INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Empirical Likelihood Inference under Stratified Sampling

by

Bob(Chongxin) Zhong

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University
Ottawa, Ontario, Canada
December 1997

© Copyright 1997
Bob(Chongxin) Zhong

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-26890-X

Canada

The undersigned hereby recommend to the Faculty of
Graduate Studies and Research acceptance of the thesis

Empirical Likelihood Inference under Stratified Sampling

Submitted by

Bob(Chongxin) Zhong, B.Sc., M.Sc.

in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

Kenneth S. Williams

Chair, Department of Mathematics and Statistics

J.N.K. Rao

Dr. J.N.K. Rao, Thesis Supervisor

John Chan

External Examiner

Carleton University

December 1997

Abstract

This dissertation consists of seven chapters. The first chapter is an introduction to empirical likelihood inference, the last chapter gives conclusions and further research areas, while the remaining five chapters deal with specific topics related to empirical likelihood method. Stratified random sampling is discussed in chapters 2 to 5.

In chapter 2, empirical likelihood under stratified random sampling is discussed for finite population inference problems. Using population auxiliary information, such as overall total, this method is shown to lead to more efficient estimators. Some limited simulation results are also given. Empirical likelihood ratio tests are also discussed.

In chapter 3, empirical likelihood and stratified sampling are combined with estimating equations to make non-parametric inferences. The results show that the empirical likelihood for parameters of interest have properties similar to those under a parametric likelihood. These results cover empirical likelihood ratio estimation and testing. Parameters subject to constraints are also studied. Illustrations of the methods are provided.

In chapter 4, a pseudo empirical likelihood method is developed in order to handle complex survey sampling designs. Asymptotic properties of these estimators are also discussed. Pseudo empirical likelihood ratio testing is also included. Illustrations of

the methods are provided.

In chapter 5, we apply the empirical likelihood technique to propose a new class of M-estimators in the presence of auxiliary information under a nonparametric setting, using stratified random sampling. It is shown that the proposed M-estimators are consistent and asymptotically normally distributed with smaller asymptotic variances than those of the usual M-estimators.

In chapter 6, we consider the following scenario: several different imperfect instruments and one perfect instrument are used independently to measure a characteristic of interest of a target population. We wish to combine the information from these independent samples to make statistical inference on parameters of interest, such as the population mean and population distribution function. We develop empirical likelihood estimators and empirical likelihood ratio tests and study their asymptotic properties.

Finally, some suggestions for further research are presented in chapter 7.

Acknowledgements

First of all, I would like to express my sincere and deeply grateful thanks to my thesis supervisor, Professor J. N. K. Rao for his supervision and encouragement during the course of this work, as well as for his kindness and support during the past four years.

My thanks also go to Prof. Shisong Mao, my M.sc. supervisor, and to Prof. J. Shao. Their support helped to make it possible for me to come to Carleton University.

My sincere thanks are also extended to Professor M. Csörgo, for his helpful discussions and suggestions.

This research was partially supported by Chand and Ratna Devi Marwah Memorial Scholarship and NSERC research grants of Professor J. N. K. Rao.

Finally, special thanks go to my wife Ci Jiang for her persistent support and to my parents for their understanding.

Contents

Acceptance Sheet	ii
Abstract	iii
Acknowledgements	v
Contents	vi
List of Tables	viii
1 Introduction	1
1.1 Empirical likelihood and asymptotic setup	2
1.1.1 Empirical likelihood	2
1.1.2 Asymptotic setup	3
1.2 Outline of Chapters 2 to 4	4
1.2.1 Introduction to Chapter 2	4
1.2.2 EL and estimating equations	7
1.2.3 Pseudo EL and estimating equations	9
1.3 EL and M-estimation	12

1.4	EL in the Presence of Measurement Error	14
2	EL for Finite Populations	17
2.1	Introduction	17
2.2	Maximum Empirical Likelihood Estimator	19
2.2.1	Numerical Evaluation for MELE	21
2.2.2	Asymptotic Results	25
2.2.3	Variance Estimation	27
2.3	Estimation of Small Area Means	28
2.4	Empirical Likelihood Ratio Test	30
2.5	Simulation study	34
2.6	Proofs	41
3	EL and GEE	55
3.1	Introduction	55
3.2	MELR Estimators and Their Properties	58
3.2.1	MELR estimators	58
3.2.2	Numerical method for given θ	60
3.2.3	Asymptotic Results	61
3.3	MELR Estimators and Testing with Constraints	64
3.4	Examples	67
3.5	Proofs	70
4	Pseudo EL and GEE	90
4.1	Introduction	90

4.2	MPELR Estimators	91
4.3	Asymptotic Results	95
4.4	An example	101
4.5	Proofs	103
5	M-estimation and EL	114
5.1	Introduction	114
5.2	Profile empirical likelihood	115
5.3	M-estimators	118
5.4	Proofs	121
6	EL in the presence of measurement error	127
6.1	Introduction	127
6.2	Empirical likelihood method in the presence of measurement error .	129
6.3	Asymptotic properties of MELRE	132
6.4	Likelihood Ratio Tests	134
6.5	Examples	135
6.6	Proofs	141
7	Conclusions and Further Research	150
7.1	Conclusions	150
7.2	Further Research	152
	References	154

List of Tables

2.1	Parameter settings for generated finite populations	37
2.2	Ratios of MSE 's of the $\hat{Y}$ to $\hat{Y}_{GREG}$, $\bar{y}_{st}$, $\hat{Y}_{OPT}$	38
2.3	Ratios of MSE 's of $\hat{F}(y_p)$ to $\hat{F}_{st}(y_p)$ and $\hat{F}_{opt}(y_p)$	39
2.4	Empirical Error Rate(%) for MELRT	40
2.5	Empirical Error Rate(%) for ANT	40

Chapter 1

Introduction

Likelihood and estimating equations provide common approaches to parametric inference. Recently, these approaches have been shown to be useful in nonparametric contexts.

Likelihood in nonparametric contexts, called empirical likelihood (described in Section 1 below), was recently introduced by Owen(1988, 1990, 1991). Owen has shown that the empirical likelihood ratio statistics have limiting chi-square distributions in certain situations, and has shown how to obtain tests and confidence limits for parameters, expressed as functionals, $\theta(F)$ of an unknown distribution function F . Other asymptotic properties and the possibility of correcting likelihood ratio statistics or their signed roots have been studied by DiCiccio and Romano(1989), Hall(1990), DiCicco, Hall and Romano(1989, 1991) and others.

Estimating equations is another important method in statistics. Qin and Lawless(1994, 1995) showed how to link empirical likelihood and estimating equations, and have shown that empirical likelihoods lead to asymptotic results similar to those

under parametric likelihoods.

Actually, Hartley and Rao(1968) gave the original idea of empirical likelihood in sample survey context as early as 1968. They called the method “scale-load” approach.

The most of above discussions deal with simple random sampling. Since simple random sampling is rarely used in practice from both practical and theoretical consideration, we will consider the case of stratified random sampling. The sampling scheme in each stratum maybe simple random sampling or a more complex sampling method.

In this chapter, we will first describe the empirical likelihood(EL) and an asymptotic set-up in Section 1.1. Section 1.2 describes briefly the problems discussed in chapters 2 to 4 and our main results. Section 1.3 covers the problems discussed in Chapter 5. Section 1.4 describes the problems studied in Chapter 6.

1.1 Empirical likelihood and asymptotic setup

1.1.1 Empirical likelihood

Suppose that a target population is divided into H strata with known weights W_h for all strata h , $\sum_h W_h = 1$. In stratum h there are N_h units with values x_{hi} ($i = 1, \dots, N_h; h = 1, \dots, H$), where x_{hi} is d -dimension variable. Denote the h^{th} stratum population mean, median and distribution function by $\bar{X}_h$, m_{X_h} and

$$F_{hN_h}(x) = N_h^{-1} \{1(x_{h1} < x) + \dots + 1(x_{hN_h} < x)\}$$

respectively, and $1(x_{hi} < x) = (1(x_{hi,1} < x_1), \dots, 1(x_{hi,d} < x_d))^T$, $x_{hi,j}$ and x_j are j -th component of vector x_{hi} and x respectively, $1(x_{hi,j} < x_j)$ is the indicator of set $(x_{hi,j} < x_j)$. Also, let $\bar{X}$, m_X and $F_X(x) = \sum_{h=1}^H W_h F_{hN_h}(x)$ be the mean, median and distribution function of the target population respectively, where $N = \sum_{h=1}^H N_h$, and $\bar{x} = \sum_{h=1}^H W_h \bar{X}_h$.

Suppose that $x_{h1}, \dots, x_{hn_h}$ is a simple random sample(SRS) without replacement from stratum h with distribution function F_{hN_h} for all h , and that the samples are selected independently across the strata. As argued by Chen and Qin (1993), the empirical likelihood for the above sampling scheme can be approximated by

$$L = \prod_{h=1}^H \prod_{i=1}^{n_h} p_{hi} \quad (1.1)$$

if $n_h \ll N_h$ and N_h is large, where $p_{hi} = Pr(x_h = x_{hi})$ and x_h has the distribution function F_{hN_h} .

We will call (1.1) the empirical likelihood. Maximizing (1.1) subject to some additional conditions according to available auxiliary information will give us more efficient estimators of parameters of interest, denoted as maximum empirical likelihood estimators or MELE.

1.1.2 Asymptotic setup

It is difficult to study the finite sample properties of MELE theoretically. Hence, we study their asymptotic properties. The finite sample properties of MELE will be studied through limited simulations. Let $\{P_\nu, \nu = 1, 2, \dots\}$ be a sequence of finite populations. Throughout the paper ν is used as the index of the finite population. The asymptotic setup here is that both the sample size n_h and the stratum size N_h

for each h go to infinity as $\nu \rightarrow \infty$, and F_{h,N_h} goes to distribution function F_h , but we will suppress the index ν for convenience. We can suppose the population of stratum h as a SRS sample from a super population F_h , hence $F_{h,N_h} \rightarrow F_h$. We also suppose that $n/n_h \rightarrow k_h > 0$, $h = 1, \dots, H$, where $n = \sum_{i=1}^H n_h$.

1.2 Outline of Chapters 2 to 4

In this section, we present an outline of Chapters 2 to 4. First, Chapters 2 to 4 all deal with stratified sampling. However, in Chapters 2, 3, we use simple random sampling in each stratum, while in Chapter 4 we permit more complex sampling in each stratum. Samples are all taken independently across strata.

1.2.1 Introduction to Chapter 2

Data set and auxiliary information

In order to be consistent with the convention of using x as the auxiliary variable and y as the response variable, we change the notation for characteristics of population to $z = (x^r, y^r)^r$, where x is auxiliary variable of dimension $d - p$ and y is the response variable of dimension p .

We suppose that samples are taken in each stratum by simple random sampling without replacement. Samples are denoted by

$$z_{h1}, \dots, z_{hn_h}, \quad h = 1, \dots, H.$$

Suppose some auxiliary information about x is available, such as population mean $\bar{X}$ or m_x when x is a scalar. In the following we suppose that only the overall mean

$\bar{X}$ is known from external sources such as demographic projections.

MEL estimators

The information $\bar{X}$ can be used to adjust our estimators for $p_{11}, \dots, p_{Hn_H}$, that is, to force these estimators to satisfy

$$\sum_h W_h \sum_i p_{hi} x_{hi} = \bar{X}. \quad (2.1)$$

Maximizing (1.1) subject to conditions (2.1) and $\sum_i p_{hi} = 1, p_{hi} > 0$ to get estimators $\hat{p}_{hi}$, we estimate the population mean $\bar{Y}$ by maximum empirical likelihood(MEL)

$$\hat{\bar{Y}} = \sum_h W_h \sum_i \hat{p}_{hi} y_{hi}$$

and the MEL estimator of distribution function $F_Y(y)$ as

$$\hat{F}_Y(y) = \sum_h W_h \sum_i \hat{p}_{hi} 1(y_{hi} < y).$$

Asymptotic properties of MELE

Under appropriate conditions, we can prove that the estimator $\hat{\bar{Y}}$ has the same asymptotic variance as the optimal regression estimator

$$\hat{\bar{Y}}_{opt} = \bar{y}_{st} + \hat{B}_0(\bar{X} - \bar{x}_{st}),$$

where $\hat{B}_0$ is given by Rao and Liu(see Chapter 2. Section 2). Further, $\hat{F}_Y(y)$ is asymptotically more efficient than $\hat{F}_n(y) = \sum_h W_h \hat{F}_{h,n_h}(y)$, the customary cumulative distribution function. It is possible to construct an optimal regression estimator of $F_Y(y)$ by changing y_{hi} to the indicator variable $1(y_{hi} \leq y)$ in the formula for $\hat{\bar{Y}}_{opt}$ (Rao and Liu, 1992). This estimator, called $\hat{F}_{opt}(y)$, is asymptotically equivalent to $\hat{F}_Y(y)$, but it may not be monotone unlike the MEL estimator $\hat{F}_Y(y)$.

Our simulation studies show that $\hat{Y}$ outperforms the GREG estimator, especially for populations with widely varying intercept terms across strata. It is also considerably more efficient than the stratified mean $\bar{y}_{st}$, except in the case of population with a weak relationship between y and x . As expected, $\hat{Y}$ and the optimal estimator are equally efficient. However, the MEL estimator $\hat{F}_Y(y)$ outperforms the empirical distribution function $\hat{F}_{st}(y)$. Although $\hat{F}_Y(y)$ and $\hat{F}_{opt}(y)$ are asymptotically equivalent, our simulation results indicate that $\hat{F}_Y(y)$ is often significantly more efficient than $\hat{F}_{opt}(y)$ in finite samples.

Empirical likelihood ratio test

Since (1.1) is defined as empirical likelihood, we can conduct empirical likelihood ratio test (ELRT) on parameter $\bar{Y}$. Let $\theta = (\theta_k^r, \theta_u^r)^r$, $\theta_k = \bar{X}$ and $\theta_u = \bar{Y}$, then $E(z) = \sum_h W_h E(z_h) = \theta$ and we define the profile empirical likelihood ratio function as

$$R_E(\theta) = \sup\{\prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi} \mid p_{hi} \geq 0, \sum_i p_{hi} = 1, \sum_h W_h \sum_i p_{hi} z_{hi} = \theta\}. \quad (2.2)$$

Maximizing $R_E(\theta)$ will give an estimator $\tilde{\theta}_u$ of the parameter θ_u . hence we may define

$$W_E(\theta) = 2 \log R_E(\tilde{\theta}) - 2 \log R_E(\theta_0)$$

as the empirical likelihood ratio statistic, where $\theta_0 = (\bar{X}^r, \bar{Y}_0^r)^r$ and $\bar{Y}_0$ is specified.

Under appropriate conditions, we get

$$W_E(\theta_0) \xrightarrow{L} \chi^2(p) \quad \text{as } \nu \rightarrow \infty$$

under the hypothesis $H_0 : \bar{Y} = \bar{Y}_0$.

1.2.2 EL and estimating equations

In Chapter 3, we combine EL, estimating equations and stratified random sampling together. The auxiliary information here is in the form of $r \geq p$ functionally independent unbiased estimation functions, that is functions $g_j(x, \theta)$, $j = 1, \dots, r$ such that $E_F\{g_j(x, \theta)\} = 0$. In vector form, we have

$$g(x, \theta) = (g_1(x, \theta), \dots, g_r(x, \theta))^T,$$

where $g(x, \theta)$ satisfies

$$E_F\{g(x, \theta)\} = \sum_h W_h E_{F_h}\{g(x_h, \theta)\} = 0. \quad (2.3)$$

The basic idea is to maximize empirical likelihood subject to constraints provided by (2.3). Since we need to consider empirical likelihood ratio tests, we first define empirical likelihood ratio function, then get the empirical likelihood ratio estimator for the parameter of interest, θ . The estimator of θ obtained by this method is called the MELR estimator.

MELR estimators

Since empirical likelihood function is

$$L(F) = \prod_{h=1}^H \prod_{i=1}^{n_h} p_{hi}. \quad (2.4)$$

and (2.4) is maximized by the empirical distribution function $F_n(x) = \sum_h W_h F_{n_h}(x)$, where $n = \sum_h n_h$, $F_{n_h} = \frac{1}{n_h} \sum_i 1(x_{hi} \leq x)$, the empirical likelihood ratio is then defined as $R(F) = L(F)/L(F_n)$, which reduces to

$$R(F) = \prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi}. \quad (2.5)$$

Since we are interested in estimating the parameter θ , and since we know the estimating equation (2.3), we define the empirical likelihood ratio function

$$R_E(\theta) = \sup\{\Pi_{h=1}^H \Pi_{i=1}^{n_h} n_h p_{hi} \mid p_{hi} \geq 0, \sum_i p_{hi} = 1, \sum_h W_h \sum_i p_{hi} g_{hi}(\theta) = 0\}, \quad (2.6)$$

where $g_{hi}(\theta) = g(x_{hi}, \theta)$ for all h and i . We may maximize $R_E(\theta)$ to obtain the MELR estimate $\tilde{\theta}$ of the parameter θ . In addition, this yields estimates $\tilde{p}_{hi}$'s of p_{hi} 's, and an estimate for the distribution F as

$$\tilde{F}_n(s) = \sum_h W_h \sum_i \tilde{p}_{hi} 1(x_{hi} \leq x).$$

The asymptotic properties of $\tilde{\theta}$ and $\tilde{F}_n(y)$ have also been studied. The results are as follows:

$$\sqrt{n}(\tilde{\theta} - \theta_0) \rightarrow N(0, (M_{21} M_{11}^{-1} M_{12})^{-1}),$$

$$\sqrt{n}(\tilde{F}_n(x) - F(x)) \rightarrow N(0, W), W = \sum_h W_h^2 k_h F_h(x)[1 - F_h(x)] - B U B^T.$$

The terms involved in the above asymptotic distributions are given in Chapter 3. From the asymptotic variance of $\sqrt{n} \tilde{\theta}$, we can see that it is of the usual sandwich form. The matrix U is non-negative definite so that $\tilde{F}_n(s)$ is at least as efficient as $F_n(y)$, the sample cumulative distribution function.

