55	0.457	0.142
		0.95	1.94	3.35	94.71	0.565	0.142
		0.99	0.37	0.76	98.87	0.812	0.142
	30	0.90	3.33	7.67	89.00	0.376	0.126
		0.95	1.76	4.03	94.21	0.462	0.126
		0.99	0.39	0.92	98.69	0.648	0.126
PROPOSED METHOD 1	20	0.90	2.45	6.20	91.35	0.502	0.142
		0.95	1.06	3.15	97.32	0.629	0.142
		0.99	0.21	0.68	99.11	0.914	0.142
	30	0.90	2.31	7.74	89.95	0.404	0.126
		0.95	1.06	4.19	94.75	0.501	0.126
		0.99	0.25	1.03	98.72	0.709	0.126
PROPOSED METHOD 2	20	0.90	2.62	4.94	92.44	0.555	0.116
		0.95	1.13	2.67	96.20	0.706	0.116
		0.99	0.12	0.57	99.31	1.198	0.116
	30	0.90	2.78	5.12	92.10	0.442	0.091
		0.95	1.18	2.77	96.05	0.547	0.091
		0.99	0.15	0.65	99.20	0.828	0.091

Table 4.8: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is chi-square with double exponential error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	7.29	9.83	82.88	0.345	0.102
		0.95	4.46	6.61	88.93	0.411	0.102
		0.99	1.55	3.00	95.45	0.540	0.102
	30	0.90	7.07	8.61	84.32	0.292	0.084
		0.95	4.12	5.49	90.39	0.348	0.084
		0.99	1.32	2.03	96.65	0.457	0.084
LINDER & BABU	20	0.90	4.04	6.43	89.53	0.419	0.136
		0.95	2.22	3.20	94.58	0.517	0.136
		0.99	0.59	0.65	98.76	0.742	0.136
	30	0.90	3.39	6.49	90.12	0.340	0. 117
		0.95	1.72	3.13	95.15	0.417	0. 117
		0.99	0.40	0.53	99.07	0.584	0. 117
PROPOSED METHOD 1	20	0.90	2.76	5.79	91.45	0.459	0.136
		0.95	1.29	2.95	95.76	0.574	0.136
		0.99	0.33	0.56	99.11	0.835	0.136
	30	0.90	2.66	6.20	91.14	0.361	0.117
		0.95	1.05	3.27	95.68	0.449	0.117
		0.99	0.23	0.59	99.18	0.642	0.117
PROPOSED METHOD 2	20	0.90	2.77	4.33	92.90	0.509	0.121
		0.95	1.39	2.42	96.19	0.639	0.121
		0.99	0.24	0.53	99.23	1.042	0.121
	30	0.90	2.73	3.90	93.37	0.418	0.091
		0.95	1.20	1.72	97.08	0.523	0.091
		0.99	0.15	0.26	99.59	0.812	0.091

Table 4.9: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is chisquare with contaminated normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	9.24	9.53	81.23	0.048	0.014
APPROX.		0.95	6.16	6.51	87.33	0.057	0.014
		0.99	2.52	2.83	94.65	0.075	0.014
	30	0.90	7.84	8.10	84.06	0.040	0.011
		0.95	5.04	4.87	90.09	0.047	0.011
		0.99	2.06	1.66	96.28	0.062	0.011
LINDER	20	0.90	4.96	5.13	89.91	0.063	0.018
& BABU		0.95	2.41	2.47	95.12	0.078	0.018
		0.99	0.51	0.42	99.07	0.111	0.018
	30	0.90	4.77	4.78	90.45	0.048	0.014
		0.95	2.49	2.07	95.44	0.059	0.014
		0.99	0.63	0.37	99.00	0.082	0.014
PROPOSED	20	0.90	3.91	4.58	91.51	0.069	0.049
METHOD 1		0.95	1.87	2.05	96.08	0.088	0.049
		0.99	0.39	0.32	99.29	0.128	0.049
	30	0.90	4.20	4.47	91.33	0.051	0.048
		0.95	2.04	1.95	96.01	0.036	0.048
		0.99	0.52	0.34	99.14	0.090	0.048
PROPOSED	20	0.90	3.76	3.96	92.28	0.076	0.016
METHOD 2		0.95	1.76	1.67	96.57	0.097	0.016
		0.99	0.26	0.20	99.54	0.169	0.016
	30	0.90	4.01	3.85	92.14	0.054	0.011
		0.95	2.14	1.81	96.05	0.067	0.011
		0.99	0.47	0.30	99.23	0.100	0.011

Table 4.10: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is mixture normal with normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	5.78	10.04	84.18	0.538	0.148
APPROX.		0.95	3.34	6.81	89.85	0.642	0.148
		0.99	1.00	3.20	95.80	0.843	0.148
	30	0.90	5.52	8.43	86.05	0.441	0.117
		0.95	2.88	5.59	91.53	0.526	0.117
		0.99	0.77	2.28	96.95	0.691	0.117
LINDER	20	0.90	3.65	7.77	88.85	0.616	0.208
& BABU		0.95	2.01	3.88	94.11	0.765	0.208
		0.99	0.49	0.74	98.77	1.111	0.208
	30	0.90	2.98	8.46	88.56	0.475	0.169
		0.95	1.61	4.35	94.04	0.581	0.169
		0.99	0.48	0.90	98.62	0.819	0.169
PROPOSED	20	0.90	1.77	7.81	90.42	0.680	0.208
METHOD 1		0.95	0.74	4.03	95.23	0.845	0.208
		0.99	0.14	0.73	99.13	1.210	0.208
	30	0.90	1.45	8.76	89.79	0.513	0.169
		0.95	0.64	4.60	94.76	0.632	0.169
		0.99	0.17	1.07	98.76	0.885	0.169
PROPOSED	20	0.90	2.65	6.84	90.51	0.675	0.142
METHOD 2		0.95	1.26	4.24	94.50	0.818	0.142
		0.99	0.22	1.27	98.51	1.178	0.142
	30	0.90	3.03	6.64	90.33	0.511	0.111
		0.95	1.40	4.00	94.60	0.610	0.111
		0.99	0.39	1.42	98.19	0.828	0.111

Table 4.11: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is mixture normal with error double exponential distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	5.90	9.43	84.67	0.519	0.147
		0.95	3.15	6.18	90.67	0.618	0.147
		0.99	0.69	2.48	96.83	0.813	0.147
	30	0.90	5.40	7.96	86.64	0.428	0.115
		0.95	2.87	4.61	92.52	0.510	0.115
		0.99	0.61	1.54	97.85	0.670	0.115
LINDER & BABU	20	0.90	3.85	7.12	89.03	0.581	0.207
		0.95	1.92	3.56	94.52	0.715	0.207
		0.99	0.31	0.71	98.98	1.027	0.207
	30	0.90	3.39	8.05	88.56	0.453	0.166
		0.95	1.83	3.68	94.49	0.550	0.166
		0.99	0.47	0.67	98.86	0.762	0.166
PROPOSED METHOD 1	20	0.90	1.74	6.92	91.34	0.639	0.207
		0.95	0.74	3.36	95.90	0.789	0.207
		0.99	0.07	0.60	99.33	1.126	0.207
	30	0.90	1.65	8.16	90.19	0.490	0.166
		0.95	0.71	3.77	95.52	0.597	0.166
		0.99	0.14	0.75	99.11	0.823	0.166
PROPOSED METHOD 2	20	0.90	2.48	6.05	91.47	0.648	0.144
		0.95	1.10	3.48	95.42	0.783	0.144
		0.99	0.20	1.02	98.78	1.125	0.144
	30	0.90	2.72	5.54	91.74	0.492	0.110
		0.95	1.15	3.15	95.70	0.586	0.110
		0.99	0.30	0.84	98.86	0.794	0.110

Table 4.12: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of β. The design is mixture normal with contaminated error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	7.32	7.80	84.88	0.071	0.019
		0.95	4.35	4.93	90.72	0.084	0.019
		0.99	1.64	1.97	96.39	0.111	0.019
	30	0.90	6.15	6.58	87.27	0.061	0.016
		0.95	3.63	4.03	92.34	0.073	0.016
		0.99	1.40	1.36	97.24	0.095	0.016
LINDER & BABU	20	0.90	4.81	5.25	89.94	0.083	0.024
		0.95	2.32	2.69	94.99	0.102	0.024
		0.99	0.52	0.53	98.95	0.147	0.024
	30	0.90	4.42	5.20	90.38	0.068	0.020
		0.95	2.36	2.60	95.04	0.083	0.020
		0.99	0.59	0.50	98.91	0.115	0.020
PROPOSED METHOD 1	20	0.90	3.61	4.61	91.78	0.088	0.093
		0.95	1.83	2.17	96.00	0.110	0.093
		0.99	0.31	0.42	99.27	0.159	0.093
	30	0.90	3.83	4.88	91.29	0.070	0.066
		0.95	1.95	2.37	95.68	0.086	0.066
		0.99	0.44	0.43	99.13	0.121	0.066
PROPOSED METHOD 2	20	0.90	4.66	5.26	90.08	0.085	0.017
		0.95	2.68	3.05	94.27	0.101	0.017
		0.99	0.85	0.99	98.16	0.139	0.017
	30	0.90	4.66	4.96	90.38	0.069	0.015
		0.95	2.60	2.82	94.58	0.081	0.015
		0.99	0.83	0.87	98.30	0.109	0.015

Table 4.13: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of β. The design is uniform with "moderate" heteroscedastic normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	MED-LGT	BIAS
NORMAL APPROX.	20	0.90	4.72	7.73	87.55	0.395	0.102
		0.95	2.45	4.91	92.64	0.471	0.102
		0.99	1.01	2.40	97.66	0.619	0.102
	30	0.90	5.59	7.72	86.69	0.343	0.090
		0.95	2.91	4.80	92.29	0.408	0.090
		0.99	0.68	1.78	97.54	0.537	0.090
LINDER & BABU	20	0.90	4.20	7.35	88.45	0.418	0.131
		0.95	2.24	3.81	93.85	0.511	0.131
		0.99	0.66	0.88	98.46	0.715	0.131
	30	0.90	3.67	7.71	88.62	0.362	0.100
		0.95	1.99	3.98	94.03	0.439	0.100
		0.99	0.56	1.05	98.39	0.601	0.100
PROPOSED METHOD 1	20	0.90	2.45	7.07	90.48	0.452	0.131
		0.95	1.05	3.70	95.25	0.554	0.131
		0.99	0.28	0.84	98.88	0.776	0.131
	30	0.90	2.29	7.77	89.94	0.382	0.100
		0.95	1.05	4.14	94.81	0.463	0.100
		0.99	0.15	1.13	98.72	0.636	0.100
PROPOSED METHOD 2	20	0.90	4.39	2.05	93.56	0.422	0.093
		0.95	2.12	0.85	97.03	0.494	0.093
		0.99	0.49	0.23	99.28	0.648	0.093
	30	0.90	3.82	6.82	89.36	0.367	0.084
		0.95	2.12	4.22	93.66	0.431	0.084
		0.99	0.42	1.55	98.03	0.559	0.084

Table 4.14: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of β. The design is uniform with "heavy" heteroscedastic normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	MED-LGT	BIAS
NORMAL	20	0.90	4.19	9.94	85.87	0.847	0.234
APPROX.		0.95	2.23	6.82	90.95	1.009	0.234
		0.99	0.38	3.03	96.59	1.326	0.234
	30	0.90	3.80	9.00	87.20	0.685	0.180
		0.95	1.58	6.08	92.34	0.817	0.180
		0.99	0.21	2.51	97.28	1.073	0.180
LINDER	20	0.90	4.35	8.79	86.86	0.862	0.402
& BABU		0.95	2.60	4.54	92.86	1.060	0.402
		0.99	0.85	0.93	98.22	1.520	0.402
	30	0.90	2.63	10.23	87.14	0.680	0.236
		0.95	1.49	5.65	92.86	0.827	0.236
		0.99	0.39	1.37	98.24	1.140	0.236
PROPOSED	20	0.90	1.08	8.99	89.93	0.984	0.397
METHOD 1		0.95	0.45	4.57	94.98	1.199	0.397
		0.99	0.07	0.77	99.16	1.669	0.397
	30	0.90	0.50	10.69	88.81	0.761	0.236
		0.95	0.17	5.97	93.86	0.924	0.236
		0.99	0.01	1.47	98.52	1.264	0.236
PROPOSED	20	0.90	1.08	8.99	89.93	0.975	0.227
METHOD 2		0.95	0.45	4.57	94.98	1.159	0.227
		0.99	0.07	0.77	99.16	1.571	0.227
	30	0.90	1.90	8.21	89.89	0.743	0.173
		0.95	0.66	5.46	93.88	0.876	0.173
		0.99	0.06	2.23	97.71	1.149	0.173

Table 4.15: Summary of coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$ for Tables 4.1–4.14.