MELR testing with constraints

The empirical likelihood methods can be extended to deal with the case in which there are constraints on parameters and even to testing on these constraints. Suppose that q -dimensional constraints on θ are of the form

$$r(\theta) = 0, \quad (2.7)$$

where $r(\theta)$ is a $q \times 1$ ($q \leq p$) and the $q \times p$ matrix $R(\theta) = \frac{\partial r}{\partial \theta^T}$ is of full rank q . We first consider how to estimate θ . This can be done by defining

$$G_2 = -\frac{1}{n} \log R_E(\theta) + \nu^T r(\theta),$$

where ν is a $q \times 1$ vector of Lagrange multipliers, and then differentiating G_2 with respect to θ and ν to get the estimator for θ denoted by $\tilde{\theta}_r$.

Now the problem about how to test $H_0 : r(\theta) = 0$ can be handled by any of the three most popular methods based on likelihood: likelihood-ratio test, Lagrange-multiplier test and Wald test. Of course here they should be based on empirical likelihood. We can prove that these methods are asymptotically equivalent. The conclusions in this chapter show that the parametric likelihood and non-parametric likelihood have similar properties.

1.2.3 Pseudo EL and estimating equations

Sampling Design

Our sampling design in this chapter is stratified sampling, where samples are taken from each stratum according to some specific design, such as simple random sampling(SRS), probability proportional to size(PPS) sampling and samples are independent across strata.

Pseudo EL

Suppose P_ν with d -dimension characteristic x_ν has the unknown distribution F_ν and a p -dimension parameter θ associated with it, and the h^{th} stratum of P_ν has d -dimension characteristic $x_{\nu h}$ which has the unknown distribution $F_{\nu h}$. Suppose $x_{h1}, \dots, x_{hN_h}$, the values of units in h^{th} stratum of P_ν , is a random sample

from a super-population, say $F_{\nu_h}^s$. If the entire population P_ν were available, the corresponding likelihood function would be

$$L(F) = \prod_{h=1}^H \prod_{i=1}^{N_h} p_{hi}, \quad (2.8)$$

where $p_{hi} = F_{\nu_h}^s(x_{hi}) - F_{\nu_h}^s(x_{hi}^-)$. The log-likelihood function then is

$$l(F) = \sum_{h=1}^H \sum_{i=1}^{N_h} \log p_{hi}. \quad (2.9)$$

where $N = \sum_{h=1}^H N_h$. But we only have a sample, say, s , from the finite population and if we view (2.9) as a finite population total, then we may have available a design unbiased estimator of $l(F)$, namely,

$$\hat{l}(F) = N \sum_h \sum_{i \in s_h} d_{hi}(s) \log p_{hi}, \quad (2.10)$$

where $d_{hi}(s)$ are design weights with $E[\sum_{i \in s_h} d_{hi}(s) \log p_{hi}] = N^{-1} \sum_{i=1}^{N_h} \log p_{hi}$ for all h , and E refers to expectation with respect to the sampling design. Obviously we have $E(\hat{l}(F)) = l(F)$. We will call (2.10) the “pseudo empirical likelihood”.

Using the pseudo empirical likelihood (2.10), we can obtain a maximum pseudo empirical likelihood ratio estimator for the parameter of interest. These estimators will be called MPELR estimators.

MPELR estimators

Since $\sum_i p_{hi} = 1$, we know (2.10) is maximized by $F_n(x) = \sum_h W_h F_{n_h}(x)$, where $n = \sum_h n_h$, $F_{n_h}(x) = \sum_{i \in s_h} \omega_{hi}(s) 1(x_{hi} < x)$, $\omega_{hi}(s) = d_{hi}(s) / \sum_{i \in s_h} d_{hi}(s)$. The log-empirical likelihood ratio is then defined as

$$\begin{aligned} \log \hat{R}(F) &= \log \hat{l}(F) - \log \hat{l}(F_n) \\ &= N \left(\sum_h \sum_{i \in s_h} d_{hi}(s) \log p_{hi} - \sum_h \sum_{i \in s_h} d_{hi}(s) \log \omega_{hi}(s) \right). \end{aligned} \quad (2.11)$$

We also assume that information about θ and F is available in the form of $r \geq p$ functionally independent unbiased estimation functions, that is. functions $g_j(x, \theta), j = 1, \dots, r$ such that $E_F\{g_j(x, \theta)\} = 0$. In vector form, we have

$$g(x, \theta) = (g_1(x, \theta), \dots, g_r(x, \theta))^T,$$

where $g(x, \theta)$ satisfies

$$E_F\{g(x, \theta)\} = \sum_h W_h E_{F_h}\{g(x_h, \theta)\} = 0. \quad (2.12)$$

Now we define the (negative) log-pseudo empirical likelihood ratio function of θ by

$$l_E(\theta) = -\frac{1}{N} \sup\{\log \hat{R}(F) | p_{hi} \geq 0, \sum_i p_{hi} = 1, \sum_h W_h \sum_i p_{hi} g_{hi}(\theta) = 0\}, \quad (2.13)$$

where $g_{hi}(\theta) = g(x_{hi}, \theta)$ for all h and i .

We may minimize $l_E(\theta)$ to obtain an estimator $\tilde{\theta}$ of the parameter θ (called MPELR estimator). In addition, this yields estimators $\tilde{p}_{hi}$'s, and therefore an estimator of the distribution function F as

$$\tilde{F}_n(x) = \sum_h W_h \sum_i \tilde{p}_{hi} 1(x_{hi} < x) \quad (2.14)$$

Asymptotic Results

Some conditions are needed to study the existence of solutions and the asymptotic properties of MPELR estimators. We assume that $n \rightarrow \infty$, $n_h/n \rightarrow k_h > 0$ and

$$\frac{\max_{1 \leq h \leq H, 1 \leq i \leq n_h} d_{hi}(s)}{\min_i d_{hi}(s)} = O\left(\frac{1}{n}\right), \quad \frac{\max_i d_{hi}(s)}{\min_i d_{hi}(s)} = O(1) \quad (2.15)$$

as $\nu \rightarrow \infty$. The assumption (2.15) means that no survey weight is disproportionately large. Due to some technical difficulty, only stratified SRS and PPS with replacement are discussed.

Since our strata weights W_h are known and samples from different strata are independent, we should also construct the weights to satisfy

$$E\left(\sum_{i \in s_h} d_{hi}(s)\right) = W_h.$$

Under the conditions above and some other conditions, we establish asymptotic normality of the estimator for both the parameter θ and the distribution function $\tilde{F}_n(x)$.

Pseudo empirical likelihood ratio test(PELRT) is discussed. Here the asymptotic distribution of PELRT is the sum of weighted χ_1^2 distribution. The reason for that is that pseudo-EL is used as the likelihood. Methods of approximating the asymptotic distribution are also discussed.

1.3 EL and M-estimation

M-estimation plays an important role in robust parametric inference and in non-parametric inference. We consider a new class of M-estimators here. Suppose that some auxiliary information in the form of general estimating equations is available. We can use this information to improve our estimator for distribution function, then replace F in M-functional $\theta(F)$ by this improved estimator for distribution function to get an estimator $\hat{\theta}$. Since this estimator $\hat{\theta}$ combines the information from auxiliary information and M-function, it will be more efficient than the usual M-estimator.

Auxiliary information

We assume that auxiliary information is in the form of r (≥ 1) functions $g_1(x)$,

$\cdots, g_r(x)$ such that

$$E_F g(x) = \sum_h W_h E_h g(x) = 0, \quad (3.1)$$

where $g(x) = (g_1(x), \cdots, g_r(x))^T$, E_h denotes the expectation with respect to the distribution function of stratum h . Here $g(x)$ does not contain any unknown parameters, hence is different from the problems discussed in chapters 2 to 4.

M-estimation

An M-functional $\theta(F)$ associated with distribution function F is defined as a root θ_0 of the equation

$$E_F \psi(x, \theta_0) = \int \psi(x, \theta_0) dF(x) = 0, \quad (3.2)$$

where $x \sim F$ and $x, \theta_0, \psi(x, \theta_0) \in R$. For a stratified sample $x_{11}, \cdots, x_{1n_1}; \cdots; x_{H1}, \cdots, x_{Hn_H}$ from $F(\cdot)$, the usual estimator of θ_0 is given by $\hat{\theta}_n = \theta(F_n)$, where $F_n(\cdot)$ is the natural stratified empirical distribution function, i.e., $F_n(x) = \sum_h \frac{W_h}{n_h} \sum_i 1(x_{hi} < x)$. $\theta(F_n)$ is called the M-estimator corresponding to ψ .

EL and M-estimation

The method to combine EL and M-estimation can be divided into two steps. First, we use auxiliary information (3.1) to estimate $F(x)$. That is to define EL as before as

$$L = \prod_{h=1}^H \prod_{i=1}^{n_h} p_{hi}, \quad (3.3)$$

where $p_{hi} = Pr(x_h = x_{hi})$. Under condition (3.1), i.e.,

$$\sum_h W_h \sum_i p_{hi} g(x_{hi}) = 0, \quad (3.4)$$

and

$$\sum_i p_{hi} = 1, \quad p_{hi} \geq 0$$

maximizing L to get estimators for p_{hi} :

$$\hat{p}_{hi} = \frac{1}{n_h} \frac{1}{1 + m_h t^\tau (g(x_{hi}) - \hat{g}_h)},$$

where

$$\hat{g}_h = \sum_{i=1}^{n_h} \hat{p}_{hi} g(x_{hi}), \quad m_h = n W_h n_h^{-1}, \quad n = \sum_{h=1}^H n_h,$$

t is the solution to

$$\sum_h W_h \hat{g}_h = \sum_h W_h \sum_i \hat{p}_{hi} g(x_{hi}) = 0.$$

Now let

$$\hat{F}_n(x) = \sum_h W_h \sum_i \hat{p}_{hi} 1[x_{hi} \leq x],$$

then $\hat{F}_n(x)$ can be regarded as an alternative estimator of the distribution function of $F(x)$.

The new M-estimator to θ_0 is then defined as by $\hat{\theta}_n$, which is a root u_0 of the equation

$$\int \psi(x, u_0) d\hat{F}_n(x) = \sum_h W_h \sum_i \hat{p}_{hi} \psi(x_{hi}, u_0) = 0. \quad (3.5)$$

We will also call $\hat{\theta}_n$ the ELM-estimator corresponding to ψ .

Asymptotic properties

We prove that $\hat{\theta}_n$ is asymptotic normal and that it has smaller asymptotic variance compared to the one that does not use the auxiliary information.

1.4 EL in the Presence of Measurement Error

Sometime we are in the situation to measure a characteristic of interest of a target population. We have several different “instruments”, one of them is “accurate”,

or it has smaller measurement error compared to the other ones. In practice we treat this instrument as “perfect”, or no measurement error. All other instruments will have larger measurement error and we treat them as “imperfect”. The perfect instrument can be used only very limited times due to cost, whereas the imperfect ones, however, can be used more frequently. Now the question is how to combine all the data available to make inference on parameters of interest.

Data set

Suppose that there are H different instruments used in measuring the characteristic of interest, the distribution function associated with instrument h ($h = 1, \dots, H$) is $F_h(x) = P[x_h \leq a]$, with unknown parameter θ , where θ is p -dimension vector. and the H -th measuring instrument is taken as perfect. These H different populations are related by the common parameter θ . We also assume that information about θ and F_h is available in the form of $r_h \geq p$ functionally independent unbiased estimation functions, that is functions $g_h(x_h, \theta)$ such that

$$E_{F_h}(g_h(x_h, \theta)) = 0, \quad (4.1)$$

where $g_h(x, \theta)$ is a r_h -dimension vector function.

We further suppose that $x_{h1}, \dots, x_{hn_h}$ is an *i.i.d.* sample from $F_h(x)$, and samples measured by different instruments are independent.

EL in the presence of measurement error

The empirical likelihood based on the independent samples

$$\{x_{h1}, \dots, x_{hn_h}; h = 1, \dots, H\}$$

is given by

$$L(F_1, \dots, F_H) = \prod_{i=1}^H \prod_{j=1}^{n_h} p_{hi}, \quad (4.2)$$

where $p_{hi} = Pr(x_h = x_{hi})$ and $\sum_{i=1}^{n_h} p_{hi} = 1$ for each h . The empirical likelihood ratio is then defined as

$$R(F_1, \dots, F_H) = \prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi}. \quad (4.3)$$

Since we know the estimating equation (4.1), we define the empirical likelihood ratio function for θ as

$$R_E(\theta) = \sup \{ \prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi} \mid p_{hi} \geq 0, \sum_i p_{hi} = 1, \sum_i p_{hi} g_{hi}(\theta) = 0 \}, \quad (4.4)$$

where $g_{hi}(\theta) = g_h(x_{hi}, \theta)$ for all h and i .

We can maximize $R_E(\theta)$ to obtain an estimator $\tilde{\theta}$ of the parameter θ , called the maximum empirical likelihood ratio estimator (MELRE). In addition, this yields estimators $\tilde{p}_{hi}$, and an estimator for the true distribution function F_H as

$$\tilde{F}_{H,n_H}(x) = \sum_{i=1}^{n_H} \tilde{p}_{Hi} 1(x_{Hi} < x). \quad (4.5)$$

Asymptotic properties

We prove that $\tilde{F}_{H,n_H}(x)$ is asymptotically more efficient than $F_{H,n_H}(x)$, the cumulative distribution function of $x_{H1}, \dots, x_{Hn_H}$. The estimator $\tilde{\theta}$ usually is significantly more efficient than the estimator using only $x_{H1}, \dots, x_{Hn_H}$.

Chapter 2

Empirical Likelihood Inference for Finite Populations with Auxiliary Information Using Stratified Random Sampling

2.1 Introduction

Suppose that a target population is divided into H strata with known weights W_h for all strata h , $\sum_h W_h = 1$. In stratum h there are N_h units with values z_{hi} ($i = 1, \dots, N_h; h = 1, \dots, H$), where z_{hi} is d -dimension variables, $z_{hi} = (x_{hi}^\tau, y_{hi}^\tau)^\tau$. x_{hi} and y_{hi} are $d - p$ and p dimensions respectively and τ denotes the transpose. Denote the h^{th} stratum population mean, median and distribution function by

$(\bar{X}_h^\tau, \bar{Y}_h^\tau)^\tau, (m_{x_h}^\tau, m_{y_h}^\tau)^\tau$ and

$$F_{hN_h}(z) = N_h^{-1} \{1(z_{h1} < z) + \cdots + 1(z_{hN_h} < z)\}$$

respectively, where $1(z_{hi} < z) = (1(z_{hi,1} < z_1), \dots, 1(z_{hi,d} < z_d))^\tau$, $z_{hi,j}$ and z_j are the j -th component of vectors z_{hi} and z respectively, $1(z_{hi,j} < z_j)$ is the indicator of set $(z_{hi,j} < z_j)$. Also, let $(\bar{X}^\tau, \bar{Y}^\tau)^\tau, (m_{\bar{X}}^\tau, m_{\bar{Y}}^\tau)$ and $F_Z(z) = (F_{\bar{X}}^\tau(x), F_{\bar{Y}}^\tau(y))^\tau = \sum_{h=1}^H W_h F_{hN_h}(z)$ be the mean, median and distribution function of the target population respectively, where $N = \sum_{h=1}^H N_h$, $\bar{X} = \sum_{h=1}^H W_h \bar{X}_h$ and $\bar{Y} = \sum_{h=1}^H W_h \bar{Y}_h$. The reason we separate z_{hi} into two parts is that we are going to suppose that some auxiliary information such as $\bar{X}$ or m_x is available when we want to make inference about target population parameters such as $\bar{Y}$ or m_Y .

In many practical problems, some population values of the auxiliary variable x are known, such as the population mean or median from external sources such as demographic projections. By using this knowledge, we would like to provide improved inferences on the population parameters associated with the characteristic of interest y , such as the population mean $\bar{Y}$, the population distribution function of y , $F_Y(y)$, and the population median m_Y . Empirical likelihood methods, recently introduced by Owen (1988, 90, 91) in the context of iid random variables, provide a systematic nonparametric approach to utilizing auxiliary information in making inference on the parameters of interest. Hartley and Rao (1968) gave the original idea of empirical likelihood in sample survey context, using their “scale-load” approach. Under simple random sampling, they obtained the maximum empirical likelihood estimator of $\bar{Y}$ when only $\bar{X}$ is known and showed that it is approximately equal to the regression estimator as the sample size, n , increases. Chen and Qin (1993)

extended these results to cover distribution function and median of Y , i.e., $F_Y(y)$ and m_Y . Qin and Lawless (1994, 95) used empirical likelihood and estimating equations in the iid case to deal with interval estimation and hypothesis testing. They obtained an empirical likelihood ratio test statistic (ELRS) and its asymptotic distribution under null hypothesis. The main aim of this chapter is to deal with the case of stratified simple random sampling.

In section 2.2, we derive the maximum empirical likelihood estimator (MELE) for the parameters of interest. Some large sample properties of MELE are also established. Application of MELE to small area estimation will be given in section 2.3. In section 2.4, empirical likelihood ratio test (ELRT) on $\bar{Y}$ is considered and its asymptotic distribution under the null hypothesis is obtained. In section 2.5, we present the result of a limited simulation study on the finite sample properties of the empirical likelihood method. Proofs of Theorems in section 2.2 and 2.4 are given in section 2.6.

2.2 Maximum Empirical Likelihood Estimator

Suppose that $z_{h1}, \dots, z_{hn_h}$ is a simple random sample without replacement from stratum h with distribution function F_{hN_h} for all h , and that the samples are selected independently across the strata. As argued by Chen and Qin (1993) for simple random sampling, the empirical likelihood for the above sampling scheme can be approximated by

$$L = \prod_{h=1}^H \prod_{i=1}^{n_h} p_{hi} \quad (2.1)$$

where $p_{hi} = Pr(z_h = z_{hi})$ and $z_h = (x_h^T, y_h^T)^T$ has the distribution function F_{hN_h} .

We consider the case of known vector of population means, $\bar{X}$, of the variable $x = (x_1, \dots, x_{d-p})^T$:

$$\bar{X} = \sum_{h=1}^H W_h E(x_h) = \sum_{h=1}^H W_h N_h^{-1} \sum_{i=1}^{N_h} x_{hi}. \quad (2.2)$$

Our results can be easily generalized to other forms of auxiliary information such as a known population median in the case of scalar x .