METHOD	n	$1-\alpha$	CP		LGT		BIAS	
			MIN	MAX	MIN	MAX	MIN	MAX
NORMAL APPROX.	20	0.90	81.23	87.55	0.048	0.847	0.014	0.234
		0.95	87.33	92.64	0.057	1.009	0.014	0.234
		0.99	94.65	97.66	0.075	1.326	0.014	0.234
	30	0.90	83.28	87.90	0.040	0.685	0.010	0.180
		0.95	89.49	93.21	0.047	0.817	0.010	0.180
		0.99	95.69	98.01	0.062	1.073	0.010	0.180
LINDER & BABU	20	0.90	86.86	89.94	0.063	0.862	0.180	0.402
		0.95	92.86	95.16	0.078	1.060	0.180	0.402
		0.99	98.22	99.07	0.111	1.520	0.180	0.402
	30	0.90	87.14	90.45	0.480	0.680	0.140	0.236
		0.95	92.86	95.44	0.059	0.827	0.140	0.236
		0.99	98.24	99.07	0.082	1.140	0.140	0.236
PROPOSED METHOD 1	20	0.90	89.93	91.78	0.069	0.984	0.490	0.397
		0.95	94.98	97.32	0.088	1.199	0.490	0.397
		0.99	98.88	99.33	0.128	1.669	0.490	0.397
	30	0.90	88.81	91.33	0.051	0.761	0.037	0.236
		0.95	93.86	96.01	0.036	0.924	0.037	0.236
		0.99	98.52	99.19	0.090	1.264	0.037	0.236
PROPOSED METHOD 2	20	0.90	89.65	93.56	0.076	0.975	0.016	0.227
		0.95	93.60	97.03	0.089	1.159	0.016	0.227
		0.99	97.72	99.54	0.117	1.571	0.016	0.227
	30	0.90	89.15	93.37	0.054	0.743	0.011	0.173
		0.95	93.66	97.08	0.067	0.876	0.011	0.173
		0.99	97.71	99.59	0.095	1.149	0.011	0.173

Chapter 5

One-way random effects model: To estimate exceedances over a threshold

5.1 Introduction

Random effects models, also known as variance-components models. are widely used in several different fields of research. In epidemiologic research, they are commonly used to measure the degree of familial resemblance with respect to biological characteristics and in genetics these models play a central role in estimating the heritability of selected traits in animal and plant populations. A comprehensive review of the developments in the area of variance components can be found in Khuri and Sahai (1985), Sahai, Khuri, and Kapadia (1985). Donner (1986), Searle, Casella, and McCulloch (1992) and Box and Tiao (1992).

In the usual analysis of the random effects model, the random effect term and the error term are assumed to be independently and normally distributed. The main objective in these analysis usually centers on the estimation of the variance components or some functions of them. Several methods, namely, Hartley and Rao (1967), Searle *et al.* (1992), Harville (1977), Rao and Kleffe (1988) have been proposed to estimate variance components. Townsend and Searle (1968), Harville (1977), Rao and Kleffe (1988), and Chaubey (1984) dealt with the minimum variance quadratic unbiased estimators of variance components (MINQUE). MINQUE estimation requires no distributional properties of the random effects or error term in the model. However, the estimators obtained by MINQUE are functions of *a priori* values. For inference on the ratio of variance components, see Spjøtvoll (1968), Seely and El-Bassiouni (1983), Harville and Fenech (1985), Khuri and Littell (1987). Prasad and Rao (1988) considered the situation in which the random effects and errors were not necessarily normally distributed and used the weighted jackknife method to obtain robust inference on the ratio of variance components.

Solomon (1989) used the random effects model and assumed normality on both components to estimate the expected number of exceedances and other quantities in systolic blood pressure over a given threshold from a sample of 16 individuals. She used parametric approach to this problem. However, her method has some limitations. It requires theoretical calculations, approximations and depends heavily on the distributions of random effects and random errors. In this chapter, we propose a weighted bootstrap procedure to estimate these quantities where in which minimal theoretical calculations are needed and inferences obtained are robust to distributional assumption on the random effects term and error term.

In Section 5.2, we review Solomon's method on estimation of exceedances over a threshold in a balanced random effects model. In Section 5.3, we propose a weighted bootstrap procedure along with theoretical justifications are also provided. A Monte Carlo study is presented in Section 5.4.

5.2 The Model

Consider the following unbalanced one-way random effects model

$$y_{ij} = \mu + u_{ij}, \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, n_i, \tag{5.2.1}$$

where

$$u_{ij} = v_i + e_{ij}, \tag{5.2.2}$$

with y_{ij} being the j-th observation in the i-th class, μ being an unknown parameter to be estimated and u_{ij} being the random error associated with y_{ij}. Here u_{ij} is assumed to be the sum of the random effects, v_i, associated with i-th class and the random errors, e_{ij}, associated with j-th observation for the i-th class. The random errors $\{e_{ij}\}$ are assumed to be independent, identically distributed with mean 0 and variance σ_e^2 ($i.e. e_{ij} \overset{ind}{\sim} (0, \sigma_e^2)$) and the random effects $\{v_i\}$ are independent, identically distributed with mean 0 and variance σ_v^2 ($i.e. v_i \overset{ind}{\sim} (0, \sigma_v^2)$). Further, $\{v_i\}$ and $\{e_{ij}\}$ are uncorrelated so that the variance-covariance structure of u_{ij} is given by

$$E(u_{ij}u_{i'j'}) = \begin{cases} \sigma_v^2 + \sigma_e^2, & \text{for } i = i' \text{ and } j = j' \\ \sigma_v^2, & \text{for } i = i' \text{ and } j \neq j' \\ 0 & \text{otherwise.} \end{cases} \tag{5.2.3}$$

Let h be a given threshold value, and define

$$I_{ij}(h) = \begin{cases} 1 & y_{ij} > h \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad T = \sum_{i=1}^{a} T_i, \text{ where } T_i = \sum_{j=1}^{n_i} I_{ij}(h). \quad (5.2.4)$$

The goal is to estimate the average number of exceedances, $E(T)$, defined as the expected number of values that exceed a given threshold, its variance, $Var(T)$ and the probability of no exceedance $Pr(T = 0)$. In this chapter, we propose a weighted bootstrap procedure to estimate these quantities. Theoretical justifications and finite sample properties of the proposed bootstrap procedure are also given.

Solomon (1989) assumed a balanced ($n_i = n$ for all i) one-way random effects model with $v_i \overset{ind}{\sim} N(0, \sigma_v^2)$, $e_{ij} \overset{ind}{\sim} N(0, \sigma_e^2)$ and derived the following expressions:

1. The expected number T of exceedances over a threshold h is given by

$$E(T) = n\Phi\left(\frac{\mu - h}{\sigma}\right), \quad (5.2.5)$$

with the variance

$$Var(T) = n\Phi\left(\frac{\mu - h}{\sigma}\right)\left(1 - \Phi\left(\frac{\mu - h}{\sigma}\right)\right)\{1 + (n-1)\rho_I\}, \quad (5.2.6)$$

where

$$\rho_I = \frac{\Phi_2\left(\frac{\mu-h}{\sigma}, \frac{\mu-h}{\sigma}; \rho = \frac{\sigma_v^2}{\sigma^2}\right) - \Phi^2\left(\frac{\mu-h}{\sigma}\right)}{\Phi\left(\frac{\mu-h}{\sigma}\right)\left(1 - \Phi\left(\frac{\mu-h}{\sigma}\right)\right)}, \quad (5.2.7)$$

with $\Phi_2(\cdot)$ denoting the standardized bivariate normal distribution function with correlation ρ.

2. The probability of observing no exceedances is given by

$$Pr(T=0) = \int_{-\infty}^{\infty} \phi(x) \left\{ \Phi\left(\frac{\gamma - x}{\tau}\right) \right\}^n dx, \qquad (5.2.8)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal probability density and distribution function, respectively, $\gamma = (h-\mu)/\sigma_v$, $\tau = \sigma_e/\sigma_v$, and $\sigma^2 = \sigma_v^2 + \sigma_e^2$.

The estimates of (5.2.5)–(5.2.8) can be obtained by replacing μ, σ_v^2 and σ_e^2 by their respective maximum likelihood estimators and the integral (5.2.8) can be evaluated by numerical approximation using these estimates. Note that this method requires theoretical calculations and approximations and heavily depends on the normality assumption on v_i and e_{ij}.

In the next section, the use of a weighted bootstrap is considered to estimate the above parameters.

5.3 The Proposed Method

The use of the bootstrap method for estimating the unknown distribution of pivotal quantities to obtain robust confidence intervals for unknown parameters has become an important topic of recent statistical research both in theory and applications. For detailed discussions on this subject, we refer readers to Efron (1982, 1987) and Wu (1986). In this section, we explore a weighted bootstrap procedure (see Wu, 1986 and Liu, 1988) to estimate the expected number of exceedances by noting that the above model is unbalanced and warrants the use of a bootstrap procedure suitable for non-i.i.d. models. In the context of drawing inferences for regression models, Wu (1986) suggested a modification to the classical bootstrap procedure to accommodate the non-i.i.d. nature of the model by drawing i.i.d. observations from an external population having mean zero and unit variance. Liu (1988) proposed to draw i.i.d. observations $\{t_i\}$ from a population with $Et_i = 0$ and $Et_i^2 = Et_i^3 = 1$. This modification made Wu's bootstrap procedure share the usual second order asymptotic properties of the classical bootstrap for i.i.d. models.

To apply Liu's weighted bootstrap procedure to our problem, we need the following notation:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}, \quad \bar{y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}, \quad \hat{\alpha}_i = 1 - \sqrt{\frac{1}{1+n_i\hat{\Delta}}}, \tag{5.3.1}$$

$$N = \sum_{i=1}^{a} n_i, \quad \text{and} \quad \hat{\Delta} = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_e^2}. \tag{5.3.2}$$

Then our bootstrap procedure involves the following steps:

1. Generate t_i from any distribution with mean 0 and variance $n_i\hat{\Delta}(1-\hat{\alpha}_i)^2$, i.e., $t_i \overset{ind}{\sim} (0, n_i\hat{\Delta}(1-\hat{\alpha}_i)^2)$ and t_{ij} from any distribution with mean zero and variance 1,i.e., $t_{ij} \overset{ind}{\sim} (0,1)$ for $i = 1, \cdots, a$, $j = 1, \cdots, n_i$. Here t_i and t_{ij} are generated independently.

2. Compute the bootstrap data

$$y_{ij}^* = \hat{\mu} + t_i(\bar{y}_{i.} - \hat{\mu}) + t_{ij}[y_{ij} - \hat{\alpha}_i\bar{y}_{i.} - (1-\hat{\alpha}_i)\hat{\mu}].$$

3. Compute the bootstrap estimator $\bar{\hat{T}}^* = a^{-1}\sum_{i=1}^{a}\hat{T}_i^*$, where

$$\hat{T}_i^* = \sum_{j=1}^{n_i} I_{ij}^*(h), \quad I_{ij}^*(h) = \begin{cases} 1 & y_{ij}^* > h \\ 0 & \text{otherwise,} \end{cases}$$

and the bootstrap variance of $\bar{\hat{T}}^*$ using

$$Var_*(\bar{\hat{T}}^*) = (a-1)^{-1}\sum(\hat{T}_i^* - \bar{\hat{T}}^*)^2.$$

4. Obtain the proportion of individuals with no exceedances using

$$Pr_*(\bar{\hat{T}}^* = 0) = \frac{\#\ \{y_{ij}^* \leq h, i = 1, \dots, a; j = 1, \dots, n_i\}}{N}.$$

5. Independently replicate steps (1)–(3) B times, where B is large, and calculate the corresponding estimates $\bar{\hat{T}}^*_{(1)}, \dots, \bar{\hat{T}}^*_{(B)}$, $Var_{*(1)}(\bar{\hat{T}}^*), \dots,$ $Var_{*(B)}(\bar{\hat{T}}^*)$ and $Pr_{*(1)}(\bar{\hat{T}}^* = 0), \dots, Pr_{*(B)}(\bar{\hat{T}}^* = 0)$.

6. The bootstrap estimator of $E_*(\bar{\hat{T}}^*)$ can be approximated by $\bar{\hat{T}}^*_{(\cdot)} = B^{-1}\sum_{b=1}^{B}\bar{\hat{T}}^*_{(b)}$, its variance by $Var_{*(\cdot)}(\bar{\hat{T}}^*) = B^{-1}\sum_{i=1}^{B} Var_{*(b)}(\bar{\hat{T}}^*)$ and $Pr_{*(\cdot)}(\bar{\hat{T}}^* = 0)$ by $B^{-1}\sum_{b=1}^{B} Pr_{*(b)}(\bar{\hat{T}}^* = 0)$.

The histogram of the B estimates for $\bar{\hat{T}}^*_{(b)}$ for $b = 1, \dots, B$ is used to form the 100(1- 2α)% confidence interval $(\bar{\hat{T}}_L(\alpha), \bar{\hat{T}}_U(\alpha))$, where $\bar{\hat{T}}_L(\alpha) = C\hat{D}F(\alpha)$. $\bar{\hat{T}}_U(\alpha) = C\hat{D}F(1-\alpha)$ and

$$C\hat{D}F(z) = \frac{\#\{\bar{\hat{T}}^*_{(b)} \leq z; b = 1, \dots, B\}}{B}$$

Theorem 5.1. *Suppose the one-way random effects model (5.2.1)-(5.2.3) hold. Then*

1. $E_F[E_*(y_{ij}^*)] = \mu$,

2. $$E_F[Cov_*(y_{ij}^*, y_{i'j'}^*)] = \begin{cases} \sigma_v^2 + \sigma_e^2 + O(a^{-1}), & \textit{for } i = i' \textit{ and } j = j' \\ \sigma_v^2 + O(a^{-1}), & \textit{for } i = i' \textit{ and } j \neq j' \\ 0 & \textit{otherwise,} \end{cases}$$

where $Cov_*(y^*_{ij}, y^*_{i'j'}) = E_*(y^*_{ij} - \hat{\mu})(y^*_{i'j'} - \hat{\mu})$ and E_* and E_F represent expectation with respect to the distributions induced by bcotstrap sampling and the model, respectively.

Proof. Since the expected values of t_i and t_{ij} are zero, we immediately have the first result.

In the following proof, we assume that σ_v^2 and $\sigma_e^2 < \infty$ and the n_i's are fixed for $i = 1, \ldots, a$. Furthermore, we assume that there exists positive constants M and m such that $n_i < M$ for all i and $n_i > m$ for all i. In view of $\hat{\alpha}$ and $\hat{\Delta}$ being consistent estimators of α and Δ, respectively, it can be easily verified that

$$\begin{aligned} E_F[Var_*(y^*_{ij})] = n_i(1-\alpha_i)^2\Delta E_F(\bar{y}_{i.} - \hat{\mu})^2 + E_F[y_{ij} - \alpha_i\bar{y}_{i.} - (1-\alpha_i)\hat{\mu}]^2 \\ + O(a^{-1}). \end{aligned} \tag{5.3.3}$$

Now, consider the first term of right hand side of equation (5.3.3)

$$\begin{aligned} E_F(\bar{y}_{i.} - \hat{\mu})^2 &= E_F(\bar{y}_{i.} - \mu)^2 - 2E_F(\bar{y}_{i.} - \mu)(\hat{\mu} - \mu) + E_F(\hat{\mu} - \mu)^2 \\ &= \left(\sigma_v^2 + \frac{\sigma_e^2}{n_i}\right) - 2\frac{n_i}{N}\left(\sigma_v^2 + \frac{\sigma_e^2}{n_i}\right) + \frac{1}{N^2}\sum_{i=1}^{a} n_i^2\left(\sigma_v^2 - \frac{\sigma_e^2}{n_i}\right) \\ &= \left(\sigma_v^2 + \frac{\sigma_e^2}{n_i}\right) - C_1 + C_2. \end{aligned} \tag{5.3.4}$$

We will show that C_1 and C_2 are of order $O(a^{-1})$

$$\begin{aligned} C_2 &\leq \left(\frac{1}{ma}\right)^2 \left(M^2 a\sigma_v^2 + Ma\sigma_e^2\right) \\ &= \left(\frac{M}{m}\right)^2 \left(\sigma_v^2 + \frac{\sigma_e^2}{M}\right)\frac{1}{a} \\ &= O(a^{-1}). \end{aligned} \tag{5.3.5}$$

Similarly,

$$\begin{aligned} C_1 &\leq 2\left(\frac{M}{m}\right)\left(\sigma_v^2 - \frac{\sigma_e^2}{n_i}\right)\left(\frac{1}{a}\right) \\ &= O(a^{-1}), \end{aligned} \tag{5.3.6}$$

and $E_F(\bar{y}_{i.} - \hat{\mu})^2 = (\sigma_v^2 + \sigma_e^2/n_i) + O(a^{-1})$. To evaluate the second term of right hand side of equation (5.3.3), consider

$$\begin{aligned} E_F[y_{ij} - \alpha_i\bar{y}_{i.} - (1-\alpha_i)\hat{\mu}]^2 &= Var_F(y_{ij} - \alpha_i\bar{y}_{i.}) \\ &= \sigma_v^2 + \sigma_e^2 - 2\alpha_i\left(\sigma_v^2 + \frac{\sigma_e^2}{n_i}\right) + \alpha_i^2\left(\sigma_v^2 + \frac{\sigma_e^2}{n_i}\right) \\ &= \left(\sigma_e^2 - \frac{\sigma_e^2}{n_i}\right) + \left(\sigma_v^2 + \frac{\sigma_e^2}{n_i}\right)(1 + \alpha_i^2 - 2\alpha_i) \\ &= \left(\sigma_e^2 - \frac{\sigma_e^2}{n_i}\right) + \frac{\sigma_e^2}{n_i}(1 + n_i\Delta)(1-\alpha)^2 \\ &= \left(\sigma_e^2 - \frac{\sigma_e^2}{n_i}\right) + \frac{\sigma_e^2}{n_i}\left(\frac{1}{1-\alpha_i}\right)^2(1-\alpha_i)^2 \\ &= \sigma_e^2. \end{aligned} \tag{5.3.7}$$

Therefore, combining (5.3.4) and (5.3.7) we have the desired result for $i = i'$ and $j = j'$. Turning to $E_F[Cov_*(y_{ij}^*, y_{i'j'}^*)]$ for $i = i'$ and $j \neq j'$, we have

$$\begin{aligned} E_F[Cov_F(y_{ij}^*, y_{i'j'}^*] &= n_i(1-\alpha_i)^2\Delta E_F(\bar{y}_{i.} - \hat{\mu})^2 + O(a^{-1}) \\ &= n_i(1-\alpha_i)^2\Delta\left(\sigma_v^2 + \frac{\sigma_e^2}{n_i} + O(a^{-1})\right) + O(a^{-1}) \\ &= n_i(1-\alpha_i)^2\Delta\left(\frac{\sigma_e^2}{n_i(1-\alpha_i)^2} + O(a^{-1})\right) + O(a^{-1}) \\ &= \sigma_v^2 + O(a^{-1}). \end{aligned} \tag{5.3.8}$$

□

5.4 A Simulation Study

To study the relative performance of the proposed resampling method. we consider the actual population originally given by Solomon (1989). This population consists of $a = 25$, $n_i = 16$ for all i with $\mu = 91.70$. The following distributions are considered for $\{v_i\}$ and $\{e_{ij}\}$:

1. $v_i \sim N(0, 26.86)$ and $e_{ij} \sim N(0, 52.02)$,

2. $v_i \sim 0.1N(0, 26.86) + 0.9N(0, 52.02)$ and
   $e_{ij} \sim 0.1N(0, 26.86) + 0.9N(0, 52.02)$,

3. $v_i \sim$ Cauchy and $e_{ij} \sim$ Cauchy,

4. $v_i \sim$ Exp(1) and $e_{ij} \sim$ Exp(1).

We then generate 1,000 sets of $\{v_i\}$ and $\{e_{ij}\}$ according to these four distributions with number of bootstrap B=1,000.

The results are given in Tables 5.1–5.4. For the case when the random effects and random errors are both independently normally distributed, the expected value of T is 6.626 for Solomon method, 6.520 for bootstrap method and 6.624 for true value. Its variance is 22.013 for Solomon method, 22.869 for bootstrap method and 22.846 for true value. The probability of no exceedance is 0.078 for Solomon method, 0.083 for bootstrap method 0.098 for true value. It is observed that the Solomon and bootstrap methods perform well in tracking the true values of the average number of exeedances, variances and the probabilities of no exceedances; while the bootstrap performs slightly better than the Solomon method in tracking the variance. It is not surprising that the Solomon method is doing well here since the variance components are normally distributed. However, similar results do not hold for other distributions, like Cauchy and exponential, where the bootstrap method performs better than the Solomon method in tracking their respective true values.

Turning to the coverage probability for $E(T)$, the percentile method for bootstrap confidence intervals is considered and the coverage probabilities are consistently below the respective nominal rates. We try some modifications in constructing the confidence interval in a transformed scale. However, the changes had little effect on the coverage probabilities.

To summarize, the maximum likelihood yields good results provided both v_i and e_{ij} are independently normally distributed. However, the bootstrap method is a good competitor in obtaining point estimates even when v_i and e_{ij} are not independently normally distributed. The question of constructing confidence intervals needs further study and this is under investigation.

Table 5.1: The number of exceedances, its variance and probability of no exceedance for the model with $v_i \sim N(0, 26.86)$ and $e_{ij} \sim N(0, 52.02)$ for $i = 1, \ldots, 25$, $j = 1, \ldots, 16$, N=1,000 simulations and B=1,000 bootstrap samples.