The maximum empirical likelihood estimator should be sought among distribution functions satisfying (2.2). Using the same argument as in Owen (1990), we need consider only estimators of F_{hN_h} whose support are contained in the set of observations. The problem therefore reduces to maximizing

$$l = \log L = \sum_h \sum_i \log p_{hi} \quad (2.3)$$

subject to

$$\sum_i p_{hi} = 1 \quad (h = 1, \dots, H; p_{hi} > 0) \quad (2.4)$$

and

$$\sum_h W_h \sum_i p_{hi} x_{hi} = \bar{X}, \quad (2.5)$$

where $\sum_h = \sum_{h=1}^H$ and $\sum_i = \sum_{i=1}^{n_h}$. A solution for the above problem exists with probability tending to one as the sample sizes go to infinity in each stratum. The proof of this conclusion will be given in section 2.6.

Using the Lagrange multiplier method, let

$$l_1 = \sum_h \sum_i \log p_{hi} - \sum_h \lambda_h (\sum_i p_{hi} - 1) - n\psi^T (\sum_h W_h \sum_i p_{hi} x_{hi} - \bar{X}),$$

where $n = \sum_h n_h$ and ψ is a $(d-p)$ -dimension vector. Then

$$\frac{\partial l_1}{\partial p_{hi}} = p_{hi}^{-1} - \lambda_h - nW_h \psi^T x_{hi} = 0, \quad i = 1, \dots, n_h; h = 1, \dots, H. \quad (2.6)$$

From (2.6) we get

$$(\lambda_h + nW_h\psi^\tau x_{hi})p_{hi} = 1,$$

and summing this expression over i leads to

$$\lambda_h = n_h - nW_h\psi^\tau \sum_i p_{hi}x_{hi},$$

using (2.4). Hence

$$\begin{aligned} p_{hi} &= [\lambda_h + nW_h\psi^\tau x_{hi}]^{-1} \\ &= n_h^{-1}[1 + m_h\psi^\tau(x_{hi} - \hat{X}_h)]^{-1}, \end{aligned}$$

where $\hat{X}_h = \sum_i p_{hi}x_{hi}$ and $m_h = nW_h n_h^{-1}$. Therefore, the estimators of p_{hi} 's are the solutions to the following system of equations:

$$p_{hi} = n_h^{-1}[1 + m_h\psi^\tau(x_{hi} - \hat{X}_h)]^{-1}, \quad (2.7)$$

$$\hat{X}_h = \sum_i p_{hi}x_{hi}, \quad h = 1, \dots, H, \quad (2.8)$$

$$\sum_h W_h \sum_i p_{hi}x_{hi} = \bar{X}, \quad (2.9)$$

subject to $0 \leq p_{hi} \leq 1$. We denote the solutions of (2.7), (2.8) and (2.9) as $\hat{p}_{hi}$ and $\hat{\psi}$. We call the above method as the empirical likelihood method. The estimator we get by using this method is called the Maximum Empirical Likelihood Estimator (MELE). We will now discuss how to solve (2.7), (2.8) and (2.9) in the following.

2.2.1 Numerical Evaluation for MELE

It is difficult to solve (2.7), (2.8) and (2.9) directly. But they can be solved indirectly as follows. Using Lagrange multiplier method, we maximize (2.3) subject to (2.4)

and

$$\sum_i p_{hi}(x_{hi} - \bar{X}_h) = 0; \quad h = 1, \dots, H$$

for each given $(\bar{X}_1, \dots, \bar{X}_H)$ such that $\sum W_h \bar{X}_h = \bar{X}$. We get the maximum as

$$-\sum_h \sum_i \log[1 + \xi_h^\tau(x_{hi} - \bar{X}_h)] - \sum_h \log n_h$$

with the Lagrange multiplier ξ_h satisfying

$$\sum_i \frac{x_{hi} - \bar{X}_h}{1 + \xi_h^\tau(x_{hi} - \bar{X}_h)} = 0.$$

where implicitly

$$p_{hi} = \frac{1}{n_h[1 + \xi_h^\tau(x_{hi} - \bar{X}_h)]}.$$

Hence, we can say that ξ_h is a function of $\bar{X}_h$ and we then maximize

$$-\sum_h \sum_i \log[1 + \xi_h^\tau(\bar{X}_h)(x_{hi} - \bar{X}_h)]$$

subject to

$$\sum_h W_h \bar{X}_h = \bar{X}.$$

Using the Lagrange method again, let

$$l_2(\bar{X}_1, \dots, \bar{X}_H, \phi) = -\sum_h \sum_i \log[1 + \xi_h^\tau(\bar{X}_h)(x_{hi} - \bar{X}_h)] - n\phi^\tau(\sum_h W_h \bar{X}_h - \bar{X}),$$

solving

$$\begin{aligned} \frac{\partial l_2}{\partial \bar{X}_h} &= -\sum_i \frac{(\frac{\partial \xi_h}{\partial \bar{X}_h})^\tau(x_{hi} - \bar{X}_h) - \xi_h}{1 + \xi_h^\tau(\bar{X}_h)(x_{hi} - \bar{X}_h)} - nW_h\phi \\ &= n_h\xi_h - nW_h\phi = 0, \end{aligned}$$

we get $\xi_h = m_h\phi$. Therefore,

$$\sum_i \frac{x_{hi} - \bar{X}_h}{1 + m_h\phi^\tau(x_{hi} - \bar{X}_h)} = 0, \quad h = 1, \dots, H, \quad (2.10)$$

$$p_{hi} = n_h^{-1} [1 + m_h \phi^\tau (x_{hi} - \bar{X}_h)]^{-1}, \quad 0 \leq p_{hi} \leq 1. \quad (2.11)$$

Or $\bar{X}_h$ is a function of ϕ , and ϕ should be chosen such that

$$\sum_h W_h \bar{X}_h(\phi) = \bar{X}. \quad (2.12)$$

It is evident that the solutions ϕ , $\bar{X}_h(\phi)$, p_{hi} to (2.10), (2.11) and (2.12), defined as $\hat{\phi}$, $\bar{X}_h(\hat{\phi})$, $\hat{p}_{hi}$, are the same as the solutions to (2.7), (2.8) and (2.9). Hence $\bar{X}_h(\hat{\phi}) = \hat{X}_h$ satisfies (2.10), i.e.

$$\sum_i \frac{x_{hi} - \hat{X}_h}{1 + m_h \hat{\phi}^\tau (x_{hi} - \hat{X}_h)} = 0. \quad (2.13)$$

We now prove that (2.10) and (2.12) have a unique solution. Denote

$$W(\bar{X}_h, \phi) = n_h^{-1} \sum_i \frac{x_{hi} - \bar{X}_h}{1 + m_h \phi^\tau (x_{hi} - \bar{X}_h)}. \quad (2.14)$$

then, noting (2.10),

$$\begin{aligned} \frac{\partial W(\bar{X}_h, \phi)}{\partial \bar{X}_h} &= -I_{d-p} + n W_h n_h^{-2} \sum_i \frac{(x_{hi} - \bar{X}_h) \phi^\tau}{[1 + m_h \phi^\tau (x_{hi} - \bar{X}_h)]^2} \\ &= -I_{d-p} + n W_h n_h^{-2} \sum_i \frac{(x_{hi} - \bar{X}_h)}{1 + m_h \phi^\tau (x_{hi} - \bar{X}_h)} \phi^\tau \\ &\quad - \frac{m_h^2}{n_h} \left(\sum_i \frac{(x_{hi} - \bar{X}_h)(x_{hi} - \bar{X}_h)^\tau}{[1 + m_h \phi^\tau (x_{hi} - \bar{X}_h)]^2} \right) \phi \phi^\tau \\ &= -I_{d-p} - \frac{m_h^2}{n_h} \left(\sum_i \frac{(x_{hi} - \bar{X}_h)(x_{hi} - \bar{X}_h)^\tau}{[1 + m_h \phi^\tau (x_{hi} - \bar{X}_h)]^2} \right) \phi \phi^\tau, \end{aligned}$$

where I_{d-p} is the identity matrix of order $d - p$. For a given ϕ , $\frac{\partial W}{\partial \bar{X}_h}$ is nonsingular since the matrix

$$\sum_i \frac{(x_{hi} - \bar{X}_h)(x_{hi} - \bar{X}_h)^\tau}{[1 + m_h \phi^\tau (x_{hi} - \bar{X}_h)]^2}$$

is positive definite. If a solution $\bar{X}_h$ to $W(\bar{X}, \phi) = 0$ exists for each given ϕ , then $\bar{X}_h$ is an implicit function of ϕ , say $\bar{X}(\phi)$, by implicit theorem. Since

$$\begin{aligned} & \frac{\partial W(\bar{X}_h(\phi), \phi)}{\partial \phi} \\ &= \sum_i \frac{-\frac{\partial \bar{X}_h(\phi)}{\partial \phi} [1 + m_h \phi^\tau (x_{hi} - \bar{X}_h(\phi))] - m_h (x_{hi} - \bar{X}_h(\phi)) (x_{hi} - \bar{X}_h(\phi))^\tau}{n_h [1 + m_h \phi^\tau (x_{hi} - \bar{X}_h(\phi))]^2} \\ &= 0, \end{aligned}$$

we obtain, noting (2.4) and (2.11),

$$\frac{\partial \bar{X}_h}{\partial \phi} = -n W_h n_h^{-2} \frac{\sum_i (x_{hi} - \bar{X}_h(\phi)) (x_{hi} - \bar{X}_h(\phi))^\tau}{[1 + m_h \phi^\tau (x_{hi} - \bar{X}_h(\phi))]^2} < 0$$

for all h . By applying the theorem of the mean we will know that

$$T(\phi) = \sum_h W_h \bar{X}_h(\phi) - \bar{X} = 0$$

will not have two different roots. Hence if a solution ϕ exists, then it is unique and we can find the ϕ satisfying (2.12).

We solve (2.10)-(2.12) by combining (2.10) and (2.12) as a nonlinear system of equations and then using NAG library function to get a solution for appropriately selected initial values, e.g., $\phi = 0$ and $\bar{X}_h = \bar{x}_h$. For this solution the condition $0 \leq p_{hi} \leq 1$ usually holds as seen in our simulation study.

Once we have the estimators $\hat{p}_{hi}$, we can estimate $\bar{Y}$ by

$$\hat{\bar{Y}} = \sum_h W_h \sum_i \hat{p}_{hi} y_{hi},$$

and $F_Y(y)$ by

$$\hat{F}_Y(y) = \sum_h W_h \sum_i \hat{p}_{hi} 1(y_{hi} < y).$$

2.2.2 Asymptotic Results

In this section, we discuss the asymptotic behavior of the MELE. We use $\| \cdot \|$ to denote Euclidean norm and $\sigma_2 > \sigma_1$ means that $\sigma_2 - \sigma_1$ is a positive definite matrix. Also, we assume that both the sample size n_h and the stratum size N_h for each h go to infinity as the index ν attached to n_h and N_h goes to infinity, i.e., $n_{h\nu}$ and $N_{h\nu}$ go to infinity as $\nu \rightarrow \infty$. However, for convenience, we will suppress the index ν in the following whenever possible. And we will denote the solution $\hat{\psi}$ by $\hat{\psi}_n$ since we are going to deal with large sample problems.

Theorem 1. *Suppose that as $\nu \rightarrow \infty$, $N_h, n_h, N_h - n_h$ go to infinity, $n/n_h \rightarrow k_h$, $k_h > 0$, and both $\sum_h W_h^2 N_h^{-1} \sum_i \|x_{hi}\|^3$ and $\sum_h W_h^2 N_h^{-1} \sum_i \|y_{hi}\|^3$ have an upper bound independent of ν , and $\sigma_{h,xx} = \text{Cov}(x_h, x_h) > \sigma_1 > 0$ for all h and ν , where σ_1 is a fixed positive definite matrix, then*

$$\sigma_\nu^{-\frac{1}{2}}(\hat{Y}_n - \bar{Y}) \xrightarrow{L} N_p(0, I_p),$$

$$\sigma_{F_\nu}^{-\frac{1}{2}}(\hat{F}_Y(y) - F_Y(y)) \xrightarrow{L} N_d(0, I_p),$$

as $\nu \rightarrow \infty$, where

$$\begin{aligned} \hat{F}_Y(y) &= \sum_h W_h \sum_i \hat{p}_{hi} 1(y_{hi} < y), \quad \hat{Y}_n = \sum_h W_h \sum_i \hat{p}_{hi} y_{hi}, \\ \sigma_\nu &= \sum_h W_h^2 (n_h^{-1} - N_h^{-1}) [\text{Var}(y_h) - 2B \text{Cov}^\tau(y_h, x_h) + B \text{Var}(x_h) B^\tau], \\ B &= \sum_h k_h W_h^2 \text{Cov}(y_h, x_h) [\sum_h k_h W_h^2 \text{Var}(x_h)]^{-1}, \\ \sigma_{xx} &= \sum_h k_h W_h^2 \text{Var}(x_h), \\ \sigma_{F_\nu} &= \sum_h W_h^2 (n_h^{-1} - N_h^{-1}) [F_{h,N_h}(1 - F_{h,N_h}) - 2G \text{Cov}^\tau(1(y_h < y), x_h) + G \text{Var}(x_h) G^\tau] \end{aligned}$$

$$G = \sum_h W_h^2 k_h \text{Cov}(1(y_h < y), x_h) \sigma_{xx}^{-1}.$$

From the results of Theorem 1, we find that the estimator $\hat{Y}$ has the same asymptotic variance as the optimal regression estimator

$$\hat{Y}_{opt} = \bar{y}_{st} + \hat{B}_0(\bar{X} - \bar{x}_{st}),$$

where

$$\hat{B}_0 = \sum_h W_h^2 n_h^{-1} s_{xy,h} / \sum_h W_h^2 n_h^{-1} s_{xx,h}$$

with

$$s_{xy,h} = \sum_i (y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h)^T / (n_h - 1),$$

$$s_{xx,h} = \sum_i (x_{hi} - \bar{x}_h)(x_{hi} - \bar{x}_h)^T / (n_h - 1).$$

Further, $\hat{F}_Y(y)$ is asymptotically more efficient than the customary estimator $\hat{F}_n(y) = \sum_h W_h \hat{F}_{h,n_h}(y)$. It is possible to construct an optimal regression estimator of $F_Y(y)$ by changing y_{hi} to the indicator variable $1(y_{hi} \leq y)$ in the formula for $\hat{Y}_{opt}$ (Rao and Liu, 1992). This estimator, $\hat{F}_{opt}(y)$, is asymptotically equivalent to $\hat{F}_Y(y)$, but it may not be monotone unlike the MEL estimator $\hat{F}_Y(y)$. Monotonicity of $\hat{F}_Y(y)$ in the case of scalar y ensures the calculation of quantiles, etc.

In above discussion we considered auxiliary information of the form of (2.2), but we can easily generalize it to the auxiliary information of the form

$$\sum_h W_h N_h^{-1} \sum_{i=1}^{N_h} w(x_{hi}) = 0.$$

The choice $w(x_{hi}) = x_{hi} - \bar{X}$ gives (2.2). When the population median m_x is known and x is a scalar, we let $w(x_{hi}) = 1_{[x_{hi} \leq m_x]} - 0.5$.

2.2.3 Variance Estimation

In this section, we consider the estimation of variance of $\hat{\bar{Y}}$ and $\hat{F}_Y(y)$. Under the conditions of Theorem 1 and from its proof, we can easily see that

$$\hat{\sigma}_\nu = \sum_h W_h^2 \frac{n_h^{-1} - N_h^{-1}}{n_h - 1} \sum_{i=1}^{n_h} [y_{hi} - \hat{\bar{Y}}_h - \hat{B}(x_{hi} - \hat{\bar{X}}_h)][y_{hi} - \hat{\bar{Y}}_h - \hat{B}(x_{hi} - \hat{\bar{X}}_h)]^\tau \quad (2.15)$$

is a consistent estimator of σ , the asymptotic variance of $\hat{\bar{Y}}$, in the sense that $n(\hat{\sigma} - \sigma) \rightarrow_p 0$, where

$$\begin{aligned} \hat{B} &= \sum_h n W_h^2 n_h^{-2} \sum_i (y_{hi} - \hat{\bar{Y}}_h)(x_{hi} - \hat{\bar{X}}_h)^\tau S_{xx}^{-1}, \\ S_{xx} &= \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{\bar{X}}_h)(x_{hi} - \hat{\bar{X}}_h)^\tau. \end{aligned}$$

Similarly,

$$\hat{\sigma}_{F_\nu} = [\sum_h W_h^2 \frac{1}{n_h} F_{h,n_h}(y)[1 - F_{h,n_h}(y)] - G_n S_{xx}^{-1} G_n^\tau] \quad (2.16)$$

is a consistent estimator of σ_{F_ν} , the asymptotic variance of $\hat{F}_Y(y)$, where

$$G_n = \sum_h n W_h^2 n_h^2 \sum_i (x_{hi} - \hat{\bar{X}}_h) 1(y_{hi} < y).$$

We can use (2.15) and (2.16) to get normal theory confidence intervals for $\bar{Y}$ and $F_Y(y)$, which are asymptotically correct.

Another consistent estimator of σ is obtained by the jackknife method. We give the result in the following theorem.

Theorem 2. *Under the same conditions as Theorem 1, let $\hat{\bar{Y}}_{-lj}$ be the estimator of $\bar{Y}$ when the j^{th} observation of l^{th} stratum is deleted and let*

$$\hat{\sigma}_J = \sum_{l=1}^H (n_l - 1) \left(\frac{1}{n_l} - \frac{1}{N_l} \right) \sum_j (\hat{\bar{Y}} - \hat{\bar{Y}}_{-lj}) ((\hat{\bar{Y}} - \hat{\bar{Y}}_{-lj})^\tau) \quad (2.17)$$

be the jackknife variance estimator. Then

$$n(\hat{\sigma}_J - \hat{\sigma}) \xrightarrow{P} 0.$$

A jackknife variance estimator for $\hat{F}_Y(y)$ can be obtained from (2.17) by simply changing y_{hi} to $1(y_{hi} \leq y)$. The consistency of this variance estimator follows along the lines of Theorem 2.

2.3 Estimation of Small Area Means

In some applications the strata may be regarded as small areas of interest. The term “small area” is used to denote a small geographical area, such as a county, a municipality or a census division. Sample survey data certainly can be used to derive reliable estimators of totals and means for large areas or domains. However, the usual direct survey estimators for a small area, based on data only from the sample units in the area, are likely to yield unacceptably large standard errors due to the unduly small size of the sample in the area. Sample sizes for small areas are typically small because the overall sample size in a survey is usually determined to provide specific accuracy at a much higher level of aggregation than that of small areas. Ghosh and Rao(1994) reviewed many methods of small area estimation, including those based on models that link the small area through auxiliary information. Here we do not assume any such model.