Method	$E(T)$	$Var(T)$	$Pr(T = 0)$
True Value	6.624	22.846	0.098
Solomon	6.626	22.013	0.078
Bootstrap	6.520	22.869	0.083

Table 5.2: The number of exceedances, its variances and probability of no exceedance for the model with v_i and $e_{ij} \sim 0.1N(0, 26.86) + 0.9N(0, 52.02)$ for $i = 1, \ldots, 25$, $j = 1, \ldots, 16$, N=1,000 simulations and B=1,000 bootstrap samples.

Method	$E(T)$	$Var(T)$	$Pr(T = 0)$
True Value	6.566	16.627	0.046
Solomon	6.565	16.119	0.056
Bootstrap	6.402	18.399	0.053

Table 5.3: The number of exceedances, its variance and probability of no exceedance for the model with v_i and $e_{ij} \sim$ Cauchy for $i = 1, \ldots, 25$, $j = 1, \ldots, 16$. N=1,000 simulations and B=1,000 bootstrap samples.

Method	$E(T)$	$Var(T)$	$Pr(T = 0)$
True Value	3.997	18.874	0.135
Solomon	7.518	12.120	0.244
Bootstrap	4.702	19.699	0.103

Table 5.4: The number of exceedances, its variance and probability of no exceedance for v_i and $e_{ij} \sim$ Exp(1) for $i = 1, \ldots, 25$, $j = 1, \ldots, 16$, N=1,000 simulations and B=1,000 bootstrap samples.

Method	$E(T)$	$Var(T)$	$Pr(T = 0)$
True Value	1.454	8.020	0.469
Solomon	1.279	5.795	0.248
Bootstrap	1.750	10.276	0.373

Chapter 6

General discussion and topics for further research

In this dissertation, we have treated four problems, namely,

(I) Inference on the slope parameter in a logistic regression model when values of the covariate X for a subset of the study subjects may be missing but values of a surrogate variables are available.

(II) Use of empirical likelihood to obtain variance estimator of the ratio and regression estimator under two-phase sampling.

(III) Two bootstrap methods are proposed to obtain robust inference on regression parameter for measurement error model with known error variance ratio.

(IV) A weighted bootstrap procedure is suggested to estimate exceedances over a threshold under one-way random effects model.

In Chapter 2, we studied the problem of estimation in logistic regression with a surrogate covariate. We regarded the sample units in the validation set S_{n-m} as being simple random sampling from S_n, where S_n is the set containing both validation and primary sets. We then derived an asymptotic variance of the empirical likelihood estimator. This asymptotic variance has two components. The first component represents the expected information based on $L(\beta|X_1, X_2, \ldots, X_n)$, the likelihood for the observed data if $P(Y|X)$ was completely known, while the second term represents the penalty for not having observed X_i for $i \in S_{n-m}$. The simulation results suggested that the empirical likelihood should be adopted over the three existing estimators. The usefulness of the empirical likelihood method relies on the fact that it is robust for any misspecification of the model $P(X|Z)$. This problem is also known as covariate measurement error in the literature, since we can regard Z_i as observed value of X_i for $i \in S_{n-m}$ measured with error.

As a further research on this problem, we would like to extend the results of this chapter to the case where the surrogate covariate is measured more than once on the same unit.

In Chapter 3, we considered ratio estimation under two-phase simple random sampling. We proposed two new variance estimators, namely, the empirical likelihood estimator $v_4(\bar{y}_{ts})$ and the design-based estimator $v_3(\bar{y}_{ts})$. The variance estimator $v_4(\bar{y}_{ts})$ was substantially more efficient than the standard variance estimator and the ones proposed by Rao and Sitter (1995). Whereas, The variance estimator $v_3(\bar{y}_{ts})$ was more efficient than the ones proposed by Rao and Sitter only for certain case. We have restricted our considerations to the case of simple random sampling without replacement on both phases. Sometimes even if $\bar{X}$ is unknown, it may be cheaper to obtain information on all units of the population of a second auxiliary variable Z, which is highly related to X, but remotely related to Y compared to X. For discussion on this type of situation in the empirical likelihood framework, see Benhin and Prasad (1997).

In Chapter 4, we proposed two bootstrap procedures for slope parameter under a linear structural relationship with known error variance ratio. The goal was to choose a correct bootstrap technique so that the first three bootstrap moments of $\hat{\beta}^*$ equaled the first three moments of $\hat{\beta}$. In the first bootstrap, we required an assumption that the joint distribution of X and Y was symmetrically distributed, however, this assumption was not needed in the second weighted bootstrap method. The methods presented here were simple and attractive in contrast to the Linder and Babu bootstrap procedure. Perhaps most importantly, it circumvented calculating the correction factors that were required in the Linder and Babu method in estimating the bootstrap variance. Our simulation results showed that the proposed bootstrap procedures were able to track to their respective nominal levels and were robust to heteroscedasticity, in particular, the weighted bootstrap. These results are encouraging and worthwhile to investigate the case where X and Y are measured with replications.

Turning to Chapter 5, we discussed the use of bootstrap in a one-way random effects model. We developed a weighted bootstrap algorithm to mimic the variance-covariance structure as in the original model. An example on blood pressure was used to illustrate the application of this bootstrap procedure. The bootstrap estimates performed well in tracking the true respective values. However, the coverage probabilities failed to track their respective nominal rates. We have no theoretical argument to explain why the coverage probabilities performed poorly. We tried some transformations, but the changes had little effect on the coverage probabilities. It is in our intention to investigate this further.

Bibliography

Anderson, T. W., and Sawa, T. (1982). Exact and approximate distributions of the maximum likelihood estimators of a slope coefficient. *J. Royal Statist. Soc., Ser B*, **44**, 52–62.

Arvesen, J. N. (1969). Jackknifing U-statistics. *Ann. Statist.*, **40**, 2076–2100.

Babu, G. J., and Bai, Z. (1992). Edgeworth expansions for errors-in-variables models. *J. Multivariate Anal.*, **42**, 226–224.

Babu, G. J., and Singh, K. (1983). Nonparametric inference on means using bootstrap. *Ann. Statist.*, **11**, 999–1003.

Babu, G. J., and Singh, K. (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā*, **46**, 219–232.

Barnett, V. D. (1967). A note on linear structural relationships when both residual variances are known. *Biometrika*, **54**, 670–672.

Benhin, E., and Prasad, N. G. N. (1997). Empirical likelihood estimation in two-phase sampling using two auxiliary variables. Unpublished manuscript.

Bickel, P. J., and Freedman, D. (1981). Some asymptotics theory for the bootstrap. *Ann. Statist.*, **9**, 1196–1217.

Birch, M. (1964). A note on the maximum likelihood estimation of a linear structural relationship. *J. Amer. Statist. Assoc.*, **59**, 1175–1178.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge.

Box, G. E. P., and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley and Sons, New York.

Carroll, R. J., and Wand, M. P. (1991). Semiparametric estimation in logistic measurement error models. *J. R. Statist. Soc. B*, **53**, 573–585.

Chan, N. N., and Mak, T. K. (1983). Estimation of multivariate linear functional relationships. *Biometrika*, **70**, 263–267.

Chaubey, Y. P. (1984). On the comparison of some non-negative estimators of variance components. *Commun. Statist. B: Simul. and Comp.*, **13**, 619–633.

Chen, J., and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, **80**, 107–116.

Chen, S. X. (1994). Comparing empirical likelihood functions and bootstrap hypothesis tests. *J. Multivariate Anal.*, **51**, 277–293.

Cochran, W. G. (1977). *Sampling Techniques*, third edition, John Wiley and Sons, New York.

Cook, M. B. (1951). Bivariate k-statistics and cumulants of their joint sampling distribution. *Biometrika*, **38**, 179–195.

Cox, D. R., and Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

Diciccio, T., Hall, P., and Romano, J. (1989). Comparison of parametric and empirical likelihood functions. *Biometrika*, **76**, 465–476.

Donner, A. (1986). A review of inference procedures for the intraclass correlation in one-way random effects model. *Int. Statist. Review*, **54**, 67–82.

Dorfman, A. H. (1994). A note on variance estimation for the regression estimator in double sampling. *J. Am. Statist. Assoc.*, **89**, 137–140.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**, 1–26.

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

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.*, **82**, 171–200.

Elston, R. C. (1977). Estimating "heritability" of a dichotomous trait. *Biometrics*, **33**, 232–233.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equation. *J. Am. Statist. Assoc.*, **72**, 147–148.

Fuller, W. A. (1987). *Measurement Error Models*, John Wiley and Sons, New York.

Gladen, B., and Rogan, W. (1979). Misclassification and the design of environmental studies. *Am. J. Epidem.*, **109**, 607–616.

Gleser, L. J. (1981). Estimation in a multivariate "error in variables" regression model: large sample results. *Ann. Statist.*, **9**, 24–44.

Gleser, L. J. (1983). Functional, structural and ultrastructural errors-in-variable models. *Proc. Bus. Econ. Statist. Sect.*, American Statistical Association, Washington, DC, pages 57–66.

Goldstein, H. (1979). Some models for analysing longitudinal data on educational attainment (with discussion). *J. R. Statist. Soc. A*, **142**, 407–442.

Hall, P., and Scala, B. L. (1990). Methodology and algorithms of empirical likelihood. *Int. Statist. Rev.*, **58**, 109–127.

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

Hartley, H. O., and Rao, J. N. K. (1968). Covariate measurement error in generalized linear models. *Biometrika*, **55**, 547–557.

Harville, D., and Fenech, A. (1985). Confidence intervals for a variance ratio, or for heritability, in an unbalanced linear model. *Biometrics*. **41**, 137–152.

Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems. *J. Amer. Statist. Assoc.*, **72**, 320–338.

Kalton, G., and Kasprzyk, D. (1986). Imputing for missing surveys responses. *Survey Methodology*, **12**, 1–16.

Kelly, G. (1984). The influence function in the errors in variable problem. *Ann. Statist.*, **12**, 87–100.

Kendall, M. G., and Stuart, A. (1979). *The Advanced Theory of Statistics*, 4th edition, volume 2, Griffin, London.

Khuri, A. I., and Littell, R. (1987). Exact tests for the main effects variance components in an unbalanced random two-way model. *Biometrics*, **43**, 545–560.

Khuri, A. I., and Sahai, H. (1985). Variance components analysis: A selective literature survey. *Int. Statist. Review*, **14**, 1261–1350.

Kolaczyk, E. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica*, **4**, 199–218.

Linder, E., and Babu, G. J. (1994). Bootstrapping the linear functional relationship with known error variance ratio. *Scand. J. Statist.*, **21**, 21–39.

Lindley, D. V., and El-Sayyad, G. M. (1968). The Bayesian estimation of a linear functional relationship. *J. R. Statist. Soc. B*, **30**, 190–202.

Liu, R. Y. (1988). Bootstrap procedure under some non-i.i.d. models. *Ann. Statist.*, **14**, 1697–1708.

Mak, T. K., Li, W. K., and Kuk, Y. C. (1986). The use of surrogate variables in binary regression models. *J. Statist. Comput. Simul.*, **24**, 245–254.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.*, **18**, 90–120.

Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist*, **19**, 1725–1747.

Pepe, M. S., and Fleming, T. (1991). A general nonparametric method for dealing with errors missing or surrogate data. *J. Am. Statist. Assoc.*, **86**, 108–121.

Pepe, M. S., Reilly, M., and Fleming, T. R. (1994). Auxiliary outcome data and the mean score method. *J. Statist. Plan. Inference*, **42**, 137–160.

Prasad, N. G. N., and Rao, J. N. K. (1988). Robust tests and confidence intervals for error variance in a regression model and for functions of variance components in an unbalanced one-way model. *Commun. Statist. Theory Meth.*, **17**, 1111–1133.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statist. Med.*, **8**, 431–440.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1993). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edition, Cambridge University Press, New York.