If the strata can be regarded as small areas and we only know the mean of the population, we can get more efficient estimators of the individual small area

(stratum) means $\bar{Y}_h$ by using auxiliary information in the form of the overall mean, $\bar{X}$. Under our set-up, we can use the MEL method to get MEL estimators for the strata means $\bar{Y}_h$ as

$$\hat{\hat{Y}}_h = \sum_i \hat{p}_{hi} y_{hi}, \quad h = 1, \dots, H.$$

where $\hat{p}_{hi}$'s are the same as those given in section 2.

We now study the efficiency of $\hat{\hat{Y}}_h$ relative to the sample mean $\bar{y}_h$ that uses only the y -data in small area h . We have

$$\begin{aligned} \hat{\hat{Y}}_h &= \sum_i \hat{p}_{hi} y_{hi} \\ &= \bar{y}_h + B_h(\bar{X} - \bar{x}_{st}) + o_p(n^{-\frac{1}{2}}), \end{aligned}$$

where

$$B_h = k_h W_h \text{Cov}(y_h, x_h) \sigma_{xx}^{-1}, \quad \sigma_{xx} = \sum_h k_h W_h^2 \text{Var}(x_h).$$

Hence

$$n \text{Var}(\hat{\hat{Y}}_h) \approx k_h [\text{Var}(y_h) - 2W_h B_h \text{Cov}(y_h, x_h)] + B_h \sigma_{xx} B_h^T.$$

Therefore, the average variance over strata is

$$\begin{aligned} &H^{-1} \sum_h \text{Var}(\hat{\hat{Y}}_h) \\ &\approx H^{-1} \left(\sum_h n_h^{-1} [\text{Var}(y_h) - 2W_h B_h \text{Cov}(y_h, x_h)] + n^{-1} \sum_h B_h \sigma_{xx} B_h^T \right). \end{aligned}$$

We use this as a measure of precision of the MEL estimators $\hat{\hat{Y}}_h$. We consider two special cases. First, if x and y are uncorrelated for all h , then

$$H^{-1} \sum_h \text{Var}(\hat{\hat{Y}}_h) \approx H^{-1} \sum_h n_h^{-1} \text{Var}(y_h) = H^{-1} \sum_h \text{Var}(\bar{y}_h).$$

In this case, there is no gain in efficiency over $\bar{y}_h$. On the other hand, if $y_{hi} = a_h + c_h x_{hi}$ for all h , where x_{hi} is scalar, then

$$\begin{aligned}
 H^{-1} \sum_h \text{Var}(\hat{\bar{Y}}_h) &\approx H^{-1} \left\{ \sum_h n_h^{-1} \left[\text{Var}(y_h) - 2c_h^2 W_h^2 k_h (\text{Var}(x_h))^2 \sigma_{xx}^{-1} \right] \right. \\
 &\quad \left. + \sum_h c_h^2 W_h^2 k_h n_h^{-1} (\text{Var}(x_h))^2 \sigma_{xx}^{-1} \right\} \\
 &= H^{-1} \left[\sum_h n_h^{-1} \text{Var}(y_h) - \sum_h c_h^2 W_h^2 k_h n_h^{-1} (\text{Var}(x_h))^2 \sigma_{xx}^{-1} \right] \\
 &< H^{-1} \sum_h \text{Var}(\bar{y}_h).
 \end{aligned}$$

This suggests that in practice $\hat{\bar{Y}}_h$ will be more efficient than $\bar{y}_h$ if y and x are approximately linearly related in each small area h .

2.4 Empirical Likelihood Ratio Test

Now we turn to the empirical likelihood ratio test (ELRT) on $\bar{Y}$. Here we denote $E(z) = \sum_h W_h E(z_h) = \theta$, where $\theta = (\theta_k^r, \theta_u^r)^r$, $\theta_k = \bar{X}$, $\theta_u = \bar{Y}$, and k, u mean “known”, “unknown” respectively. By (2.1) we know the empirical likelihood function is

$$L(F) = \prod_{h=1}^H \prod_{i=1}^{n_h} F_{h,N_h}(z_{hi}) = \prod_{h=1}^H \prod_{i=1}^{n_h} p_{hi}, \quad (4.1)$$

where $p_{hi} = dF_{h,N_h}(z_{hi}) = Pr(z_h = z_{hi})$ and F denote the population distribution. Noting that $p_{hi} \geq 0$, $\sum_i p_{hi} = 1$ and $F(z) = \sum_h W_h F_{h,N_h}(z)$, we know (4.1) is maximized by $F_n(z) = \sum_h W_h F_{h,n_h}(z)$, where $F_{h,n_h}(z) = \frac{1}{n_h} \sum_i 1(z_{hi} < z)$. The empirical likelihood ratio is then defined as $R(F) = L(F)/L(F_n)$ which reduces to

$$R(F) = \prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi}. \quad (4.2)$$

Since we are interested in the parameter $\theta_u = \bar{Y}$, and we have auxiliary information (2.2), we define the profile empirical likelihood ratio function

$$R_E(\theta) = \sup\{\prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi} \mid p_{hi} \geq 0, \sum_i p_{hi} = 1, \sum_h W_h \sum_i p_{hi} z_{hi} = \theta\}. \quad (4.3)$$

The existence of a unique value for the right-hand side of (4.3) for a given θ_u is the same as discussed in section 2. Using the Lagrange multiplier method, let

$$l = \sum_h \sum_i \log p_{hi} - \sum_h \lambda_h (\sum_i p_{hi} - 1) - nt^\tau (\sum_h W_h \sum_i p_{hi} z_{hi} - \theta)$$

where $t = (t_1, t_2, \dots, t_d)^\tau$ are Lagrange multipliers. Taking derivatives with respect to p_{hi} , we have

$$\frac{\partial l}{\partial p_{hi}} = p_{hi}^{-1} - \lambda_h - n w_h t^\tau z_{hi} = 0, \quad i = 1, \dots, n_h; h = 1, \dots, H.$$

$$\begin{aligned} \sum_i p_{hi} \frac{\partial l}{\partial p_{hi}} &= n_h - \lambda_h - n W_h t^\tau \sum_i p_{hi} z_{hi} = 0 \\ \Rightarrow \lambda_h &= n_h - n W_h t^\tau \sum_i p_{hi} z_{hi}. \end{aligned}$$

Hence,

$$p_{hi} = \{n_h [1 + m_h t^\tau (z_{hi} - \hat{Z}_h)]\}^{-1}, \quad (4.4)$$

where $\hat{Z}_h = \sum_i p_{hi} z_{hi}$, $m_h = n W_h n_h^{-1}$ and t is the solution to

$$\theta = \sum_h W_h \sum_i p_{hi} z_{hi} = \sum_h \frac{W_h}{n_h} \sum_i \frac{1}{1 + m_h t^\tau (z_{hi} - \hat{Z}_h)} z_{hi}.$$

From the discussion of section 2 we know that t can be determined in terms of θ , and $t = t(\theta)$ is actually a continuous differentiable function of θ .

The empirical likelihood ratio function for θ now becomes as

$$L_E(\theta) = \prod_{h=1}^H \prod_{i=1}^{n_h} \left(\frac{1}{n_h} \right) \frac{1}{1 + m_h t^\tau (z_{hi} - \hat{Z}_h)},$$

which leads to the empirical log-likelihood ratio

$$-2 \log L_E(\theta) = l_E(\theta) = 2 \sum_h \sum_i \log[1 + m_h t^\tau (z_{hi} - \hat{Z}_h)]. \quad (4.5)$$

We minimize $l_E(\theta)$ to obtain an estimator $\tilde{\theta}_u$ of the parameter θ_u and $\tilde{\theta} = (\bar{X}^\tau, \tilde{\theta}_u^\tau)^\tau$, called the empirical likelihood ratio estimator. In addition, this yields estimators $\tilde{p}_{hi}$ from (4.4), and an estimator for the distribution function F :

$$\tilde{F}_Y(y) = \sum_i W_h \sum_i \tilde{p}_{hi} 1(y_{hi} < y). \quad (4.6)$$

Now we turn to the calculation of $\tilde{\theta}$ by minimizing of $l_E(\theta)$. From the context above we know that it is equivalent to maximizing

$$L = \prod_h \prod_i p_{hi} \quad (4.7)$$

subject to

$$\sum_i p_{hi} = 1, p_{hi} \geq 0, \quad (4.8)$$

$$\sum_h W_h \sum_i p_{hi} x_{hi} = \theta_k, \quad (4.9)$$

$$\sum_h W_h \sum_i p_{hi} y_{hi} = \theta_u \quad (4.10)$$

where θ_u is unknown. Since θ_u is unknown, (4.10) does not add any information into the estimation problem. Hence, L will be maximized by dropping (4.10). We can show this as follows. Let

$$\begin{aligned} l_1 = & \sum_h \sum_i \log p_{hi} - \sum_h \lambda_h (\sum_i p_{hi} - 1) - nt_k^\tau (\sum_h w_h \sum_i p_{hi} x_{hi} - \theta_k) \\ & - nt_u^\tau (\sum_h W_h \sum_i p_{hi} y_{hi} - \theta_u), \end{aligned}$$

where t_k, t_u are multipliers with the same dimension as θ_k, θ_u respectively. Then

$$\frac{\partial l_1}{\partial \theta_u} = nt_u = 0 \Rightarrow t_u = 0.$$

This shows that we can drop (4.10). Therefore, $\tilde{p}_{hi}$'s are the same as $\hat{p}_{hi}$'s given in section 2.2, and $\tilde{\theta}_u = \hat{Y} = \sum_h W_h \sum_i \tilde{p}_{hi} y_{hi}$. We can prove this by following the steps in section 2.2.

We note here that under $H_0 : \theta_u = \theta_{u_0}$, the estimator for p_{hi} will be changed to

$$\hat{p}_{hi} = \frac{1}{n_h [1 + m_h \hat{t}^\tau(\theta_0)(z_{hi} - \hat{\bar{Z}}_h)]},$$

where $\hat{\bar{Z}}_h = \sum_i \hat{p}_{hi} z_{hi}$, $\theta_0 = (\bar{X}^\tau, \theta_{u_0}^\tau)^\tau$. The calculation of $\hat{p}_{hi}$'s is as discussed in section 2 except for changing x_{hi} to z_{hi} , $\hat{X}_h$ to $\hat{\bar{Z}}_h$, etc.

The empirical likelihood ratio statistic for testing $H_0 : \theta_u = \theta_{u_0}$ is given by

$$\begin{aligned} W_E(\theta_0) &= l_E(\theta_0) - l_E(\tilde{\theta}) \\ &= 2 \sum_h \sum_i \log[1 + m_h \hat{t}^\tau(\theta_0)(z_{hi} - \hat{\bar{Z}}_h)] - \\ &\quad - 2 \sum_h \sum_i \log[1 + m_h \tilde{t}^\tau(\tilde{\theta})(z_{hi} - \tilde{\bar{Z}}_h)], \end{aligned}$$

where $\tilde{\bar{Z}}_h = \sum_i \tilde{p}_{hi} z_{hi}$.

Theorem 3. *Under the conditions of Theorem 1 and $\text{Var}(z_h) \geq \sigma_1 > 0$ for any h and ν , we have*

$$W_E(\theta_0) \xrightarrow{L} \chi^2(p) \quad \text{as } \nu \rightarrow \infty$$

when H_0 is true.

Using Theorem 3, we reject H_0 at level α if $W_E(\theta_0)$ exceeds the upper α -point, $\chi_\alpha^2(p)$, of $\chi^2(p)$.

2.5 Simulation study

We conducted a limited simulation study on the finite sample properties of the proposed MEL estimators of the population mean $\bar{Y}$ and the distribution function $F_Y(y)$ in the scalar case ($p = 1, d = 2$). We also considered the likelihood ratio test on $\bar{Y}$.

For the simulation study, we employed six stratified finite populations considered by Chen and Sitter (1996). Each population consisted of $L = 4$ strata with stratum sizes $N_h = 8000 - 300h$ and stratum sample sizes $n_h = 100 - 9h$ for $h = 1, 2, 3, 4$. The scalar characteristic y_{hi} for the i -th unit within the h -th stratum was generated from the model

$$y_{hi} = \alpha_h + \beta_h x_{hi} + \gamma_h x_{hi}^2 + \xi_h x_{hi}^a \epsilon_{hi} \quad (5.1)$$

for specified values of α_h , β_h , γ_h , a and b_h , where ϵ_{hi} are iid random variables for each h , generated from either a chi-squared distribution with b_h degrees of freedom, i.e., $\chi_{b_h}^2$, or a standard normal distribution, $N(0, 1)$. Further, the scalar x_{hi} are iid variables for each h , generated from $\chi_{3+h^2}^2/2$. Table 2.1 gives the parameter settings for generating the six populations. The strength of the relationship between y and x is weak for population 1, and in populations 2-4, y and x are linearly related, while in populations 5 and 6 a quadratic term is added to create departures from linearity. The populations cover both symmetric and skewed errors ϵ_{hi} .

For each of the parameter settings in Table 2.1, a stratified finite population was created and $R = 1000$ independent stratified simple random samples were drawn. The random numbers needed to generate the finite population were obtained using the NAG fortran library function. The empirical mean squared error of an estimator

$\hat{\theta}$ of a parameter θ was calculated as

$$MSE(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}^{(r)} - \theta)^2,$$

where $\hat{\theta}^{(r)}$ is the value of $\hat{\theta}$ for the r -th simulated sample.

We now study the performance of the MEL estimator $\hat{Y}$ of the mean $\bar{Y}$ relative to the stratified mean, $\bar{y}_{st} = \sum_h W_h \bar{y}_h$, a generalized regression estimator (GREG) and the optimal regression estimator $\hat{Y}_{opt}$. The GREG estimator is given by

$$\hat{Y}_{GREG} = \bar{y}_{st} + \hat{B}_G(\bar{X} - \bar{x}_{st}),$$

where

$$\hat{B}_G = \frac{\sum_h (W_h/n_h) \sum_i (y_{hi} - \bar{y}_{st})(x_{hi} - \bar{x}_{st})}{\sum_h (W_h/n_h) \sum_i (x_{hi} - \bar{x}_{st})^2}.$$

This estimator is motivated by a linear regression model with common intercept and slope coefficients and constant error variance (Sarndal et al., 1991, p.229). Table 2.2 reports the ratios of $MSE(\hat{Y})$ to $MSE(\hat{Y}_{GREG})$, $MSE(\bar{y}_{st})$ and $MSE(\hat{Y}_{opt})$ for all the six populations. It is clear from Table 2.2 that MELE outperforms the GREG estimator, especially for populations 1-4 with widely varying intercept terms across strata. It is also considerably more efficient than the stratified mean $\bar{y}_{st}$, except in the case of population 1 with a weak relation between y and x . As expected, MELE and the optimal estimator are equally efficient for all the six populations.

The second part of our simulation study deals with the estimation of population distribution function $F_Y(y)$. We selected the values of $F_Y(y)$ corresponding to the quartiles $y_p (p = \frac{1}{4}, \frac{2}{4}, \frac{3}{4})$, i.e., $F_Y(y_p) = p$. Table 2.3 reports the ratios of $MSE(\hat{F}_Y(y))$ to $MSE(\hat{F}_{st}(y))$ and $MSE(\hat{F}_{opt}(y))$, where $\hat{F}_Y(y)$ is the MELE of $F_Y(y)$, $\hat{F}_{st}(y) = \sum_h W_h F_{h,n_h}(y)$ is the empirical distribution function and $\hat{F}_{opt}(y)$ is

the optimal regression estimator of $F_Y(y)$. It is clear from Table 2.3 that the MEL estimator $\hat{F}_Y(y)$ outperforms the empirical distribution function $\hat{F}_{st}(y)$ in all the six populations. Although $\hat{F}_Y(y)$ and $\hat{F}_{opt}(y)$ are asymptotically equivalent, our simulation results indicate that $\hat{F}_Y(y)$ is often significantly more efficient than $\hat{F}_{opt}(y)$ in finite samples. For example, the ratio equals 0.62 for population 2 and 4 at $p = \frac{1}{2}$ and 0.7 for population 6 at $p = \frac{1}{4}$. The main advantage of $\hat{F}_Y(y)$, however, is that it is monotone, unlike $\hat{F}_{opt}(y)$.

The final part of our simulation study investigates the performance of the empirical likelihood ratio test and the asymptotic normality test obtained from the result of theorem 1 (called ANT). Since we know the means of the six populations, say, $\bar{Y}_0$, we can test the hypothesis $H_0 : \bar{Y} = \bar{Y}_0$ from the simulated samples. We tested H_0 using the empirical likelihood ratio test statistic $W_E(\bar{Y}_0)$ at nominal levels $\alpha = 0.05$ and 0.1. Table 2.4 reports the empirical type I error rates for $\alpha = 0.10$ and 0.05, obtained by computing the proportion of samples for which $W_E(\bar{Y}_0) > \chi^2_\alpha(1)$, the upper α -point of a χ^2 variable with 1 degree of freedom. Table 2.5 reports the empirical error rates for $\alpha = 0.10$ and 0.05, obtained by computing the proportion of samples for which $|\hat{\bar{Y}} - \bar{Y}| > u_\alpha \hat{\sigma}_\nu^{-\frac{1}{2}}$, where u_α is the upper α -point of a $N(0, 1)$ distribution. Table 2.4 shows that the empirical likelihood ratio test performs well over the six populations, although somewhat conservative. It appears that higher-order corrections to the empirical likelihood ratio test are needed to get more accurate results, as noted by Qin and Lawless(1995) in the context of simple random sampling. Table 2.5 shows that the standard normal approximation performs well over all six populations.