Rao, C. R., and Kleffe, J. (1988). *Estimation of Variance Components and Application*, North-Holland, Amsterdam.

Rao, J. N. K., and Sitter, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, **82**, 453–460.

Rao, P. S. R. S., and Rao, J. N. K. (1971). Small sample results for ratio estimators. *Biometrika*, **58**, 625–630.

Riersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica*, **18**, 375–389.

Rosner, B., Willett, W., and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statist. Med.*, **8**, 1075–1093.

Rubin, D. B. (1987). *Multiple Imputation for Non-Response in Surveys*, John Wiley and Sons, New York.

Sahai, H., Khuri, A. I., and Kapadia, C. H. (1985). A second bibliography on variance components. *Commun. Statist. A*, **14**, 63–115.

Schafer, D. W. (1987). Covariate measurement error in generalized linear models. *Biometrika*, **74**, 385–391.

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

Seely, J. F., and El-Bassiouni, Y. (1983). Applying Wald's variance component test. *Ann. Statist.*, **11**, 197–201.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, New York.

Sitter, R. R. (1997). Variance estimation for the regression estimator in two-phase sampling. *J. Am. Statist. Assoc.*, **92**, 780–787.

Smith, C. A. B. (1980). A review of inference procedures for the intraclass correlation in one-way random effects model. *Ann. Hum. Gen.*, **21**, 363–373.

Solari, M. E. (1969). The maximum likelihood solution of the problem of estimating a linear functional relationship. *J. R. Statist. Soc. B*, **31**, 372–375.

Solomon, P. J. (1989). On components of variance and modelling exceedances over a threshold. *Austral. J. Statist.*, **31**, 18–24.

Spjøtvoll, E. (1968). Confidence intervals and tests for variance ratios in unbalanced variance components models. *Ann. Statist.*, **11**, 197–201.

Stefanski, L. A., and Carroll, R. J. (1985). Covariate measurement error in generalized linear models. *Ann. Statist.*, **13**, 1335–1351.

Sukhatme, P. V., and Sukhatme, B. V. (1970). *Sampling Theory of Surveys with Applications*, second edition, Iowa State University Press, Ames.

Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Statist. Assoc.*, **82**, 528–540.

Townsend, E. C., and Searle, S. R. (1968). Best quadratic unbiased estimation of variance components from unbalanced data in one-way classification. *Biometrics*, **27**

Wittes, J., Lakatos, E., and Probstfield, J. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statist. Med.*, **8**, 415–425.

Woodhouse, G., Yang, M., Goldstein, H., and Rasbash, J. (1996). Adjusting for measurement error in multilevel analysis. *J. R. Statist. Soc. A*, **159**, 201–212.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.*, **14**, 1261–1350.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, New York.

Appendix A

Proof of Theorem 4.1

We first obtain the bias of $\hat{h}$ by writing $\hat{h}$ in terms of bivariate k-statistics. That is,

$$\begin{aligned}\hat{h} &= \frac{1}{2}\hat{\mu}_{11}^{-1}\left(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20}\right)\\ &= \frac{1}{2}k_{11}^{-1}\left(k_{02} - \lambda^2 k_{20}\right),\end{aligned}$$

where $k_{rs} = n(n-1)^{-1}\hat{\mu}_{rs}$ for $r, s = 0, 1$ and $\lambda^2 = \sigma_\varepsilon^2/\sigma_\delta^2$. By a multivariate Taylor series expansion around $\hat{h} = \hat{h}(k_{02}, k_{20}, k_{11})$, we have

$$\hat{h} = h + \frac{1}{2}\kappa_{11}^{-1}h_1 + \frac{1}{2}\kappa_{11}^{-2}h_2 + \frac{1}{2}\kappa_{11}^{-3}h_3 + O_p(n^{-2}), \qquad (A.0.1)$$

where

$$h_1 = (k_{02} - \kappa_{02}) - \lambda^2(k_{20} - \kappa_{20}) - 2h(k_{11} - \kappa_{11}),$$

$$h_2 = -[(k_{02} - \kappa_{02})(k_{11} - \kappa_{11}) - \lambda^2(k_{20} - \kappa_{20})(k_{11} - \kappa_{11}) - 2h(k_{11} - \kappa_{11})^2],$$

$$h_3 = (k_{02} - \kappa_{02})(k_{11} - \kappa_{11})^2 - \lambda^2(k_{20} - \kappa_{20})(k_{11} - \kappa_{11})^2 - 2h(k_{11} - \kappa_{11})^3,$$

and the κ's are the respective population cumulants. Now, we have the first central moment of $\hat{h}$ given by

$$
\begin{aligned}
E_F(\hat{h} - h) = &-\frac{1}{2}\kappa_{11}^{-2}(\kappa[(02)] - \lambda^2\kappa[(20)] - 2h\kappa[(11)^2]) \\
&+ \frac{1}{2}\kappa_{11}^{-3}\{\kappa[(02)(11)] - \lambda^2\kappa[(20)(11) - 2h\kappa[(11)^2]\} + O(n^{-3}),
\end{aligned}
$$

where

$$\kappa[(\alpha\alpha')^r(\beta\beta')^s] = E_F[(k_{\alpha\alpha'} - \kappa_{\alpha\alpha'})^r(k_{\beta\beta'} - \kappa_{\beta\beta'})^s], \quad r + s \leq 3.$$

The evaluation of the terms on the right-hand side involves heavy algebra. We use the method of bivariate k-statistics and product cumulants. The relevant formulae have been tabulated by Cook (1951) . Using these formulae and writing in terms of cumulants, we have

$$E_F(\hat{h} - h) = -\frac{1}{2n}\kappa_{11}^{-2}a_{1\kappa} + \frac{1}{2n^2}\kappa_{11}^{-3}(a_{2\kappa} + a_{3\kappa} + a_{4\kappa}) + O(n^{-3}),$$

where

$$
\begin{aligned}
a_{1\kappa} &= \kappa_{13} + 2\kappa_{02}\kappa_{11} - \lambda^2(\kappa_{31} + 2\kappa_{20}\kappa_{11}) - 2h(\kappa_{22} + \kappa_{20}\kappa_{02} + \kappa_{11}^2), \\
a_{2\kappa} &= \kappa_{24} + 5\kappa_{22}\kappa_{02} + 6\kappa_{13}\kappa_{11} + \kappa_{04}\kappa_{20} + 2\kappa_{12}^2 + 2\kappa_{03}\kappa_{21} \\
&\quad + 6\kappa_{02}\kappa_{11}^2 + 2\kappa_{02}^2\kappa_{20}, \\
a_{3\kappa} &= -\lambda^2(\kappa_{42} + 5\kappa_{22}\kappa_{20} + 6\kappa_{31}\kappa_{11} + \kappa_{40}\kappa_{02} + 2\kappa_{21}^2 + 2\kappa_{30}\kappa_{12} \\
&\quad + 6\kappa_{20}\kappa_{11}^2 + 2\kappa_{20}^2\kappa_{02}), \\
a_{4\kappa} &= -2h(\kappa_{33} + 6\kappa_{22}\kappa_{11} + 3\kappa_{20}\kappa_{13} + 3\kappa_{31}\kappa_{02} + 3\kappa_{21}\kappa_{12} + \kappa_{30}\kappa_{03}, \\
&\quad + 2\kappa_{11}^3 + 6\kappa_{20}\kappa_{02}\kappa_{11}).
\end{aligned}
$$

In virtue of the following identities given by Cook (1951)

$$
\begin{aligned}
\kappa_{20} &= \mu_{20}, \\
\kappa_{11} &= \mu_{11}, \\
\kappa_{30} &= \mu_{30}, \\
\kappa_{21} &= \mu_{21}, \\
\kappa_{40} &= \mu_{40} - 3\mu_{20}^2, \\
\kappa_{31} &= \mu_{31} - 3\mu_{20}\mu_{11}, \\
\kappa_{22} &= \mu_{22} - \mu_{20}\mu_{02} - 2\mu_{11}, \\
\kappa_{50} &= \mu_{50} - 10\mu_{30}\mu_{20}, \\
\kappa_{41} &= \mu_{41} - 4\mu_{30}\mu_{11} - 6\mu_{21}\mu_{20}, \\
\kappa_{32} &= \mu_{32} - \mu_{30}\mu_{02} - 6\mu_{21}\mu_{11} - 3\mu_{20}\mu_{12}, \\
\kappa_{60} &= \mu_{60} - 15\mu_{40}\mu_{20} - 10\mu_{30}^2 + \mu_{20}^2, \\
\kappa_{51} &= \mu_{51} - 5\mu_{40}\mu_{11} - 10\mu_{31}\mu_{20} - 10\mu_{30}\mu_{21} + 30\mu_{20}^2\mu_{11}, \\
\kappa_{42} &= \mu_{42} - \mu_{40}\mu_{02} - 8\mu_{31}\mu_{11} - 4\mu_{30}\mu_{12} - 6\mu_{22}\mu_{20} - 6\mu_{21}^2 \\
&\quad + 6\mu_{20}^2\mu_{02} + 24\mu_{20}\mu_{11}^2, \\
\kappa_{33} &= \mu_{33} - 3\mu_{31}\mu_{02} - \mu_{30}\mu_{03} - 9\mu_{22}\mu_{11} - 9\mu_{21}\mu_{12} \\
&\quad - 3\mu_{20}\mu_{02} + 18\mu_{20}\mu_{11}\mu_{02} + 12\mu_{11}^3,
\end{aligned}
$$

the bias of $\hat{h}$ can be written in terms of moments as

$$
\begin{aligned}
E_F(\hat{h} - h) = -\frac{1}{2n}\mu_{11}^{-2}\left\{\mu_{13} - \lambda^2\mu_{31} - 2h\mu_{22}\right\} + \frac{1}{2n^2}\mu_{11}^{-3}(a_{2\mu} + a_{3\mu} + a_{4\mu}) \\
+ O(n^{-3}),
\end{aligned}
$$

where

$$
\begin{aligned}
a_{2\mu} &= \mu_{24} - 2\mu_{13}\mu_{11} - 2\mu_{03}\mu_{21} - \mu_{02}\mu_{22} - 4\mu_{12}^2 + 2\mu_{02}\mu_{11}^2,\\
a_{3\mu} &= -\lambda^2(\mu_{42} - 2\mu_{31}\mu_{11} - 2\mu_{30}\mu_{12} - \mu_{20}\mu_{22} - 4\mu_{21}^2 + 2\mu_{20}\mu_{11}^2),\\
a_{4\mu} &= -2h(\mu_{33} - 3\mu_{22}\mu_{11} - 6\mu_{12}\mu_{21} + 2\mu_{11}^3).
\end{aligned}
$$

Upon simplification, the bias of $\hat{h}$ is given by

$$E_F(\hat{h} - h) = -\frac{1}{2n}\mu_{11}^{-2}\left\{\mu_{13} - \lambda^2\mu_{31} - 2h\mu_{22}\right\} + O(n^{-2}).$$