Table 2.1: Parameter settings for generated finite populations

Population	h	α_h	β_h	γ_h	ξ_h	a	ϵ_{hi}
I	1	2	0	0	0.2	-0.5	χ_3^2
	2	6	0	0	0.2	-0.5	χ_4^2
	3	10	0	0	0.2	-0.5	χ_5^2
	4	14	0	0	0.2	-0.5	χ_6^2
II	1	2	0.5	0	0.2	-0.5	χ_3^2
	2	6	1.0	0	0.2	-0.5	χ_4^2
	3	10	-0.5	0	0.2	-0.5	χ_5^2
	4	14	1.0	0	0.2	-0.5	χ_6^2
III	1	2	0.5	0	0.1	0.5	χ_3^2
	2	6	1.0	0	0.1	0.5	χ_4^2
	3	10	-0.5	0	0.2	0.5	χ_5^2
	4	14	1.0	0	0.1	0.5	χ_6^2
IV	1	2	0.5	0	0.1	0.5	$N(0, 1)$
	2	6	1.0	0	0.2	0.5	$N(0, 1)$
	3	10	-0.5	0	0.2	0.5	$N(0, 1)$
	4	14	1.0	0	0.2	0.5	$N(0, 1)$
V	1	2	0.5	0.05	0.1	0.5	$N(0, 1)$
	2	6	1.0	-0.05	0.2	0.5	$N(0, 1)$
	3	10	-0.5	0.05	0.2	0.5	$N(0, 1)$
	4	14	1.0	-0.05	0.2	0.5	$N(0, 1)$
IV	1	4	0.5	0.05	0.1	0.5	$N(0, 1)$
	2	6	1.0	-0.05	0.2	0.5	$N(0, 1)$
	3	8	-0.5	0.05	0.2	0.5	$N(0, 1)$
	4	10	1.0	-0.05	0.2	0.5	$N(0, 1)$

Table 2.2: Ratios of MSE 's of the $\hat{Y}$ to $\hat{Y}_{GREG}$, $\bar{y}_{st}$, $\hat{Y}_{OPT}$

Population	$\hat{Y}$ to $\hat{Y}_{GREG}$	$\hat{Y}$ to $\bar{y}_{st}$	$\hat{Y}$ to $\hat{Y}_{OPT}$
1	0.0247	0.938	1.001
2	0.474	0.786	0.999
3	0.445	0.888	0.994
4	0.434	0.791	0.996
5	0.629	0.604	1.009
6	0.818	0.556	1.006

Table 2.3: Ratios of MSE 's of $\hat{F}(y_p)$ to $\hat{F}_{st}(y_p)$ and $\hat{F}_{opt}(y_p)$

Population	Ratio	$p = 1/4$	$p = 2/4$	$p = 3/4$
1	$\hat{F}(y_p)$ to $\hat{F}_{st}(y_p)$	0.497	0.344	0.374
	$\hat{F}(y_p)$ to $\hat{F}_{opt}(y_p)$	0.996	0.999	0.991
2	$\hat{F}(y_p)$ to $\hat{F}_{st}(y_p)$	0.944	0.472	0.920
	$\hat{F}(y_p)$ to $\hat{F}_{opt}(y_p)$	0.933	0.618	0.982
3	$\hat{F}(y_p)$ to $\hat{F}_{st}(y_p)$	0.999	0.699	0.979
	$\hat{F}(y_p)$ to $\hat{F}_{opt}(y_p)$	0.975	0.787	0.989
4	$\hat{F}(y_p)$ to $\hat{F}_{st}(y_p)$	0.960	0.485	0.947
	$\hat{F}(y_p)$ to $\hat{F}_{opt}(y_p)$	0.938	0.625	0.993
5	$\hat{F}(y_p)$ to $\hat{F}_{st}(y_p)$	0.486	0.857	0.982
	$\hat{F}(y_p)$ to $\hat{F}_{opt}(y_p)$	0.755	0.901	0.967
6	$\hat{F}(y_p)$ to $\hat{F}_{st}(y_p)$	0.349	0.881	0.933
	$\hat{F}(y_p)$ to $\hat{F}_{opt}(y_p)$	0.708	0.983	0.969

Table 2.4: Empirical Error Rate(%) for MELRT

Population	$\alpha = 10\%$	$\alpha = 5\%$
1	7.0	2.2
2	8.1	3.2
3	9.2	3.3
4	9.4	3.4
5	8.6	1.5
6	9.1	2.4

Table 2.5: Empirical Error Rate(%) for ANT

Population	$\alpha = 10\%$	$\alpha = 5\%$
1	9.9	4.1
2	9.5	4.7
3	10.1	5.1
4	8.8	4.8
5	9.3	4.5
6	9.0	5.2

2.6 Proofs

We first summarize the existence conclusion of maximizing problem at the beginning of section 2 into a lemma and then prove it.

Lemma 1 *Under conditions of Theorem 1, there exists with probability tending to one, as ν goes to infinity, a solution to the problem of maximizing $\sum_h \sum_i \log p_{hi}$ subject to*

$$\sum_i p_{hi} = 1 \ (h = 1, \dots, H, \ p_{hi} \geq 0), \ \sum_h W_h \sum_i p_{hi} x_{hi} = \bar{X}.$$

Proof of Lemma 1

In order to simplify our discussion, we suppose that x is a scalar here. The discussion is similar when x is multi-dimensional.

Denote

$$p' = (p'_{11}, \dots, p'_{1n_1}; \dots; p'_{H1}, \dots, p'_{Hn_H})^T.$$

$$D = \{p' | p'_{hi} \geq 0, \sum_h \sum_i p'_{hi} = 1, \sum_h \sum_i p'_{hi} x_{hi} = \bar{X}\},$$

and

$$D_1 = \{p' | p'_{hi} \geq 0, \sum_i p'_{hi} = W_h, \sum_h \sum_i p'_{hi} x_{hi} = \bar{X}\},$$

then obviously D_1 is a convex subset of D . As shown by Owen (1990), a unique solution exists for maximizing $\sum_h \sum_i \log p'_{hi}$ within D , provided that $\bar{X}$ is within the convex hull of $x_{11}, \dots, x_{Hn_H}$. Since $D_1 \subset D$, and if we can show that D_1 is not empty, we can get a maximum for $\sum_h \sum_i \log p'_{hi}$ within D_1 , which is equivalent to maximizing problem in the lemma since $p'_{hi} = W_h p_{hi}$.

Under the conditions of Theorem 1, it is obvious that

$$P(\bar{X} \in (x_{(1)}, x_{(n)})) \rightarrow 1 \text{ as } \nu \rightarrow \infty.$$

where $x_{(1)} = \min_h \min_i x_{hi}$, $x_{(n)} = \max_h \max_i x_{hi}$. Since H is finite, and samples are independent across strata, we can also conclude that

$$P(\bar{X}_1 \in (x_{1(1)}, x_{1(n_1)}), \dots, \bar{X}_H \in (x_{H(1)}, x_{H(n_H)})) \rightarrow 1 \text{ as } \nu \rightarrow \infty.$$

where $x_{h(1)} = \min_i x_{hi}$, $x_{h(n_h)} = \max_i x_{hi}$, $h = 1, \dots, H$. Suppose $\bar{X}_h \in (x_{h(1)}, x_{h(n_h)})$, $h = 1, \dots, H$, then we have solution $(p_{h1}^0, \dots, p_{hn_h}^0) (\sum_i p_{hi}^0 = 1)$, such that $\sum_i p_{hi}^0 x_{hi} = \bar{X}_h$. Since $\sum_h W_h \bar{X}_h = \bar{X}$, it follows that

$$p^0 = (W_1 p_{11}^0, \dots, W_1 p_{1n_1}^0; \dots; W_H p_{H1}^0, \dots, W_H p_{Hn_H}^0)^\tau \in D_1,$$

The conditions $\bar{X}_h \in (x_{h(1)}, x_{h(n_h)})$ ($h = 1, \dots, H$) hold with probability tending to 1, hence D_1 is not empty with probability tending to 1. This completes the proof.

We need the following Lemma to prove Theorem 1.

Lemma 2 *Under conditions of Theorem 1, we have*

$$\hat{\psi}_n = \hat{\psi}_n(\bar{X}) = S_{xx}^{-1}(\bar{x}_{st} - \bar{X}) + o_p(A^{\frac{1}{2}}) = O_p(A^{\frac{1}{2}}), \quad (6.1)$$

where

$$A = \sum_h W_h^2 (n_h^{-1} - N_h^{-1}), \quad S_{xx} = \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau,$$

$$\bar{x}_{st} = \sum_h W_h \bar{x}_h, \quad \hat{X}_h = \sum_i \hat{p}_{hi} x_{hi},$$

$$\hat{p}_{hi} = \{n_h [1 + m_h \hat{\psi}_n^\tau (x_{hi} - \hat{X}_h)]\}^{-1}.$$

Proof of Lemma 2

By using the decomposition

$$x_{hi} = x_{hi}[1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)] - (x_{hi} - \hat{X}_h + \hat{X}_h)m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)$$

and noting (2.13) we have

$$\begin{aligned} \hat{X}_h &= \sum_i \hat{p}_{hi} x_{hi} \\ &= \sum_i \frac{1}{n_h [1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)]} x_{hi} \\ &= \bar{x}_h - n W_h n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} \hat{\psi}_n. \end{aligned} \quad (6.2)$$

Since $\sum_h W_h \hat{X}_h = \bar{X}$, we have:

$$\begin{aligned} \bar{X} &= \sum_h W_h \bar{x}_h - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} \hat{\psi}_n \\ &= \bar{x}_{st} - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} \hat{\psi}_n, \end{aligned} \quad (6.3)$$

or

$$\bar{x}_{st} - \bar{X} = \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} \hat{\psi}_n. \quad (6.4)$$

From (6.2) and (6.4) we get

$$\begin{aligned} W_h^2 \| \bar{x}_h - \hat{X}_h \|^2 &= \hat{\psi}_n^\tau [n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)}]^2 \hat{\psi}_n \\ &\leq \hat{\psi}_n^\tau [\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)}]^2 \hat{\psi}_n \\ &= \| \bar{x}_{st} - \bar{X} \|^2 \xrightarrow{P} 0, \end{aligned} \quad (6.5)$$

where $\| \bar{x}_{st} - \bar{X} \|^2 \xrightarrow{P} 0$ follows from standard results for stratified random sampling theory (e.g., Bickel and Freeman, 1984). Hence

$$S_{xx} = \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \bar{x}_h + \bar{x}_h - \hat{X}_h)(x_{hi} - \bar{x}_h + \bar{x}_h - \hat{X}_h)^\tau$$

$$\begin{aligned}
&= \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \bar{x}_h)(x_{hi} - \bar{x}_h)^\tau \\
&\quad + \sum_h n W_h^2 n_h^{-1} (\bar{x}_h - \hat{X}_h)(\bar{x}_h - \hat{X}_h)^\tau \\
&\xrightarrow{P} \sum_h W_h^2 k_h \sigma_{h,xx}.
\end{aligned} \tag{6.6}$$

Therefore,

$$S_{xx} = \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau = O_p(1). \tag{6.7}$$

From (6.4) we have

$$\begin{aligned}
\| \bar{x}_{st} - \bar{X} \|^2 &= \hat{\psi}_n^\tau \left[\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau (x_{hi} - \hat{X}_h)} \right]^2 \hat{\psi}_n \\
&\geq \hat{\psi}_n^\tau \left[\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \| \hat{\psi}_n \| \| x_{hi} - \hat{X}_h \|} \right]^2 \hat{\psi}_n \\
&\geq \hat{\psi}_n^\tau \left[\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \| \hat{\psi}_n \| 2x_{h,n_h}^*} \right]^2 \hat{\psi}_n
\end{aligned} \tag{6.8}$$

where $x_{h,n_h}^* = \max_i \| x_{hi} \|$. Similar to the proof of Owen's (1988) Lemma, we can show that $x_{h,n_h}^* = o(n_h^{\frac{1}{2}})$ almost surely (a.s.). Hence

$$\max_h (2W_h n_h^{-1} x_{h,n_h}^*) = \max_h o(n_h^{-\frac{1}{2}}) = o(n_{(1)}^{-\frac{1}{2}}) = o(n^{-\frac{1}{2}}) (a.s.), \tag{6.9}$$

where $n_{(1)} = \min\{n_1, \dots, n_H\}$. By (6.8) we get

$$\begin{aligned}
\| \bar{x}_{st} - \bar{X} \|^2 &\geq \hat{\psi}_n^\tau \left[\frac{1}{1 + \| \hat{\psi}_n \| o(n^{\frac{1}{2}})} \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau \right]^2 \hat{\psi}_n \\
&\geq \hat{\psi}_n^\tau \left[\frac{1}{1 + \| \hat{\psi}_n \| o(n^{\frac{1}{2}})} \right]^2 \left[\sum_h k_h W_h^2 \sigma_1 / 2 \right]^2 \hat{\psi}_n \quad (a.s.) \\
&\geq \left[\frac{\lambda}{1 + \| \hat{\psi}_n \| o(n^{\frac{1}{2}})} \right]^2 \| \hat{\psi}_n \|^2 \quad (a.s.),
\end{aligned}$$

where $\lambda = \sum_h k_h W_h^2 \sigma_{(1)} / 2 > 0$, $\sigma_{(1)}$ is the smallest eigenvalue of σ_1 . Hence,

$$\begin{aligned}
\| \hat{\psi}_n \| &\leq \| \bar{x}_{st} - \bar{X} \| [1 + \| \hat{\psi}_n \| o(n^{\frac{1}{2}})] / \lambda \quad (a.s.) \\
&= \lambda^{-1} (\| \bar{x}_{st} - \bar{X} \| + \| \bar{x}_{st} - \bar{X} \| \| \hat{\psi}_n \| o(n^{\frac{1}{2}})) \quad (a.s.).
\end{aligned} \tag{6.10}$$

Since

$$Pr\{\|\bar{x}_{st} - \bar{X}\| \geq cA^{\frac{1}{2}}\} \leq (cA^{\frac{1}{2}})^{-2} \sum_h W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) tr(\sigma_{h,xx}) \leq c^{-2} max_h tr(\sigma_{h,xx}),$$

we know $\|\bar{x}_{st} - \bar{X}\| = O_p(A^{\frac{1}{2}})$ and $\|\bar{x}_{st} - \bar{X}\| o_p(n^{\frac{1}{2}}) = o_p(1)$. Therefore, (6.10) becomes $\|\hat{\psi}_n\| \leq \lambda^{-1}(\|\bar{x}_{st} - \bar{X}\| + \|\hat{\psi}_n\| o_p(1))$, i.e., $\|\hat{\psi}_n\| = O_p(A^{\frac{1}{2}})$, which establishes the last part of (6.1). From (6.4) we have

$$\begin{aligned} \bar{x}_{st} - \bar{X} &= [\sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^T] \hat{\psi}_n \\ &\quad - [\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^T}{1 + m_h \hat{\psi}_n^T (x_{hi} - \hat{X}_h)} m_h \hat{\psi}_n^T (x_{hi} - \hat{X}_h)] \hat{\psi}_n, \end{aligned}$$

and

$$\begin{aligned} &\|\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^T m_h \hat{\psi}_n^T (x_{hi} - \hat{X}_h)}{1 + m_h \hat{\psi}_n^T (x_{hi} - \hat{X}_h)} \hat{\psi}_n\| \\ &\leq \|\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^T |m_h \hat{\psi}_n^T (x_{hi} - \hat{X}_h)|}{1 + m_h \hat{\psi}_n^T (x_{hi} - \hat{X}_h)} \hat{\psi}_n\| \\ &\leq max_h max_i |m_h \hat{\psi}_n^T (x_{hi} - \hat{X}_h)| \\ &\quad \cdot \|\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^T}{1 + m_h \hat{\psi}_n^T (x_{hi} - \hat{X}_h)}\| \\ &= max_h o_p(n_h^{\frac{1}{2}}) \|\hat{\psi}_n\| \|\bar{x}_{st} - \bar{X}\| \\ &= o_p(1) O_p(A^{\frac{1}{2}}) \\ &= o_p(A^{\frac{1}{2}}). \end{aligned}$$

Therefore,

$$\bar{x}_{st} - \bar{X} = [\sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^T] \hat{\psi}_n + o_p(A^{\frac{1}{2}}),$$

or

$$\hat{\psi}_n = S_{xx}^{-1}(\bar{x}_{st} - \bar{X}) + o_p(A^{\frac{1}{2}}). \square$$

Proof of theorem 1

We have

$$\begin{aligned}
\hat{Y}_n &= \sum_h W_h \sum_i \hat{p}_{hi} y_{hi} \\
&= \sum_h W_h \sum_i \frac{1}{n_h [1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)]} y_{hi} \\
&= \sum_h \frac{W_h}{n_h} \sum_i y_{hi} - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{\hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} y_{hi} \\
&= \bar{y}_{st} - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(y_{hi} - \hat{Y}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} \hat{\psi}_n \\
&= \bar{y}_{st} - \sum_h n W_h^2 n_h^{-2} \sum_i (y_{hi} - \hat{Y}_h)(x_{hi} - \hat{X}_h)^\tau \hat{\psi}_n \\
&\quad + \sum_h n^2 W_h^3 n_h^{-3} \sum_i \frac{\hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau \hat{\psi}_n}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} (y_{hi} - \hat{Y}_h),
\end{aligned}$$

where the last term, noting (6.9), is

$$\begin{aligned}
&\| \sum_h n^2 W_h^3 n_h^{-3} \sum_i \frac{\hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau \hat{\psi}_n}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} (y_{hi} - \hat{Y}_h) \| \\
&\leq \sum_h n^2 W_h^3 n_h^{-3} \sum_i \frac{\hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau \hat{\psi}_n}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} \| y_{hi} - \hat{Y}_h \| \\
&\leq o_p(n^{-\frac{1}{2}}) n \| \hat{\psi}_n^\tau \| \| \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)} \hat{\psi}_n \| \\
&= o_p(1) \| \bar{x}_{st} - \bar{X} \| \\
&= o_p(A^{\frac{1}{2}}).
\end{aligned}$$

Hence, from (6.1) and (6.7) we get

$$\hat{Y}_n = \bar{y}_{st} - \left[\sum_h n W_h^2 n_h^{-2} \sum_i (y_{hi} - \hat{Y}_h)(x_{hi} - \hat{X}_h)^\tau \right].$$

$$\begin{aligned}
& \cdot S_{xx}^{-1}(\bar{x}_{st} - \bar{X}) + o_p(A^{\frac{1}{2}}) \\
&= \bar{y}_{st} - \sum_h k_h W_h^2 \text{Cov}(y_h, x_h) \left[\sum_h k_h W_h^2 \text{Cov}(x_h, x_h) \right]^{-1} (\bar{x}_{st} - \bar{X}) + o_p(A^{\frac{1}{2}}) \\
&= \bar{y}_{st} - B(\bar{x}_{st} - \bar{X}) + o_p(A^{\frac{1}{2}}).
\end{aligned}$$

Since $E[\bar{y}_{st} - B(\bar{x}_{st} - \bar{X})] = \bar{Y}$ and

$$\begin{aligned}
& \text{Var}(\bar{y}_{st} - B(\bar{x}_{st} - \bar{X})) \\
&= \sum_h W_h^2 \frac{n_h^{-1} - N_h^{-1}}{N_h - 1} \sum_{i=1}^{N_h} [y_{hi} - \bar{Y}_h - B(x_{hi} - \bar{X}_h)][y_{hi} - \bar{Y}_h - B(x_{hi} - \bar{X}_h)]^\tau \\
&= \sum_h W_h^2 (n_h^{-1} - N_h^{-1}) [\text{Var}(y_h) - 2B \text{Cov}^\tau(y_h, x_h) + B \text{Var}(x_h) B^\tau] \\
&= \sigma_\nu,
\end{aligned}$$

by appealing to Bickel & Freedman (1984), we get

$$\sigma_\nu^{-\frac{1}{2}}(\hat{\bar{Y}}_n - \bar{Y}) \xrightarrow{L} N_p(0, I_p).$$

Now we turn to the second result of theorem 1. Denoting

$$F_{h,n_h}(y) = \frac{1}{n_h} \sum_i 1(y_{hi} < y),$$

we have

$$\begin{aligned}
\hat{F}_Y(y) &= \sum_h W_h \sum_i \frac{1}{n_h [1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X})]} 1(y_{hi} < y) \\
&= \sum_h W_h n_h^{-1} \sum_i 1(y_{hi} < y) \\
&\quad - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{\hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X})} 1(y_{hi} < y) \\
&= \sum_h W_h F_{h,n_h}(y) - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{\hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X})} 1(y_{hi} < y) \\
&= \sum_h W_h F_{h,n_h}(y) - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau(x_{hi} - \hat{X})} 1(y_{hi} < y) \hat{\psi}_n
\end{aligned}$$

$$\begin{aligned}
&= \sum_h W_h F_{h,n_h}(y) - \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{X}_h)^\tau 1(y_{hi} < y) \hat{\psi}_n \\
&\quad + \hat{\psi}_n^\tau \sum_h n^2 W_h^3 n_h^{-3} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau (x_{hi} - \hat{X}_h)} 1(y_{hi} < y) \hat{\psi}_n,
\end{aligned}$$

where by (6.4) the third term is

$$\begin{aligned}
&|\hat{\psi}_n^\tau \sum_h n^2 W_h^3 n_h^{-3} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau (x_{hi} - \hat{X}_h)} 1(y_{hi} < y) \hat{\psi}_n| \\
&\leq \frac{n}{n_{(1)}} \hat{\psi}_n^\tau \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau}{1 + m_h \hat{\psi}_n^\tau (x_{hi} - \hat{X}_h)} \hat{\psi}_n \\
&= \frac{n}{n_{(1)}} \hat{\psi}_n^\tau (\bar{x}_{st} - \bar{X}) \\
&= o_p(A^{\frac{1}{2}}).
\end{aligned}$$

Therefore

$$\begin{aligned}
\hat{F}_Y(y) &= \sum_h W_h F_{h,n_h}(y) - \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{X}_h)^\tau 1(y_{hi} < y) \hat{\psi}_n + o_p(A^{\frac{1}{2}}) \\
&= \sum_h W_h F_{h,n_h}(y) - G'_n \hat{\psi}_n + o_p(A^{\frac{1}{2}}) \\
&= \sum_h W_h F_{h,n_h}(y) - G'_n S_{xx}^{-1} (\bar{x}_{st} - \bar{X}) + o_p(A^{\frac{1}{2}}) \\
&= \sum_h W_h F_{h,n_h}(y) - G' \sigma_{xx}^{-1} (\bar{x}_{st} - \bar{X}) + o_p(A^{\frac{1}{2}}),
\end{aligned}$$

where

$$\begin{aligned}
G'_n S_{xx}^{-1} &= \sum_h n W_h^2 n_h^{-2} \sum_i (x_{hi} - \hat{X}_h)^\tau 1(y_{hi} < y) S_{xx}^{-1} \\
&\xrightarrow{P} \sum_h W_h^2 k_h E[(x_h - \bar{X}_h) 1(y_h < y)]^\tau \sigma_{xx}^{-1} \\
&= \sum_h W_h^2 k_h \text{Cov}(1(y_h < y), x_h^\tau) \sigma_{xx}^{-1} = G.
\end{aligned}$$

Since $E[\sum_h W_h F_{h,n_h}(y) - G(\bar{x}_{st} - \bar{X})] = F_{h,N_h}(y)$,

$$\text{Var}\left(\sum_h W_h F_{h,n_h}(y) - G(\bar{x}_{st} - \bar{X})\right)$$

$$\begin{aligned}
&= \sum_h W_h^2 (n_h^{-1} - N_h^{-1}) \frac{1}{N_h - 1} \sum_{i=1}^{N_h} [1(y_{hi} < y) - F_{h,N_h}(y) - G(x_{hi} - \bar{X}_h)] \\
&\quad \cdot [1(y_{hi} < y) - F_h(y) - G(x_{hi} - \bar{X}_h)]^\tau \\
&= \sum_h W_h^2 (n_h^{-1} - N_h^{-1}) [F_{h,N_h}(1 - F_{h,N_h}) \\
&\quad - 2GCov^\tau(1(y_h < y), x_h) + GVar(x_h)G^\tau] \\
&= V_\nu
\end{aligned}$$

Hence,

$$\sigma_{F_\nu}^{-\frac{1}{2}}(\hat{F}_Y(y) - F_Y(y)) \rightarrow N_p(0, I_p). \square$$

Proof of theorem 2

Denote $\nu_{hi} = nW_h\hat{\psi}_n^\tau(x_{hi} - \hat{X}_h)$,

$$\nu_{hi,-lj} = (n-1)W_h\hat{\psi}_{-lj}^\tau(x_{hi} - \hat{X}_{h,-lj})$$

for convenience, where $\hat{\psi}_{-lj}$ is the Lagrange multiplier ψ when j -th observation of l -th stratum is deleted. In the following $o_p(1)$ means it holds uniformly over h, i, l, j .

From Lemma 2.1 it can be shown that the following hold uniformly over h, i, l, j :

$$\begin{aligned}
\hat{p}_{hi} &= \frac{1}{n_h + \nu_{hi}} = \frac{1}{n_h}(1 + o_p(1)), \\
\hat{p}_{hi,-lj} &= \frac{1}{n_h + \nu_{hi,-lj}} = \frac{1}{n_h}(1 + o_p(1)), \quad h \neq l, h = 1, \dots, H, \\
\hat{p}_{li,-lj} &= \frac{1}{n_l - 1 + \nu_{li,-lj}} = \frac{1}{n_l - 1}(1 + o_p(1)), \\
\hat{p}_{li} \cdot \hat{p}_{li,-lj} &= \frac{1}{n_l(n_l - 1)}(1 + o_p(1)), \\
\hat{p}_{hi,-lj} \cdot \hat{p}_{hi} &= \frac{1}{n_h^2}(1 + o_p(1)), \quad h \neq l, i \neq j.
\end{aligned}$$

Though MELE for p_{lj} is 0 when we delete x_{lj} from the data, we redefine $\hat{p}_{lj, -lj}$ as above for convenience. We have

$$\begin{aligned}
& \hat{X}_l - \hat{X}_{l, -lj} \\
&= \sum_i (\hat{p}_{li} - \hat{p}_{li, -lj}) x_{li} + \hat{p}_{lj, -lj} x_{lj} \\
&= \sum_i (\hat{p}_{li} - \hat{p}_{li, -lj}) (x_{li} - \hat{X}_l) + \hat{p}_{lj, -lj} (x_{lj} - \hat{X}_l) \\
&= \left[\frac{1}{n_l + \nu_{li}} - \frac{1}{n_l - 1 + \nu_{li, -lj}} \right] (x_{li} - \hat{X}_l) + \frac{1}{n_l - 1} (x_{lj} - \hat{X}_l) (1 + o_p(1)) \\
&= -\frac{1}{(n_l + \nu_{li})(n_l - 1 + \nu_{li, -lj})} (x_{li} - \hat{X}_l) \\
&\quad + \sum_i \frac{\nu_{li, -lj} - \nu_{li}}{(n_l + \nu_{li})(n_l - 1 + \nu_{li, -lj})} (x_{li} - \hat{X}_l) + \frac{1}{n_l - 1} (x_{lj} - \hat{X}_l) + o_p\left(\frac{1}{n}\right) \\
&= -\sum_i \frac{1}{n_l(n_l - 1)} (1 + o_p(1)) (x_{li} - \hat{X}_l) \\
&\quad + W_l \sum_i \frac{n[\hat{\psi}_{-lj}^\tau(x_{li} - \hat{X}_{l, -lj}) - \hat{\psi}_n^\tau(x_{li} - \hat{X}_l)]}{(n_l + \nu_{li})(n_l - 1 + \nu_{li, -lj})} (x_{li} - \hat{X}_l) \\
&\quad - W_l \sum_i \frac{(x_{li} - \hat{X}_l)(x_{li} - \hat{X}_{l, -lj})^\tau}{(n_l + \nu_{li})(n_l - 1 + \nu_{li, -lj})} \hat{\psi}_{-lj} + \frac{1}{n_l - 1} (x_{lj} - \hat{X}_l) + o_p\left(\frac{1}{n}\right) \\
&= -\frac{1}{n_l - 1} (\bar{x}_l - \hat{X}_l) + W_l \sum_i \frac{n(x_{li} - \hat{X}_l)(x_{li} - \hat{X}_l)^\tau (\hat{\psi}_{-lj} - \hat{\psi})}{(n_l + \nu_{li})(n_l - 1 + \nu_{li, -lj})} \\
&\quad + W_l \sum_i \frac{n(x_{li} - \hat{X}_l)(\hat{X}_l - \hat{X}_{l, -lj})^\tau}{[n_l + \nu_{li}][n_l - 1 + \nu_{li, -lj}]} \hat{\psi}_{-lj} \\
&\quad - W_l \sum_i \frac{(x_{li} - \hat{X}_l)(x_{li} - \hat{X}_{l, -lj})^\tau}{n_l(n_l - 1)} \hat{\psi}_{-lj} (1 + o_p(1)) + \frac{1}{n_l - 1} (x_{lj} - \hat{X}_l) + o_p\left(\frac{1}{n}\right) \\
&= \frac{1}{n_l - 1} (x_{lj} - \bar{x}_l) + W_l \sum_i \frac{n}{n_l(n_l - 1)} (x_{li} - \hat{X}_l)(x_{li} - \hat{X}_l)^\tau (\hat{\psi}_{-lj} - \hat{\psi}) (1 + o_p(1)) \\
&\quad + o_p\left(\frac{1}{n}\right) \\
&= \frac{1}{n_l - 1} (x_{lj} - \bar{x}_l) + W_l \sum_i \frac{n}{n_l(n_l - 1)} (x_{li} - \hat{X}_l)(x_{li} - \hat{X}_l)^\tau (\hat{\psi}_{-lj} - \hat{\psi}) + o_p\left(\frac{1}{n}\right)
\end{aligned}$$

$$= \frac{1}{n_l - 1}(x_{lj} - \bar{x}_l) + W_l \sum_i \frac{n}{n_l^2} (x_{li} - \hat{X}_l)(x_{li} - \hat{X}_l)^\tau (\hat{\psi}_{-lj} - \hat{\psi}) + o_p\left(\frac{1}{n}\right), \quad (6.11)$$

where the third term in the sixth equality sign has the order

$$\begin{aligned} & \left\| W_l \sum_i \frac{n(x_{li} - \hat{X}_l)(\hat{X}_l - \hat{X}_{l,-lj})^\tau}{(n_l + \nu_{li})(n_l - 1 + \nu_{li,-lj})} \hat{\psi}_{-lj} \right\| \\ & \leq \left\| \hat{\psi}_{-lj} \right\| \left\| \hat{X}_l - \hat{X}_{l,-lj} \right\| \cdot \left\| \sum_i \frac{n(x_{li} - \hat{X}_l)}{(n_l + \nu_{li})(n_l - 1 + \nu_{li,-lj})} \right\| \\ & \leq O_p(1) \left\| \sum_i \frac{n(x_{li} - \hat{X}_l)}{n_l + \nu_{li}} \cdot \frac{1}{n_l - 1} (1 + o_p(1)) \right\| \\ & = O_p(1) o_p\left(\frac{1}{n}\right) \\ & = o_p\left(\frac{1}{n}\right). \end{aligned}$$

The fourth term of the same equation has also the same order.

Using the same method we can prove for $h \neq l$ that

$$\hat{X}_h - \hat{X}_{h,-lj} = W_h \sum_i \frac{n}{n_l^2} (x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau (\hat{\psi}_{-lj} - \hat{\psi}) + o_p\left(\frac{1}{n}\right). \quad (6.12)$$

Since

$$\sum_h W_h \hat{X}_h = \bar{X}, \quad \sum_h W_h \hat{X}_{h,-lj} = \bar{X},$$

by (6.11) and (6.12) we have

$$\begin{aligned} 0 &= \sum_h W_h (\hat{X}_h - \hat{X}_{h,-lj}) \\ &= \frac{W_l}{n_l - 1} (x_{lj} - \bar{x}_l) + \sum_h W_h^2 \sum_i \frac{n}{n_h^2} (x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau (\hat{\psi}_{-lj}^\tau - \hat{\psi}) + o_p\left(\frac{1}{n}\right). \end{aligned}$$

Therefore

$$\hat{\psi} - \hat{\psi}_{-lj}^\tau = \left[\sum_h W_h^2 \sum_i \frac{n}{n_h^2} (x_{hi} - \hat{X}_h)(x_{hi} - \hat{X}_h)^\tau \right]^{-1} \frac{W_l}{n_l - 1} (x_{lj} - \bar{x}_l) + o_p\left(\frac{1}{n}\right). \quad (6.13)$$

Following the steps of (6.11) we can obtain similar results on $\hat{Y}_h - \hat{Y}_{h,-lj}$. We have

$$\begin{aligned}
& \hat{Y} - \hat{Y}_{-lj} \\
&= \sum_h W_h (\hat{Y}_h - \hat{Y}_{h,-hj}) \\
&= \sum_h n W_h^2 n_h^{-2} \sum_i (y_{hi} - \hat{Y}_h) (x_{hi} - \hat{X}_h)^\tau (\hat{\psi}_{-lj} - \hat{\psi}) \\
&\quad + \frac{W_l}{n_l - 1} (y_{lj} - \bar{y}_l) + o_p\left(\frac{1}{n}\right) \\
&= \sum_h \frac{n W_h^2}{n_h^2} \sum_i (y_{hi} - \hat{Y}_h) (x_{hi} - \hat{X}_h)^\tau \left[\sum_h \frac{n W_h^2}{n_h^2} \sum_i (x_{hi} - \hat{X}_h) (x_{hi} - \hat{X}_h)^\tau \right]^{-1} \cdot \\
&\quad \cdot \frac{W_l}{n_l - 1} (x_{lj} - \bar{x}_l) + \frac{W_l}{n_l - 1} (y_{lj} - \bar{y}_l) + o_p\left(\frac{1}{n}\right) \\
&= \frac{W_l}{n_l - 1} [y_{lj} - \bar{y}_l - \hat{B}(x_{lj} - \bar{x}_l)] + o_p\left(\frac{1}{n}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{\sigma}_J - \hat{\sigma} &= \sum_l W_h^2 (n_l - 1) \left(\frac{1}{n_l} - \frac{1}{N_l} \right) \sum_j (\hat{Y} - \hat{Y}_{-lj}) (\hat{Y} - \hat{Y}_{-lj})^\tau - \hat{\sigma} \\
&= \sum_l \frac{W_l^2}{n_l - 1} \left(\frac{1}{n_l} - \frac{1}{N_l} \right) \sum_j [y_{lj} - \bar{y}_l - \hat{B}(x_{lj} - \bar{x}_l)] [y_{lj} - \bar{y}_l - \hat{B}(x_{lj} - \bar{x}_l)]^\tau \\
&\quad + o_p\left(\frac{1}{n}\right) - \hat{\sigma} \\
&= o_p\left(\frac{1}{n}\right),
\end{aligned}$$

or $n(\hat{\sigma}_J - \hat{\sigma}) \rightarrow_p 0$.

Proof of theorem 3

The result in Lemma 2 holds if we change $\bar{X}$ to θ_0 , $\bar{x}_{st}$ to $\bar{z}_{st}$, i.e.,

$$\hat{t} = \hat{t}(\theta_0) = S_{zz}^{-1} (\bar{z}_{st} - \theta_0) + o_p(A^{\frac{1}{2}}) = O_p(A^{\frac{1}{2}}),$$

where $S_{zz} = \sum_h n w_h^2 n_h^{-2} \sum_i (z_{hi} - \hat{Z}_h)(z_{hi} - \hat{Z}_h)^\tau$. Therefore,

$$\begin{aligned}
l_E(\theta_0) &= \sum_h \sum_i \log[1 + m_h \hat{t}^\tau(\theta_0)(z_{hi} - \hat{Z}_h)] \\
&= \sum_h \sum_i m_h \hat{t}^\tau(\theta_0)(z_{hi} - \hat{Z}_h) - \frac{1}{2} \sum_h \sum_i [m_h \hat{t}^\tau(\theta_0)(z_{hi} - \hat{Z}_h)]^2 \\
&\quad + o_p(1) \\
&= \sum_h m_h \sum_i (\bar{z}_{st} - \theta_0)^\tau S_{zz}^{-1} (z_{hi} - \hat{Z}_h) \\
&\quad - \frac{1}{2} \sum_h n^2 W_h^2 n_h^{-2} \sum_i (\bar{z}_{st} - \theta_0) S_{zz}^{-1} (z_{hi} - \hat{Z}_h) (z_{hi} - \hat{Z}_h)^\tau S_{zz}^{-1} (\bar{z}_{st} - \theta_0) \\
&\quad + o_p(1) \\
&= n(\bar{z}_{st} - \theta_0)^\tau S_{zz}^{-1} (\bar{z}_{st} - \theta_0) - \frac{n}{2} (\bar{z}_{st} - \theta_0)^\tau S_{zz}^{-1} (\bar{z}_{st} - \theta_0) + o_p(1) \\
&= \frac{n}{2} (\bar{z}_{st} - \theta_0)^\tau S_{zz}^{-1} (\bar{z}_{st} - \theta_0) + o_p(1) \\
&= \frac{n}{2} (\bar{z}_{st} - \theta_0)^\tau \sigma_{zz}^{-1} (\bar{z}_{st} - \theta_0) + o_p(1),
\end{aligned}$$

where $S_{zz} \xrightarrow{P} \sum_h W_h^2 k_h \sigma_{h,zz}$.