Turning to the second central moment of $\hat{h}$, by squaring (A.0.1) then taking expectation with respect with F, we have

$$E_F(\hat{h} - h)^2 = \frac{1}{4n}\kappa_{11}^{-2}b_{1\kappa} - \frac{1}{2n^2}\kappa_{11}^{-3}\left(b_{2\kappa} + b_{3\kappa} + b_{4\kappa} + b_{5\kappa} + b_{6\kappa}\right) + O(n^{-3}),$$

where

$$
\begin{aligned}
b_{1\kappa} &= \kappa_{04} + 2\kappa_{02}^2 - 2\lambda^2(\kappa_{22} + 2\kappa_{11}^2) + \lambda^4(\kappa_{40} + 2\kappa_{20}^2)\\
&\quad - 4h[\kappa_{13} + \kappa_{02}\kappa_{11} - \lambda^2(\kappa_{31} + \kappa_{20}\kappa_{11})] + 4h^2(\kappa_{22} + \kappa_{02}\kappa_{20} + \kappa_{11}^2),\\
b_{2\kappa} &= \kappa_{15} + 8\kappa_{13}\kappa_{02} + 4\kappa_{04}\kappa_{11} + 4\kappa_{03}\kappa_{12} + 8\kappa_{02}^2\kappa_{11},\\
b_{3\kappa} &= -2\lambda^2(\kappa_{33} + 8\kappa_{22}\kappa_{11} + 2\kappa_{20}\kappa_{13} + 2\kappa_{31}\kappa_{02} + 4\kappa_{21}\kappa_{12} + 4\kappa_{11}^3 + 4\kappa_{20}\kappa_{02}\kappa_{11}),\\
b_{4\kappa} &= -4h[\kappa_{24} + 5\kappa_{22}\kappa_{02} + 6\kappa_{13}\kappa_{11} + \kappa_{04}\kappa_{20} + 2\kappa_{12}^2 + 2\kappa_{03}\kappa_{21} + 6\kappa_{02}\kappa_{11}^2\\
&\quad + 2\kappa_{02}^2\kappa_{20} - \lambda^2(\kappa_{42} + 5\kappa_{22}\kappa_{20} + 6\kappa_{31}\kappa_{11} + \kappa_{40}\kappa_{02} + 2\kappa_{21}^2 + 2\kappa_{30}\kappa_{12}\\
&\quad + 6\kappa_{20}\kappa_{11}^2 + 2\kappa_{20}^2\kappa_{02})],\\
b_{5\kappa} &= \lambda^4(\kappa_{51} + 8\kappa_{31}\kappa_{20} + 4\kappa_{40}\kappa_{11} + 4\kappa_{30}\kappa_{21} + 8\kappa_{20}^2\kappa_{11}),\\
b_{6\kappa} &= 4h^2(\kappa_{33} + 6\kappa_{22}\kappa_{11} + 3\kappa_{20}\kappa_{13} + 3\kappa_{02}\kappa_{31} + 3\kappa_{21}\kappa_{12} + \kappa_{30}\kappa_{03}\\
&\quad + 2\kappa_{11}^3 + 6\kappa_{20}\kappa_{02}\kappa_{11}),
\end{aligned}
$$

or in terms of moments, we have

$$E_F(\hat{h}-h)^2 = \frac{1}{4n}\mu_{11}^{-2}b_{1\mu} + O(n^{-2}).$$

where $b_{1\mu} = \mu_{04} + \lambda^4\mu_{40} + 2\mu_{22}(2h^2-\lambda^2) - 4h(\mu_{13}-\lambda^2\mu_{31})$. In a similar fashion. the third central moment of $\hat{h}$ is

$$E_F(\hat{h}-h)^3 = \frac{1}{8n^2}\kappa_{11}^{-3}\left(c_{1\kappa}+c_{2\kappa}+c_{3\kappa}+c_{4\kappa}+c_{5\kappa}\right) + O(n^{-3}),$$

where

$$\begin{aligned}
c_{1\kappa} &= \kappa_{06} - \lambda^6\kappa_{60} + 12(\kappa_{04}\kappa_{02} - \lambda^6\kappa_{40}\kappa_{20}) + 4(\kappa_{03}^2 - \lambda^6\kappa_{30}^2) \\
&\quad + 8(\kappa_{02}^3 - \lambda^6\kappa_{20}^3), \\
c_{2\kappa} &= -3\lambda^2[(\kappa_{24} - \lambda^2\kappa_{42}) + 4\kappa_{22}(\kappa_{02} - \lambda^2\kappa_{20}) + 8\kappa_{11}(\kappa_{13} - \lambda^2\kappa_{31}) \\
&\quad + 4(\kappa_{12}^2 - \lambda^2\kappa_{21}^2) + 8\kappa_{11}^2(\kappa_{02} - \lambda^2\kappa_{20}), \\
c_{3\kappa} &= -6h[\kappa_{15} + \lambda^4\kappa_{51} + 8(\kappa_{13}\kappa_{02} + \lambda^4\kappa_{31}\kappa_{20}) + 4\kappa_{11}(\kappa_{04} + \lambda^4\kappa_{40}) \\
&\quad + 4(\kappa_{03}\kappa_{12} + \lambda^4\kappa_{30}\kappa_{21}) + 8\kappa_{11}(\kappa_{02}^2 + \lambda^4\kappa_{20}^2) - 2\lambda^2(\kappa_{33} + 8\kappa_{22}\kappa_{11} \\
&\quad + 2\kappa_{20}\kappa_{13} + 2\kappa_{31}\kappa_{02} + 4\kappa_{21}\kappa_{12} + 4\kappa_{11}^3 + 4\kappa_{02}\kappa_{20}\kappa_{11})], \\
c_{4\kappa} &= 12h^2[\kappa_{24} - \lambda^2\kappa_{42} + 5\kappa_{22}(\kappa_{02} - \lambda^2\kappa_{20}) + 6\kappa_{11}(\kappa_{13} - \lambda^2\kappa_{31}) \\
&\quad + \kappa_{04}\kappa_{20} - \lambda^2\kappa_{40}\kappa_{02} + 2(\kappa_{12}^2 - \lambda^2\kappa_{21}^2) + 2(\kappa_{03}\kappa_{21} - \lambda^2\kappa_{30}\kappa_{12})] \\
&\quad + 6\kappa_{11}^2(\kappa_{02} - \lambda^2\kappa_{20}) + 2\kappa_{02}\kappa_{20}(\kappa_{02} - \lambda^2\kappa_{20}), \\
c_{5\kappa} &= -8h^3[\kappa_{33} + 6\kappa_{22}\kappa_{11} + 3\kappa_{20}\kappa_{13} + 3\kappa_{02}\kappa_{31} + 3\kappa_{12}\kappa_{21} + \kappa_{03}\kappa_{30} \\
&\quad + 2\kappa_{11}^3 + 6\kappa_{02}\kappa_{20}\kappa_{11}].
\end{aligned}$$

Hence, we have

$$E_F(\hat{h}-h)^3 = \frac{1}{8n^2}\mu_{11}^{-3}\left(c_{1\mu}+c_{2\mu}+c_{3\mu}+c_{4\mu}+c_{5\mu}\right) + O(n^{-3}),$$

where

$$
\begin{aligned}
c_{1\mu} &= \mu_{06} - 3\mu_{04}\mu_{02} - 6\mu_{03}^2 + 2\mu_{02}^3 - \lambda^6(\mu_{60} - 3\mu_{40}\mu_{20} - 6\mu_{30}^2 + 2\mu_{20}^3), \\
c_{2\mu} &= -3\lambda^2[\mu_{24} - \mu_{04}\mu_{20} - 4\mu_{03}\mu_{21} - 2\mu_{22}\mu_{02} - 2\mu_{12}^2 + 2\mu_{02}^2\mu_{20} \\
&\quad - \lambda^2(\mu_{42} - \mu_{40}\mu_{02} - 4\mu_{30}\mu_{12} - 2\mu_{22}\mu_{20} - 2\mu_{21}^2 + 2\mu_{20}^2\mu_{02})], \\
c_{3\mu} &= -6h[\mu_{15} - \mu_{04}\mu_{11} - 2\mu_{13}\mu_{02} - 6\mu_{03}\mu_{12} + 2\mu_{02}^2\mu_{11} \\
&\quad + \lambda^4(\mu_{51} - \mu_{40}\mu_{11} - 2\mu_{31}\mu_{20} - 6\mu_{30}\mu_{21} + 2\mu_{20}^2\mu_{11}) \\
&\quad - 2\lambda^2(\mu_{33} - \mu_{31}\mu_{02} - \mu_{03}\mu_{30} - \mu_{22}\mu_{11} - 5\mu_{12}\mu_{21} - \mu_{20}\mu_{13} + 2\mu_{02}\mu_{20}\mu_{11})], \\
c_{4\mu} &= 12h^2[\mu_{24} - 2\mu_{13}\mu_{11} - 2\mu_{03}\mu_{21} - \mu_{22}\mu_{02} - 4\mu_{12}^2 + 2\mu_{02}\mu_{11}^2 \\
&\quad - \lambda^2(\mu_{42} - 2\mu_{31}\mu_{11} - 2\mu_{30}\mu_{12} - \mu_{22}\mu_{20} - 4\mu_{21}^2 + 2\mu_{20}\mu_{11}^2)], \\
c_{5\mu} &= -8h^3[\mu_{33} - 3\mu_{22}\mu_{11} - 6\mu_{12}\mu_{21} + 2\mu_{11}^3].
\end{aligned}
$$

Upon simplification, the third central moment of $\hat{h}$ is

$$
E_F(\hat{h} - h)^3 = \frac{1}{8n^2}\mu_{11}^{-3}\left(c'_{1\mu} + c'_{2\mu} + c'_{3\mu} + c'_{4\mu} + c'_{5\mu}\right) + O(n^{-3}), \qquad (A.0.2)
$$

where

$$
\begin{aligned}
c'_{1\mu} &= \mu_{06} - 6\mu_{03}^2 - \lambda^6(\mu_{60} - 6\mu_{30}^2), \\
c'_{2\mu} &= -3\lambda^2[\mu_{24} - 4\mu_{03}\mu_{21} - 2\mu_{12}^2 - \lambda^2(\mu_{42} - 4\mu_{30}\mu_{12} - 2\mu_{21}^2)], \\
c'_{3\mu} &= -6h[\mu_{15} - 6\mu_{03}\mu_{12} + \lambda^4(\mu_{51} - 6\mu_{30}\mu_{21}) \\
&\quad - 2\lambda^2(\mu_{33} - \mu_{03}\mu_{30} - 5\mu_{12}\mu_{21})], \\
c'_{4\mu} &= 12h^2[\mu_{24} - 2\mu_{03}\mu_{21} - 4\mu_{12}^2 - \lambda^2(\mu_{42} - 2\mu_{30}\mu_{12} - 4\mu_{21}^2)], \\
c'_{5\mu} &= -8h^3[\mu_{33} - 6\mu_{12}\mu_{21}].
\end{aligned}
$$

If the joint distribution of X and Y is symmetrically distributed, we have $\mu_{30} = \mu_{03} = \mu_{12} = \mu_{21} = 0$ which in turn results

$$\begin{aligned}
c'_{1\mu} &= \mu_{06} - 6\mu_{03}^2 - \lambda^6\mu_{60}, \\
c'_{2\mu} &= -3\lambda^2[\mu_{24} - \lambda^2\mu_{42}], \\
c'_{3\mu} &= -6h[\mu_{15} + \lambda^4(\mu_{51} - 2\lambda^2\mu_{33}], \\
c'_{4\mu} &= 12h^2[\mu_{24} - \lambda^2\mu_{42}], \\
c'_{5\mu} &= -8h^3\mu_{33}.
\end{aligned}$$

Substitute these values into (A.0.2), we get

$$\begin{aligned}
E_F(\hat{h} - h)^3 = \frac{1}{8n^2}\mu_{11}^3 \{&\mu_{06} - \lambda^6\mu_{60} - 3\lambda^2(\mu_{24} - \lambda^2\mu_{42}) - 6h[\mu_{15} + \lambda^4\mu_{51} \\
&-2\lambda^2\mu_{33}] + 12h^2(\mu_{24} - \lambda^2\mu_{42}) - 8h^3\mu_{33}\} + O(n^{-3}).
\end{aligned}$$