Similarly,

$$l_E(\tilde{\theta}) = \frac{n}{2} (\bar{x}_{st} - \theta_k)^\tau \sigma_{xx}^{-1} (\bar{x}_{st} - \theta_k) + o_p(1).$$

where $\sigma_{xx} = \sum_h W_h^2 k_h \sigma_{h,xx}$. Therefore,

$$\begin{aligned}
W_E(\theta_0) &= 2[l_E(\theta_0) - l_E(\tilde{\theta})] \\
&= n[(\bar{z}_{st} - \theta_0)^\tau \sigma_{zz}^{-1} (\bar{z}_{st} - \theta_0) - (\bar{x}_{st} - \theta_k)^\tau \sigma_{xx}^{-1} (\bar{x}_{st} - \theta_k)] + o_p(1).
\end{aligned}$$

Since

$$\begin{aligned}
\sigma_{zz}^{-1} &= \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} I_{d-p} & -\sigma_{xx}^{-1} \sigma_{xy} \\ 0 & I_p \end{pmatrix} \begin{pmatrix} \sigma_{xx}^{-1} & 0 \\ 0 & \sigma_{y \cdot x}^{-1} \end{pmatrix} \begin{pmatrix} I_{d-p} & 0 \\ -\sigma_{yx} \sigma_{xx}^{-1} & I_p \end{pmatrix}.
\end{aligned}$$

where $\sigma_{y \cdot x} = \sigma_{yy} - \sigma_{yx}\sigma_{xx}^{-1}\sigma_{xy}$, $\sigma_{yy} = \sum_h W_h^2 k_h \sigma_{h,yy}$, $\sigma_{xy} = \sum_h W_h^2 k_h \text{Cov}(x_h, y_h)$, we can write

$$\begin{aligned} W_E(\theta_0) &= n[(\bar{y}_{st} - \theta_{u0}) - \sigma_{yx}\sigma_{xx}^{-1}(\bar{x}_{st} - \theta_k)]^\tau \sigma_{y \cdot x}^{-1}[(\bar{y}_{st} - \theta_{u0}) - \sigma_{yx}\sigma_{xx}^{-1}(\bar{x}_{st} - \theta_k)] \\ &\quad + o_p(1). \end{aligned} \tag{6.14}$$

Now

$$E[(\bar{y}_{st} - \theta_{u0}) - \sigma_{yx}\sigma_{xx}^{-1}(\bar{x}_{st} - \theta_k)] = 0,$$

$$\begin{aligned} &n \text{Var}(\bar{y}_{st} - \theta_{u0} - \sigma_{yx}\sigma_{xx}^{-1}(\bar{x}_{st} - \theta_k)) \\ &= n[\sum_h W_h^2(\frac{1}{n_h} - \frac{1}{N_h})\sigma_{h,yy} + \sigma_{yx}\sigma_{xx}^{-1}\sum_h W_h^2(\frac{1}{n_h} - \frac{1}{N_h})\sigma_{h,xx}\sigma_{xx}^{-1}\sigma_{xy} \\ &\quad - 2\sum_h W_h^2(\frac{1}{n_h} - \frac{1}{N_h})\text{Cov}(y_h, x_h)\sigma_{xx}^{-1}\sigma_{xy}] \\ &\rightarrow \sum_h W_h^2 k_h \sigma_{h,yy} + \sigma_{yx}\sigma_{xx}^{-1}\sum_h W_h^2 k_h \sigma_{h,xx}\sigma_{xx}^{-1}\sigma_{xy} \\ &\quad - 2\sum_h W_h^2 k_h \text{Cov}(y_h, x_h)\sigma_{xx}^{-1}\sigma_{xy} \\ &= \sigma_{yy} - \sigma_{yx}\sigma_{xx}^{-1}\sigma_{xy} \\ &= \sigma_{y \cdot x}, \end{aligned}$$

Therefore,

$$W_E(\theta_0) \xrightarrow{L} \chi^2(p),$$

noting that

$$n^{\frac{1}{2}}[(\bar{y}_{st} - \theta_{u0}) - \sigma_{yx}\sigma_{xx}^{-1}(\bar{x}_{st} - \theta_k)] \xrightarrow{L} N(0, \sigma_{y \cdot x}).$$

Chapter 3

Empirical Likelihood and General Estimating Equations under Stratified Sampling

3.1 Introduction

Likelihood and estimating equations provide the most common approaches to parametric inference. Recently, it has also been shown to be useful in non-parametric contexts (Qin and Lawless(1994,5) and the references cited therein). Our purpose in this chapter is to combine empirical likelihood, estimating equations and stratified sampling technology together. In order to simplify the discussion, we suppose i.i.d. sample in each stratum, but the discussion can be carried over to simple random sampling without replacement.

Suppose that a target population with d -dimensional characteristic x has the

unknown distribution function F and a p -dimensional parameter θ associated with F . We are interested in making inference on θ . The sampling scheme is as follows. Suppose that the target population is divided into H strata with known weight W_h for all strata. We have an i.i.d. sample $x_{h1}, \dots, x_{hn_h}$ on the characteristic x_h from stratum h with distribution function F_h , where x_{hi} 's are d -dimensional ($h = 1, \dots, H, i = 1, \dots, n_h$). The F_h 's are unknown, and the samples across strata are independent. From the context we know that

$$F = \sum_h W_h F_h. \quad (1.1)$$

We also assume that information about θ and F is available in the form of $r \geq p$ functionally independent unbiased estimation functions, that is, functions $g_j(x, \theta), j = 1, \dots, r$ such that $E_F\{g_j(x, \theta)\} = 0$. In vector form,

$$g(x, \theta) = (g_1(x, \theta), \dots, g_r(x, \theta))^T,$$

where $g(x, \theta)$ satisfies

$$E_F\{g(x, \theta)\} = \sum_h W_h E_{F_h}\{g(x_h, \theta)\} = 0. \quad (1.2)$$

In the following, we will show how to use such information to estimate θ and F in conjunction with empirical likelihood. But first for illustration we give some examples that we will return to later.

Example 1. Sometimes we have information relating to the first and second moment of a variable. For example, let x be univariate with mean θ and $E(x^2) = m(\theta)$, where $m(\cdot)$ is a known function. The information about F can be expressed

in the form (1.2) by taking $g(x, \theta) = (x - \theta, x^2 - m(\theta))^T$ and the restriction becomes

$$\begin{pmatrix} \sum_h W_h E x_h - \theta \\ \sum_h W_h E x_h^2 - m(\theta) \end{pmatrix} = 0 \quad (1.3)$$

Example 2. Let $x = (y, z)^T$ be bivariate with $E(y) = E(z) = \theta$. In this case we can take $g(x, \theta) = (y - \theta, z - \theta)^T$ and the restriction becomes

$$\begin{pmatrix} \sum_h W_h E y_h - \theta \\ \sum_h W_h E z_h - \theta \end{pmatrix} = 0. \quad (1.4)$$

A somewhat similar problem is when $E(y) = \bar{Y}$ is known and $E(z) = \theta$ is to be estimated, in which case we would have $g(x) = (y - \bar{Y}, z - \theta)^T$. This problem has been discussed in the survey context in chapter 2, where $\bar{X}$ and $\bar{Y}$ played the roles of $\bar{Y}$ and θ respectively.

We will show later how to use the empirical likelihood method to solve the above problems. The basic idea is to maximize an empirical likelihood subject to constraints provided by (1.2) to get the maximum empirical likelihood ratio estimators (MELR) for θ and F . In section 3.2, we give MELR estimators for θ and F , and establish asymptotic normality. Section 3.3 shows how the MELR method can be used in models where parameters are subject to constraints. Hypothesis testing problems will also be considered. Finally, section 3.4 presents examples to demonstrate how these methods can be used in practice. Proofs are delegated to section 3.5. .

3.2 MELR Estimators and Their Properties

3.2.1 MELR estimators

The empirical likelihood function given by

$$L(F) = \prod_{h=1}^H \prod_{i=1}^{n_h} p_{hi}, \quad (2.1)$$

where $p_{hi} = Pr(x_h = x_{hi})$. Only those F distributions with F_h 's which have an atom of probability on each x_{hi} have nonzero likelihood. We know (2.1) is maximized by the empirical distribution function $F_n(x) = \sum_h W_h F_{n_h}(x)$ by noting (1.1), where $n = \sum_h n_h$, $F_{n_h}(x) = \frac{1}{n_h} \sum_i 1(x_{hi} < x)$, $1(x_{hi} < x) = (1(x_{hi,1} < x_1), \dots, 1(x_{hi,d} < x_d))^T$, $x_{hi,j}$ and x_j are j -th component of vector x_{hi} and x respectively. $1(x_{hi,j} < x_j)$ is the indicator of set $(x_{hi,j} < x_j)$. The empirical likelihood ratio is then defined as $R(F) = L(F)/L(F_n)$, which reduces to

$$R(F) = \prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi}. \quad (2.2)$$

We remark that formulas here and elsewhere in this chapter do not require that the x_{hi} 's be distinct. Since we are interested in estimating the parameter θ , and we know the estimating equation (1.2), we define the empirical likelihood ratio function

$$R_E(\theta) = \sup \left\{ \prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi} \mid p_{hi} \geq 0, \sum_i p_{hi} = 1, \sum_h W_h \sum_i p_{hi} g_{hi}(\theta) = 0 \right\}, \quad (2.3)$$

where $g_{hi}(\theta) = g(x_{hi}, \theta)$ for all h and i . As discussed in chapter 2, for any given θ in a neighbourhood of true θ , a unique value for the right side of (2.3) exists with probability tending to 1 as all $n_h \rightarrow \infty$. The maximum can be found via Lagrange multiplier method. Let

$$G = \sum_h \sum_i \log p_{hi} - \sum_h \lambda_h \left(\sum_i p_{hi} - 1 \right) - nt^T \left(\sum_h W_h \sum_i p_{hi} g_{hi}(\theta) \right).$$

where t is a r -dimension vector. Then

$$\frac{\partial G}{\partial p_{hi}} = p_{hi}^{-1} - \lambda_h - nW_h t^\tau g_{hi}(\theta) = 0, \quad i = 1, \dots, n_h; h = 1, \dots, H,$$

$$\begin{aligned} \sum_i p_{hi} \frac{\partial G}{\partial p_{hi}} &= n_h - \lambda_h - nW_h t^\tau \sum_i p_{hi} g_{hi}(\theta) = 0 \\ \Rightarrow \lambda_h &= n_h - nW_h t^\tau \hat{g}_h(\theta), \end{aligned}$$

where $\hat{g}_h(\theta) = \sum_i p_{hi} g_{hi}(\theta)$. Hence, letting $m_h = nW_h n_h^{-1}$,

$$p_{hi} = \frac{1}{n_h [1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))]}, \quad (2.4)$$

$$0 = \sum_i p_{hi} g_{hi}(\theta) - \hat{g}_h(\theta) = \sum_i \frac{g_{hi}(\theta) - \hat{g}_h(\theta)}{n_h [1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))]}, \quad (2.5)$$

with the restriction from $\frac{\partial G}{\partial t} = 0$, i.e.,

$$\sum_h W_h \sum_i p_{hi} g_{hi}(\theta) = 0,$$

or

$$\sum_h W_h \hat{g}_h(\theta) = 0, \quad (2.6)$$

from which (see section 3.2.2) t can be determined in terms of θ . Note that $t = t(\theta)$ is actually a continuous differentiable vector valued function of θ . Therefore, the empirical log-likelihood ratio statistic is given by

$$l_E(\theta) = \sum_h \sum_i \log[1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))] \quad (2.7)$$

with t being defined by (2.5). We minimize $l_E(\theta)$ to obtain an estimate $\tilde{\theta}$ of the parameter θ (called MELR). In addition, this yields estimates $\tilde{p}_{hi}$'s from (2.4), and an estimate for distribution F as

$$\tilde{F}_n(s) = \sum_h W_h \sum_i \tilde{p}_{hi} 1(x_{hi} < s). \quad (2.8)$$

Sometime in order to indicate $l_E(\theta)$ and $l_E(\tilde{\theta})$ also depend on $t(\theta)$, $\tilde{t}$, $l_E(\theta, t(\theta))$ and $l_E(\tilde{\theta}, \tilde{t})$ are used to denote $l_E(\theta)$, $l_E(\tilde{\theta})$ separately.

3.2.2 Numerical method for given θ

We first show how to solve (2.4) (2.5) and (2.6) to get p_{hi} 's for given θ , where $0 \leq p_{hi} \leq 1$. Here implicitly p_{hi} 's are functions of θ . Suppose all $E_h g(x_h, \theta) = M_h(\theta)$ are also given, where E_h indicates the expectation is taken with respect to F_h (we note here $M_h(\theta)$ may include other characteristics of x_h which we are not interested). Here, of course, $M_h(\theta)$'s must satisfy $\sum_h W_h M_h(\theta) = 0$, then using Lagrange multiplier method we find that $\sum_h \sum_i \log p_{hi}$ attains the maximum

$$-\sum_h \sum_i \log[1 + \psi_h^T(g_{hi}(\theta) - M_h(\theta))] - \sum_h \log n_h$$

with ψ_h satisfying

$$\sum_i \frac{g_{hi}(\theta) - M_h(\theta)}{1 + \psi_h^T(g_{hi}(\theta) - M_h(\theta))} = 0,$$

where implicitly

$$p_{hi} = \frac{1}{n_h[1 + \psi_h^T(g_{hi}(\theta) - M_h(\theta))]}.$$

Hence, we can say that ψ_h is a function of $M_h(\theta)$ and θ and we want to maximize

$$-\sum_h \sum_i \log[1 + \psi_h^T(M_h(\theta), \theta)(g_{hi}(\theta) - M_h(\theta))]$$

subject to $\sum_h W_h M_h(\theta) = 0$. Using Lagrange method again, let

$$\begin{aligned} G_1 &= G_1(M_1(\theta), \dots, M_H(\theta), t) \\ &= -\sum_h \sum_i \log[1 + \psi_h^T(M_h(\theta), \theta)(g_{hi}(\theta) - M_h(\theta))] - nt^T(\sum_h W_h M_h(\theta)). \end{aligned}$$

Since

$$\begin{aligned} \frac{\partial G_1}{\partial M_h(\theta)} &= -\sum_i \frac{(\frac{\partial \psi_h}{\partial M_h(\theta)})^T(g_{hi}(\theta) - M_h(\theta)) - \psi_h}{1 + \psi_h^T(M_h(\theta), \theta)(g_{hi}(\theta) - M_h(\theta))} - nW_h t \\ &= n_h \psi_h - nW_h t. \end{aligned}$$

letting $n_h \psi_h - n W_h t = 0$, we get $\psi_h = m_h t$, which implies that

$$\sum_i \frac{g_{hi}(\theta) - M_h(\theta)}{1 + m_h t^\tau (g_{hi}(\theta) - M_h(\theta))} = 0, \quad (2.9)$$

$$p_{hi} = \frac{1}{n_h [1 + m_h t^\tau (g_{hi}(\theta) - M_h(\theta))]}, 0 \leq p_{hi} \leq 1, \quad (2.10)$$

or we can say that $M_h(\theta)$ is a function of t , say $\bar{M}_h(t)$, and t should be chosen such that

$$\sum_h W_h \bar{M}_h(t) = 0. \quad (2.11)$$

In conclusion, for any given θ , we can select a value of t , solving (2.9) for each $h = 1, \dots, H$ under the conditions of $p_{hi} \geq 0$ to get $\bar{M}_h(t)$, then check whether $\bar{M}_h(t)$'s satisfy equation (2.11). Another way is combining (2.9) and (2.11) as a system of equations, using Newton's method to solve these equations for appropriately selected initial values. Usually (2.10) will be satisfied.

Using the same method as in Chapter 2, we can prove the existence of solutions to (2.9), (2.10) and (2.11). We also know that $M_h(\theta)$'s are continuous differentiable functions of θ .

3.2.3 Asymptotic Results

The following lemmas will be used to prove that at some point $\tilde{\theta}$, $l_E(\theta)$ attains its minimum and to establish asymptotic properties of $\tilde{\theta}$. We use $\|\cdot\|$ to denote Euclidean norm.

Lemma 1 *Let Y_i be i.i.d. random d -dimension variables and define $z_n = \text{Max}_{1 \leq i \leq n} \|Y_i\|$. If $E(\|Y_i\|^k) < \infty$, $k > 0$, then*

$$z_n = o(n^{\frac{1}{k}}) \text{ (a.s.)}$$

Lemma 2 Suppose as $n \rightarrow \infty$, $n/n_h \rightarrow k_h > 0$ for all h . And suppose that in a neighborhood of the true value θ_0 , $E_h[(g(x_h, \theta_0) - E_h g(x_h, \theta_0))(g(x_h, \theta_0) - E_h g(x_h, \theta_0))^T] = \sigma_h(\theta_0) > \sigma_1 > 0$ for all h , where σ_1 is positive definite. $\|g(x, \theta)\|^3$ is bounded by some integrable function $G(x)$ in this neighbourhood. Then for any given θ in this neighbourhood,

$$t(\theta) = \hat{S}_n^{-1}(\theta) \bar{g}_{st}(\theta) + o(n^{-\frac{1}{3}}) = O(n^{-\frac{1}{3}}) \text{ (a.s.)}, \quad (2.12)$$

where $\hat{S}_n(\theta) = \sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T$, $\hat{g}_h(\theta) = \sum_i \hat{p}_{hi} g_{hi}(\theta)$, $\bar{g}_{st}(\theta) = \sum_h W_h \hat{g}_h(\theta)$, $\hat{p}_{hi} = \{n_h [1 + m_h t^T(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))]\}^{-1}$. $t(\theta)$ is a solution to (2.4) (2.5) and (2.6).