Hence, the proof of Theorem 4.1 follows by noting that $\hat{\beta}$ is a continuous function of $\hat{h}$ and

$$\frac{\partial\hat{\beta}}{\partial\hat{h}} = 1 + \frac{\hat{h}}{\sqrt{\hat{h}^2 + \lambda^2}} = \frac{\hat{\beta}}{\sqrt{\hat{h}^2 + \lambda^2}}.$$

Appendix B

Proof of Theorem 4.4

Define

$$\hat{\Delta}_i^* = (X_i^* - \bar{X})(Y_i^* - \bar{Y}) - \hat{\mu}_{11} \text{ and note that } E_*(n^{-1}\sum_{i=1}^{n}\hat{\Delta}_i^*) = 0.$$

Then, $\hat{h}^*$ can be re-written as

$$\begin{aligned}\hat{h}^* &= \hat{h} + \frac{1}{2n}\hat{\mu}_{11}^*\hat{\mu}_{11}^{-2}\sum_{i=1}^{n}\hat{A}_i^* \\ &= \hat{h} + \frac{1}{2n}\hat{\mu}_{11}^*\left(\sum_{i=1}^{n}\hat{A}_i^* + \frac{1}{n\hat{\mu}_{11}}\sum_{i=1}^{n}\hat{A}_i^*\sum_{i=1}^{n}\hat{\Delta}_i^*\right). \qquad \text{(B.0.1)}\end{aligned}$$

Taking expectation of (B.0.1) with respect to the distribution induced by bootstrap sampling described in Section 4.3.1 we have $\sum_{i=1}^{n} E_*\hat{A}_i^* = 0$. The bias of $\hat{h}^*$ is given by

$$E_*(\hat{h}^* - \hat{h}) = \frac{1}{2n}\hat{\mu}_{11}^{-2}\left(\hat{\mu}_{13} - \lambda^2\hat{\mu}_{31} - 2\hat{h}\hat{\mu}_{22}\right).$$

Hence,

$$E_*(\hat{\beta}^* - \hat{\beta}) = -\frac{\hat{\beta}}{2n\hat{\mu}_{11}^2\sqrt{\hat{h}^2 + \lambda^2}}\left\{\hat{\mu}_{13} - \lambda^2\hat{\mu}_{31} - 2h\hat{\mu}_{22}\right\} + O_p(n^{-2}),$$

which yields part (i) of Theorem 4.4 with the remark that $\hat{\beta}$, $\hat{\mu}_{11}$, $\hat{\mu}_{13}$, $\hat{\mu}_{31}$ and $\hat{\mu}_{22}$ are $\sqrt{n}$ consistent estimates of β, μ_{11}, μ_{13}, μ_{31} and μ_{22}, respectively. To establish part (ii) of Theorem 4.3 consider the second moment of $\hat{h}^*$ about $\hat{h}$.

$$\begin{aligned}
E_*(\hat{h}^* - \hat{h})^2 &= E_*\left\{\frac{1}{4n^2}\hat{\mu}_{11}^{-2}\left(\sum_i^n \hat{A}_i^*\right)^2\left[1 + 2\hat{\mu}_{11}^{-1}\frac{\sum_i^n \hat{\Delta}_i^*}{n} + \hat{\mu}_{11}^{-2}\frac{(\sum_i^n \hat{\Delta}_i^*)^2}{n^2}\right]\right\} \\
&= E_*\left\{\frac{1}{4n^2}\hat{\mu}_{11}^{-2}\left(\sum_i^n \hat{A}_i^{*2} + \sum_{i\neq j}^n \hat{A}_i^*\hat{A}_j^*\right) + \frac{1}{2n^3}\hat{\mu}_{11}^{-3}\left(\sum_{i\neq j}^n \hat{A}_i^{*2}\hat{\Delta}_j\right.\right. \\
&\quad \left. + \sum_{i\neq j,k}^n \hat{A}_i^*\hat{A}_j^*\hat{\Delta}_k\right) + \frac{1}{4n^4}\hat{\mu}_{11}^{-4}\left(\sum_{i,j}^n \hat{A}_i^{*2}\hat{\Delta}_j^2 + \sum_{i\neq j,k}^n \hat{A}_i^*\hat{A}_j^*\hat{\Delta}_k^2\right. \\
&\quad \left.\left. + \sum_{i,j,k}^n \hat{A}_i^{*2}\hat{\Delta}_j\hat{\Delta}_k + \sum_{i\neq j,k\neq l}^n \hat{A}_i^*\hat{A}_j^*\hat{\Delta}_k\hat{\Delta}_l\right)\right\} \\
&= \frac{1}{4n}\hat{\mu}_{11}^{-2}\{\hat{\mu}_{04} + \lambda^4\hat{\mu}_{40} + 2\hat{\mu}_{22}(2\hat{h} - \lambda^2) - 4\hat{h}(\hat{\mu}_{13} - \lambda^2\hat{\mu}_{31})\} \\
&\quad + O_p(n^{-2}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
E_F[E_*(\hat{h}^* - \hat{h})^2] &= \frac{1}{4n}\mu_{11}^{-2}\left\{\mu_{04} + \lambda^4\mu_{40} + 2\mu_{22}(2\hat{h} - \lambda^2) - 4\hat{h}(\mu_{13} - \lambda^2\mu_{31})\right\} \\
&\quad + O(n^{-2}).
\end{aligned}$$

Now, the required result follows along the lines of the argument given in the proof of part (i). As for the third moment of $\hat{h}^*$ about $\hat{h}$, we have

$$\begin{aligned}
E_*(\hat{h}^* - \hat{h})^3 = \frac{1}{8n^2}\hat{\mu}_{11}^{-3}\{&\hat{\mu}_{06} - 3\lambda^2\hat{\mu}_{24} + 3\lambda^4\hat{\mu}_{42} - \lambda^6\hat{\mu}_{60} \\
&-6\hat{h}(\hat{\mu}_{15} - 2\lambda^2\hat{\mu}_{33} + \lambda^4\hat{\mu}_{51}) \\
&+12\hat{h}^2(\hat{\mu}_{24} - \lambda^2\hat{\mu}_{42}) - 8\hat{h}^3\hat{\mu}_{33}\Big\} + O_p(n^{-3}).
\end{aligned}$$

Hence,

$$\begin{aligned} E_F[E_*(\hat{h}^* - \hat{h})^3] = \frac{1}{8n^2}\mu_{11}^{-3}\{&\mu_{06} - 3\lambda^2\mu_{24} + 3\lambda^4\mu_{42} - \lambda^6\mu_{60} \\ &-6h(\mu_{15} - 2\lambda^2\mu_{33} + \lambda^4 51) \\ &+12h^2(\mu_{24} - \lambda^2\mu_{42}) - 8h^3\mu_{33}\} + O(n^{-3}). \end{aligned}$$

The rest of the proof follows using earlier arguments.

Appendix C

Proof of Theorem 4.6

Let

$$\hat{h}^* = \frac{\hat{\mu}_{02}^* - \lambda^2 \hat{\mu}_{20}^*}{2\hat{\mu}_{11}^*},$$

where $\hat{\mu}_{rs}^* = n^{-1} \sum_{i=1}^n \hat{\mu}_{rs}^{*(i)}$, $r, s = 0, 1$ with

$$\begin{aligned}
\hat{\mu}_{02}^{*(i)} &= \hat{\mu}_{02} + t_i[(Y_i - \bar{Y})^2 - \hat{\mu}_{02}], \\
\hat{\mu}_{20}^{*(i)} &= \hat{\mu}_{20} + t_i[(X_i - \bar{X})^2 - \hat{\mu}_{20}], \\
\hat{\mu}_{11}^{*(i)} &= \hat{\mu}_{11} + t_i[(X_i - \bar{X})(Y_i - \bar{Y}) - \hat{\mu}_{11}],
\end{aligned}$$

and $t_1, \ldots, t_n$ being i.i.d. random variables chosen completely independent of data X_i's and Y_i's such that $Et_i = 0, Et_i^2 = Et_i^3 = 1$. Consider a multivariate Taylor expansion of $\hat{h}^*$ around $\hat{h} = \hat{h}(\hat{\mu}_{20}, \hat{\mu}_{02}, \hat{\mu}_{11})$, we have

$$\hat{h}^* = \hat{h} + \frac{1}{2}\hat{\mu}_{11}^{-1}\hat{h}_1^* + \frac{1}{2}\hat{\mu}_{11}^{-2}\hat{h}_2^* + \frac{1}{2}\hat{\mu}_{11}^{-3}\hat{h}_3^* + O_p(n^{-2})$$

or equivalently,

$$\hat{h}^* - \hat{h} = \frac{1}{2}\hat{\mu}_{11}^{-1}\hat{h}_1^* + \frac{1}{2}\hat{\mu}_{11}^{-2}\hat{h}_2^* + \frac{1}{2}\hat{\mu}_{11}^{-3}\hat{h}_3^* + O_p(n^{-2}), \qquad \text{(C.0.1)}$$

where

$$\begin{aligned}
\hat{h}_1^* &= (\hat{\mu}_{02}^* - \hat{\mu}_{02}) - \lambda^2(\hat{\mu}_{20}^* - \hat{\mu}_{20}) - 2\hat{h}(\hat{\mu}_{11}^* - \hat{\mu}_{11}),\\
\hat{h}_2^* &= -[(\hat{\mu}_{02}^* - \hat{\mu}_{02})(\hat{\mu}_{11}^* - \hat{\mu}_{11}) - \lambda^2(\hat{\mu}_{20}^* - \hat{\mu}_{20})(\hat{\mu}_{11}^* - \hat{\mu}_{11}) - 2\hat{h}(\hat{\mu}_{11}^* - \hat{\mu}_{11})^2],\\
\hat{h}_3^* &= (\hat{\mu}_{02}^* - \hat{\mu}_{02})(\hat{\mu}_{11}^* - \hat{\mu}_{11})^2 - \lambda^2(\hat{\mu}_{20}^* - \hat{\mu}_{20})(\hat{\mu}_{11}^* - \hat{\mu}_{11})^2 - 2\hat{h}(\hat{\mu}_{11}^* - \hat{\mu}_{11})^3.
\end{aligned}$$

Note that the following can be easily verified

$$\begin{aligned}
E_t[(\hat{\mu}_{02}^* - \hat{\mu}_{02})] &= 0,\\
E_t[(\hat{\mu}_{20}^* - \hat{\mu}_{20})] &= 0,\\
E_t[(\hat{\mu}_{11}^* - \hat{\mu}_{11})] &= 0,\\
E_t[(\hat{\mu}_{20}^* - \hat{\mu}_{20})^2] &= n^{-1}(\hat{\mu}_{40} - \hat{\mu}_{20}^2),\\
E_t[(\hat{\mu}_{02}^* - \hat{\mu}_{02})^2] &= n^{-1}(\hat{\mu}_{04} - \hat{\mu}_{02}^2),\\
E_t[(\hat{\mu}_{11}^* - \hat{\mu}_{11})^2] &= n^{-1}(\hat{\mu}_{22} - \hat{\mu}_{11}^2),\\
E_t[(\hat{\mu}_{02}^* - \hat{\mu}_{02})^3] &= n^{-2}(\hat{\mu}_{06} - 3\hat{\mu}_{04}\hat{\mu}_{11} + 2\hat{\mu}_{02}^3),\\
E_t[(\hat{\mu}_{20}^* - \hat{\mu}_{20})^3] &= n^{-2}(\hat{\mu}_{60} - 3\hat{\mu}_{40}\hat{\mu}_{11} + 2\hat{\mu}_{20}^3),\\
E_t[(\hat{\mu}_{11}^* - \hat{\mu}_{11})^3] &= n^{-2}(\hat{\mu}_{33} - 3\hat{\mu}_{22}\hat{\mu}_{11} + 2\hat{\mu}_{11}^3),\\
E_t[(\hat{\mu}_{02}^* - \hat{\mu}_{02})(\hat{\mu}_{20}^* - \hat{\mu}_{20})] &= n^{-1}(\hat{\mu}_{22} - \hat{\mu}_{02}\hat{\mu}_{20}),\\
E_t[(\hat{\mu}_{02}^* - \hat{\mu}_{02})(\hat{\mu}_{11}^* - \hat{\mu}_{11})] &= n^{-1}(\hat{\mu}_{13} - \hat{\mu}_{02}\hat{\mu}_{11}),\\
E_t[(\hat{\mu}_{20}^* - \hat{\mu}_{20})(\hat{\mu}_{11}^* - \hat{\mu}_{11})] &= n^{-1}(\hat{\mu}_{31} - \hat{\mu}_{20}\hat{\mu}_{11}),\\
E_t[(\hat{\mu}_{02}^* - \hat{\mu}_{02})^2(\hat{\mu}_{02}^* - \hat{\mu}_{02})] &= n^{-2}(\hat{\mu}_{24} - \hat{\mu}_{04}\hat{\mu}_{02} - 2\hat{\mu}_{22}\hat{\mu}_{02} + 2\hat{\mu}_{02}^2\hat{\mu}_{20}),\\
E_t[(\hat{\mu}_{02}^* - \hat{\mu}_{02})(\hat{\mu}_{20}^* - \hat{\mu}_{20})^2] &= n^{-2}(\hat{\mu}_{42} - \hat{\mu}_{40}\hat{\mu}_{20} - 2\hat{\mu}_{22}\hat{\mu}_{20} + 2\hat{\mu}_{20}^2\hat{\mu}_{02}),\\
E_t[(\hat{\mu}_{02}^* - \hat{\mu}_{02})^2(\hat{\mu}_{11}^* - \hat{\mu}_{11})] &= n^{-2}(\hat{\mu}_{15} - \hat{\mu}_{04}\hat{\mu}_{11} - 2\hat{\mu}_{13}\hat{\mu}_{02} + 2\hat{\mu}_{02}^2\hat{\mu}_{11}),
\end{aligned}$$

$$E_t[(\hat{\mu}^*_{20} - \hat{\mu}_{20})^2(\hat{\mu}^*_{11} - \hat{\mu}_{11})] = n^{-2}(\hat{\mu}_{51} - \hat{\mu}_{40}\hat{\mu}_{11} - 2\hat{\mu}_{31}\hat{\mu}_{20} + 2\hat{\mu}^2_{20}\hat{\mu}_{11}),$$

$$E_t[(\hat{\mu}^*_{02} - \hat{\mu}_{02})(\hat{\mu}^*_{11} - \hat{\mu}_{11})^2] = n^{-2}(\hat{\mu}_{24} - \hat{\mu}_{22}\hat{\mu}_{02} - 2\hat{\mu}_{31}\hat{\mu}_{11} + 2\hat{\mu}^2_{02}\hat{\mu}_{11}),$$

$$E_t[(\hat{\mu}^*_{20} - \hat{\mu}_{20})(\hat{\mu}^*_{11} - \hat{\mu}_{11})^2] = n^{-2}(\hat{\mu}_{42} - \hat{\mu}_{22}\hat{\mu}_{20} - 2\hat{\mu}_{13}\hat{\mu}_{11} + 2\hat{\mu}^2_{20}\hat{\mu}_{11}),$$

$$\begin{aligned} E_t[(\hat{\mu}^*_{02} - \hat{\mu}_{02})(\hat{\mu}^*_{20} - \hat{\mu}_{20})(\hat{\mu}^*_{11} - \hat{\mu}_{11})] = n^{-2}(&\hat{\mu}_{33} - \hat{\mu}_{22}\hat{\mu}_{11} - \hat{\mu}_{13}\hat{\mu}_{20} - \hat{\mu}_{31}\hat{\mu}_{02} \\ &+ 2\hat{\mu}_{02}\hat{\mu}_{20}\hat{\mu}_{11}). \end{aligned}$$

The bias of $\hat{h}^*$ under the weighted bootstrap procedure is obtained by taking expectation on both sides of (C.0.1)

$$E_t(\hat{h}^* - \hat{h}) = -\frac{1}{2n}\hat{\mu}_{11}^{-2}\left\{\hat{\mu}_{13} - \lambda^2\hat{\mu}_{31} - 2\hat{h}\hat{\mu}_{22}\right\} + O_p(n^{-2}).$$

As for the second central moment for $\hat{h}^*$, it is given by

$$E_t(\hat{h}^* - \hat{h})^2 = \frac{1}{4n}\hat{\mu}_{11}^{-2}\left\{\hat{\mu}_{04} + \lambda^4\hat{\mu}_{40} + 2\hat{\mu}_{22}(2h^2 - \lambda^2) - 4h(\hat{\mu}_{13} - \lambda^2\hat{\mu}_{31}\right\} + O_p(n^{-2})$$

The third central moment of $\hat{h}^*$ is given by

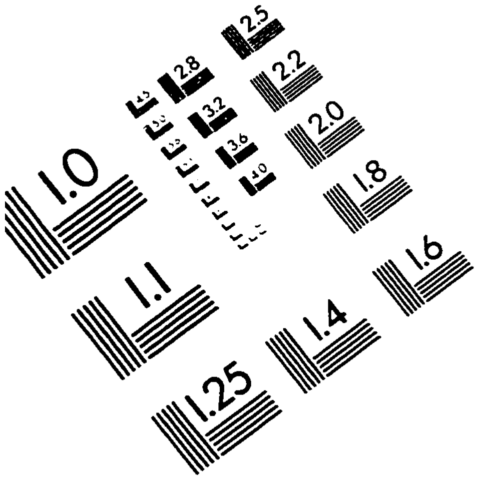
$$E_t(\hat{h}^* - \hat{h})^3 = \frac{1}{8n^2}\hat{\mu}_{11}^{-3}\left(\hat{c}''_{1\mu} + \hat{c}''_{2\mu} + \hat{c}''_{3\mu} + \hat{c}''_{4\mu} + \hat{c}''_{5\mu}\right) + O_p(n^{-3}).$$

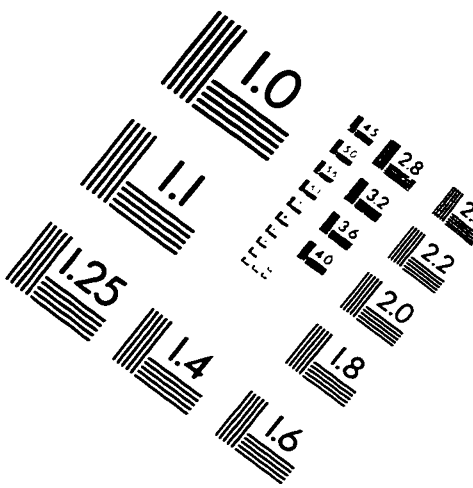
where

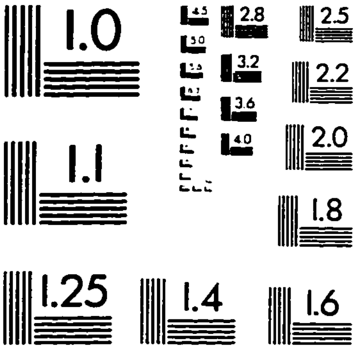
$$\begin{aligned} \hat{c}''_{1\mu} &= \hat{\mu}_{06} - 6\hat{\mu}^2_{03} - \lambda^6(\hat{\mu}_{60} - 6\hat{\mu}^2_{30}), \\ \hat{c}''_{2\mu} &= -3\lambda^2[\mu_{24} - 4\hat{\mu}_{03}\hat{\mu}_{21} - 2\hat{\mu}^2_{12} - \lambda^2(\mu_{42} - 4\hat{\mu}_{30}\hat{\mu}_{12} - 2\hat{\mu}^2_{21})], \\ \hat{c}''_{3\mu} &= -6h[\hat{\mu}_{15} - 6\hat{\mu}_{03}\hat{\mu}_{12} + \lambda^4(\hat{\mu}_{51} - 6\hat{\mu}_{30}\hat{\mu}_{21}) \\ &\qquad - 2\lambda^2(\hat{\mu}_{33} - \hat{\mu}_{03}\hat{\mu}_{30} - 5\hat{\mu}_{12}\hat{\mu}_{21})], \\ \hat{c}''_{4\mu} &= 12h^2[\hat{\mu}_{24} - 2\hat{\mu}_{03}\hat{\mu}_{21} - 4\hat{\mu}^2_{12} - \lambda^2(\hat{\mu}_{42} - 2\hat{\mu}_{30}\hat{\mu}_{12} - 4\hat{\mu}^2_{21})], \\ \hat{c}''_{5\mu} &= -8h^3[\hat{\mu}_{33} - 6\hat{\mu}_{12}\hat{\mu}_{21}]. \end{aligned}$$

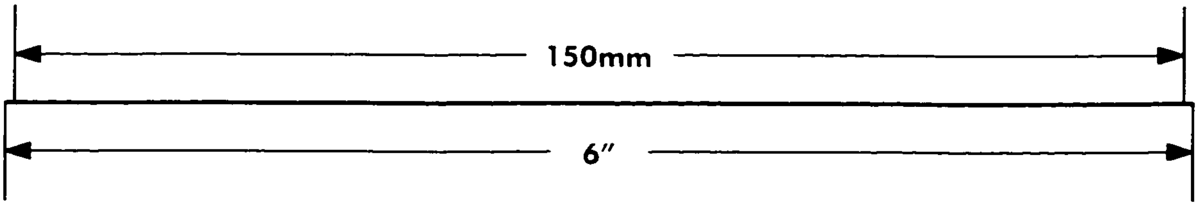
The rest of the proof follows along the lines as in Appendix B.

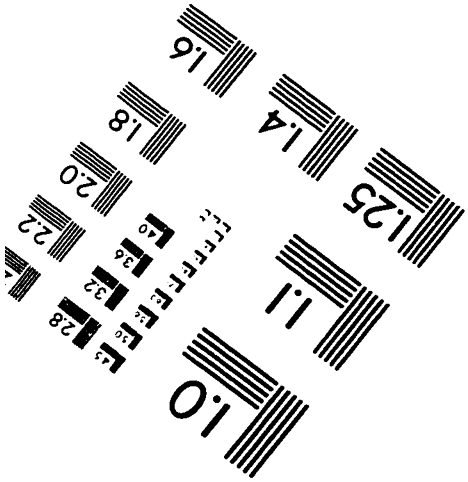
IMAGE EVALUATION TEST TARGET (QA-3)

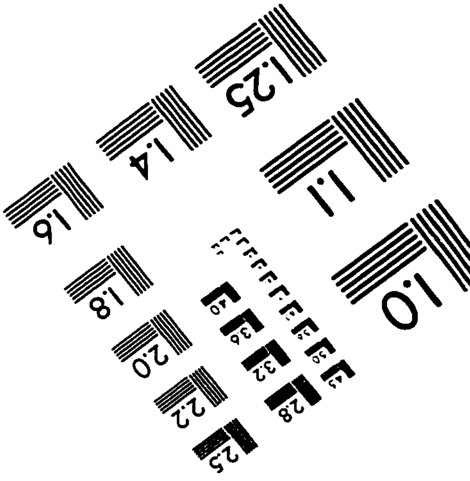












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989