Lemma 3 (Jennrich, 1969) Let g be a function on $X \times \Theta$, where X is a Euclidean space and Θ is a compact subset of a Euclidean space. Let $g(x, \theta)$ be a continuous function of θ for each x and a measurable function of x for each θ . Assume also that $|g(x, \theta)| \leq h(x)$ for all x and θ , where h is integrable with respect to a probability distribution function F on x . If $x_1, x_2, \dots$ is a random sample from F then for almost every sequence (x_t)

$$n^{-1} \sum_{t=1}^n g(x_t, \theta) \rightarrow \int g(x, \theta) dF(x) \text{ (a.s.)}$$

uniformly for all θ in Θ .

Lemma 4 In addition to the conditions of Lemma 2, we further assume that $\partial g(x, \theta)/\partial \theta$ is continuous in a neighborhood of the true value θ_0 , $\|\partial g(x, \theta)/\partial \theta\|$ is bounded by some integrable function $G(x)$ in this neighbourhood, and the rank of $\sum_h W_h E_h[\partial g(x_h, \theta_0)/\partial \theta]$ is p . Then, as $n \rightarrow \infty$, with probability 1, $l_E(\theta)$ attains its

minimum value at some point $\tilde{\theta}$ in the interior of the ball $\|\theta - \theta_0\| \leq n^{-\frac{1}{2}}$, and $\tilde{\theta}$ and $\tilde{t} = \tilde{t}(\tilde{\theta})$ satisfy

$$Q_{1n}(\tilde{\theta}, \tilde{t}) = 0, \quad Q_{2n}(\tilde{\theta}, \tilde{t}) = 0, \quad (2.13)$$

where

$$Q_{1n}(\theta, t) = \sum_h \frac{W_h}{n_h} \sum_i \frac{g_{hi}(\theta)}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]}, \quad (2.14)$$

with

$$\hat{g}_h(\theta) = \sum_i \frac{g_{hi}(\theta)}{n_h [1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))]}, \quad (2.15)$$

and

$$Q_{2n}(\theta, t) = \sum_h \frac{W_h}{n_h} \sum_i \frac{1}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]} \left[\frac{\partial g_{hi}(\theta)}{\partial \theta^\tau} \right]^\tau t. \quad (2.16)$$

Theorem 1 *In addition to the conditions of Lemma 4 above, we further assume that $\frac{\partial^2 g(x, \theta)}{\partial \theta \partial \theta^\tau}$ is continuous in θ in a neighbourhood of the true value θ_0 and $\|\frac{\partial^2 g(x, \theta)}{\partial \theta \partial \theta^\tau}\|$ is bounded by some integrable function $G(x)$ in that neighbourhood. Then*

$$\sqrt{n}(\tilde{\theta} - \theta_0) \rightarrow N(0, V), \quad \sqrt{n}(\tilde{t} - 0) \rightarrow N(0, U).$$

$$\sqrt{n}(\tilde{F}_n(s) - F(s)) \rightarrow N(0, W),$$

where $\tilde{F}_n(x) = \sum_h W_h \sum_i \tilde{p}_{hi} 1(x_{hi} < s)$, $\tilde{p}_{hi} = \frac{1}{n_h} \frac{1}{1 + m_h \tilde{t}^\tau (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))}$, and V, U, W are defined in the proof Section 3.5. Further, $\tilde{\theta}$ and $\tilde{t}$ are asymptotically independent.

Theorem 2 *In the semi-parametric model (1.2), for testing $H_0 : \theta = \theta_0$ the empirical likelihood ratio test statistic*

$$R = 2 \left\{ \sum_h \sum_i \log[1 + m_h t^\tau(\theta_0)(g_{hi}(\theta_0) - \hat{g}_h(\theta_0))] \right\}$$

$$-\sum_h \sum_i \log[1 + m_h \tilde{t}^\tau(\tilde{\theta})(g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))]$$

is asymptotically χ_p^2 under H_0 , assuming the conditions of theorem 1.

Corollary 1 Assume $r > p$. In order to test the model (1.2), i.e.,

$$E_F\{g(x, \theta)\} = \sum_h W_h E_{F_h}\{g(x_h, \theta)\} = 0,$$

we can use the empirical likelihood ratio statistic

$$R_1 = 2 \sum_h \sum_i \log[1 + m_h \tilde{t}^\tau(\tilde{\theta})g_{hi}(\tilde{\theta})].$$

Under the conditions of theorem 2, R_1 is asymptotically χ_{r-p}^2 if (1.2) is correct.

3.3 MELR Estimators and Testing with Constraints

In this section we extend the empirical likelihood methods to deal with case in which there are constraints on parameters. Suppose there is a q -dimensional constraint on θ , i.e.,

$$r(\theta) = 0, \tag{3.1}$$

where $r(\theta)$ is a $q \times 1$ ($q \leq p$) and the $q \times p$ matrix $R(\theta) = \frac{\partial r}{\partial \theta^\tau}$ is of full rank q . To minimize $l_E(\theta)$ defined by (2.7) subject to $r(\theta) = 0$, we consider

$$G_2 = \frac{1}{n} l_E(\theta) + \nu^\tau r(\theta),$$

where ν is a $q \times 1$ vector of Lagrange multipliers. Differentiating G_2 with respect to θ and ν , we have

$$\frac{1}{n} \frac{\partial l_E(\theta)}{\partial \theta^\tau} + \nu^\tau R(\theta) = 0, \quad r(\theta) = 0. \tag{3.2}$$

Consequently, to minimize $l_E(\theta)$ subject to $r(\theta) = 0$, we need the solution of

$$Q_{1n}(\theta, t, \nu) = 0, \quad Q_{2n}(\theta, t, \nu) = 0, \quad Q_{3n}(\theta, t, \nu) = 0, \quad (3.3)$$

where

$$\begin{aligned} Q_{1n}(\theta, t, \nu) &= \sum_h \frac{W_h}{n_h} \sum_i \frac{g_{hi}(\theta)}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]}, \\ Q_{2n}(\theta, t, \nu) &= \left[\sum_h \frac{W_h}{n_h} \sum_i \frac{1}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]} \left(\frac{\partial g_{hi}(\theta)}{\partial \theta} \right)^\tau \right] t + R^\tau(\theta) \nu, \\ Q_{3n}(\theta, t, \nu) &= r(\theta), \end{aligned}$$

where $\hat{g}_h(\theta)$ is defined by (2.15). We denote the solution of (3.3) as $(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r)$. To discuss the asymptotic behavior of these estimators and the test statistics based on them, we suppose that $r(\theta_0) = 0$, where θ_0 is the true value of θ . First we give two lemmas which show that the solution to (3.3) does exist with probability tending 1 as $n \rightarrow \infty$.

We need the following lemma from Aitchison and Silvey (1958) in the proof of Lemma 6:

Lemma 5 *If $\psi(\lambda)$ is a continuous function mapping $\mathbb{R}^s$ into itself with the property that, for every λ such that $\|\lambda\| = 1$, $\lambda^\tau \psi(\lambda) < 0$, then there exists a point $\hat{\lambda}$ such that $\|\hat{\lambda}\| < 1$ and $\psi(\hat{\lambda}) = 0$.*

Lemma 6 *Suppose that $g(x, \theta)$ is continuous and differentiable in the neighbourhood of θ_0 and $E \|g(x, \theta)\|^3 < \infty$. Furthermore, assume that $\sigma_h(\theta_0) > 0$ for all h and that $\sum_h W_h E_h \left\{ \frac{\partial g(x_h, \theta_0)}{\partial \theta} \right\}$ is of full rank. Then in the sphere $\{\theta : \|\theta - \theta_0\| \leq d_n\}$, the equation $Q_{1n}(\theta, t) = 0$ almost surely has roots $t = t(\theta) = O(d_n)$, and $t(\theta)$ is continuous and differentiable when θ belongs to this sphere, where $d_n = n^{-\frac{1}{3}-\delta}$, $\frac{1}{6} > \delta > 0$.*

Lemma 7 Assume that the conditions of Lemma 3.1 hold; that in a neighbourhood of θ_0 , $r(\theta)$ is a continuous differentiable function; that the $q \times p$ matrix $R(\theta)$ is of rank q ; and that

$$\frac{\partial^2 r(\theta)}{\partial \theta \partial \theta^\tau}, \quad \frac{\partial^2 g(x, \theta)}{\partial \theta \partial \theta^\tau}$$

exist and are bounded by some constant and some integrable function respectively. Then the equations (3.3) almost surely have solutions in $U_{d_n} = \{(\theta, t, \nu) : \|\theta - \theta_0\| + \|t\| + \|\nu\| \leq d_n\}$ as n goes to infinity, and any solution of (3.3) in U_{d_n} minimizes $l_E(\theta)$ subject to the condition $r(\theta) = 0$.

Theorem 3 below establishes the asymptotic normality of $\tilde{\theta}_r$.

Theorem 3 Under conditions of Lemma 3.2, we have

$$\sqrt{n} \tilde{\xi}_r = \sqrt{n} \begin{pmatrix} \tilde{\theta}_r - \theta_0 \\ \tilde{\nu}_r \end{pmatrix} \rightarrow N \left(0, \begin{pmatrix} P & 0 \\ 0 & H \end{pmatrix} \right), \quad (3.4)$$

where P, H are defined in Section 3.5.

Now we turn to the problem about how to test $H_0 : r(\theta) = 0$. There are three popular tests based on the likelihood. They are: likelihood-ratio test, Lagrange-multiplier test and Wald test. Of course here they should be based on the empirical likelihood. These statistics are defined respectively as

$$ELR = 2l_E(\tilde{\theta}_r), \quad LA = n \tilde{\nu}_r^\tau H_{\tilde{\theta}_r}^{-1} \tilde{\nu}_r, \quad WA = nr^\tau(\hat{\theta}) H_{\hat{\theta}} r(\hat{\theta}), \quad (3.5)$$

where $\hat{\theta}$ is a solution of the estimating equation $\sum_h W_h n_h^{-1} \sum_i g_{hi}(\theta) = 0$, and $H_{\hat{\theta}}$ defined in (5.27), is the asymptotic covariance matrix of $\sqrt{n} \tilde{\nu}_r$.

The following theorem gives the asymptotic behavior and relationship between the test statistics defined in (3.5).

Theorem 4 *Under the assumptions of Theorem 3.3, the three test statistics in (3.5) are asymptotically equivalent, and each of them is asymptotically distributed as χ_q^2 under H_0 .*

3.4 Examples

We consider several illustrations of the estimation procedures. We primarily consider large-sample aspects, but computational issues are also discussed by giving the equations and the method to solve them.

Example 1 (continued) We apply the approach in Section 2 to solve the problem displayed in Section 1. Equations (2.9), (2.11) and (2.16) lead to

$$\frac{1}{n_h} \sum_i \frac{x_{hi} - \theta - u_h}{1 + m_h t_1 (x_{hi} - \theta - u_h) + m_h t_2 (x_{hi}^2 - m(\theta) - v_h)} = 0, \quad h = 1, \dots, H,$$

$$\frac{1}{n_h} \sum_i \frac{x_{hi}^2 - m(\theta) - v_h}{1 + m_h t_1 (x_{hi} - \theta - u_h) + m_h t_2 (x_{hi}^2 - m(\theta) - v_h)} = 0, \quad h = 1, \dots, H,$$

$$\sum_h W_h u_h = 0, \quad \sum_h W_h v_h = 0.$$

$$t_1 + m'(\theta)t_2 = 0$$

The third equation implies $t_1 = -m'(\theta)t_2$ and by substituting this into the first and the second we get $2H + 2$ equations in t_2 , θ , $u_1, \dots, u_H$, $v_1, \dots, v_H$ to solve. We use Fortran NAG library function to solve them, for initial value $t_2 = 0$, $\theta = \sum_h W_h \bar{x}_h$, $u_h = 0$, $v_h = 0$, $h = 1, \dots, H$. Usually the solutions should satisfy (2.10).

The results of Theorem 1 show that $\sqrt{n}(\tilde{\theta} - \theta_0) \rightarrow N(0, V)$, where V is given in Theorem 1, or $\tilde{\theta} - \theta_0 \approx N(0, \frac{1}{n}V)$. Since

$$\frac{M_{11}}{n} \approx \sum_h \frac{W_h^2}{n_h} \sigma_h(\theta_0) \triangleq \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

where $a_{11} = \sum_h \frac{W_h^2}{n_h} \text{Var}(x_h)$, $a_{12} = \sum_h \frac{W_h^2}{n_h} \text{Cov}(x_h, x_h^2)$, $a_{22} = \sum_h \frac{W_h^2}{n_h} \text{Var}(x_h^2)$, and $M_{12} = (1, m'(\theta))^T$, we have

$$n M_{21} M_{11}^{-1} M_{12} \approx \frac{a_{22} - 2m'(\theta)a_{12} + a_{11}(m'(\theta))^2}{a_{11}a_{22} - a_{12}^2},$$

or

$$\frac{V}{n} = \frac{a_{11}a_{22} - a_{12}^2}{a_{22} - 2m'(\theta)a_{12} + a_{11}(m'(\theta))^2}.$$

Also, since

$$\Delta = a_{22} - 2m'(\theta)a_{12} + a_{11}(m'(\theta))^2 = \text{Var} \left(m'(\theta)\bar{x}_{st} - \sum_h \frac{W_h}{n_h} \sum_i x_{hi}^2 \right) > 0,$$

$$\begin{aligned} \frac{V}{n} &= \Delta^{-1}(a_{11}a_{22} - a_{12}^2) \\ &= a_{11} - \Delta^{-1}(\Delta a_{11} - a_{11}a_{12} + a_{12}^2) \\ &= a_{11} - \Delta^{-1}(-2m'(\theta)a_{11}a_{12} + a_{11}^2(m'(\theta))^2 + a_{12}^2) \\ &= a_{11} - \Delta^{-1}(m'(\theta)a_{11} - a_{12})^2. \end{aligned}$$

Thus $\frac{V}{n} \leq \text{Var}(\bar{x}_{st})$, which is the variance of $\bar{x}_{st} - \theta_0$. Therefore, $\tilde{\theta}$ is asymptotically at least as efficient as $\bar{x}_{st}$.

Example 2 (continued) [Two-sample problem with common mean]. In this case observations (x_{hi}, y_{hi}) $i = 1, \dots, n_h$ occur in independent pairs and $\sum_h W_h E x_h =$

$\sum_h W_h E y_h = \theta$. To estimate θ , we consider the estimating equations based on $g_1 = x - \theta$, $g_2 = y - \theta$ and we associate the empirical likelihood probability p_{hi} with (x_{hi}, y_{hi}) . After some simplification, from equations (2.9), (2.11) and (2.16), we have

$$\begin{aligned} \frac{1}{n_h} \sum_i \frac{x_{hi} - \theta - u_h}{1 + m_h t_1(x_{hi} - \theta - u_h) + m_h t_2(y_{hi} - \theta - v_h)} &= 0, \quad h = 1, \dots, H, \\ \frac{1}{n_h} \sum_i \frac{y_{hi} - m(\theta) - v_h}{1 + m_h t_1(x_{hi} - \theta - u_h) + m_h t_2(y_{hi} - \theta - v_h)} &= 0, \quad h = 1, \dots, H, \\ \sum_h W_h u_h &= 0, \quad \sum_h W_h v_h = 0, \\ t_1 + t_2 &= 0 \end{aligned}$$

or

$$\begin{aligned} \frac{1}{n_h} \sum_i \frac{x_{hi}}{1 + m_h t(x_{hi} - y_{hi})} - \theta - u_h &= 0, \quad h = 1, \dots, H, \\ \frac{1}{n_h} \sum_i \frac{y_{hi}}{1 + m_h t(x_{hi} - y_{hi})} - \theta - v_h &= 0, \quad h = 1, \dots, H, \\ \sum_h W_h u_h &= 0, \quad \sum_h W_h v_h = 0. \end{aligned}$$

The initial values may be set to

$$\theta = \frac{1}{2}(\bar{x}_{st} + \bar{y}_{st}), \quad t = 0, \quad u_h = v_h = 0, \quad h = 1, \dots, H.$$

Here

$$\begin{aligned} \frac{M_{11}}{n} &\approx \sum_h \frac{W_h^2}{n_h} \begin{pmatrix} \sigma_{x_h}^2 & \sigma_{x_h, y_h}^2 \\ \sigma_{y_h, x_h}^2 & \sigma_{y_h}^2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad M_{12} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\ (1, 1) &\begin{pmatrix} a_{22} & a_{21} \\ a_{12} & a_{11} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (a_{22} - 2a_{12} + a_{11}) \end{aligned}$$

Therefore

$$\frac{V}{n} \approx \frac{a_{11}a_{22} - a_{12}^2}{a_{22} - 2a_{12} + a_{11}}$$

In particular, note that in the case where

$$\sigma_{x_h, y_h}^2 = Cov(x_h, y_h) = 0, \quad \frac{V}{n} = \sigma_{\bar{x}_{st}}^2 \sigma_{\bar{y}_{st}}^2 / (\sigma_{\bar{x}_{st}}^2 + \sigma_{\bar{y}_{st}}^2),$$

which is the same as the variance of the optimal linear combination estimator

$$\hat{\theta} = \left(\frac{a_{22}}{a_{11} + a_{22}} \right) \bar{x}_{st} + \left(\frac{a_{11}}{a_{11} + a_{22}} \right) \bar{y}_{st}.$$

3.5 Proofs

Proof of Lemma 1

Let $Y_{i(l)}$ be the l^{th} component of Y_i , then $\|Y_i\| \leq \sum_{l=1}^d |Y_{i(l)}|$ and

$$z_n \leq \max_{1 \leq i \leq n} \sum_{l=1}^d |Y_{i(l)}| \leq \sum_{l=1}^d \max_{1 \leq i \leq n} |Y_{i(l)}|.$$

Following Owen(1988), we get $\max_{1 \leq i \leq n} |Y_{i(l)}| = o(n^{\frac{1}{k}})$ (a.s.), hence we have $z_n = o(n^{\frac{1}{k}})$ (a.s.). $\square$

Proof of Lemma 2

By using the decomposition

$$\begin{aligned} g_{hi}(\theta) &= g_{hi}(\theta)[1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))] \\ &\quad - (g_{hi}(\theta) - \hat{g}_h(\theta) + \hat{g}_h(\theta))m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta)) \end{aligned}$$

and noting (2.5), we have

$$\begin{aligned} \hat{g}_h(\theta) &= \sum_i \hat{p}_{hi} g_{hi}(\theta) \\ &= \sum_i \frac{1}{n_h [1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))]} g_{hi}(\theta) \\ &= \bar{g}_h(\theta) - n W_h n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta), \quad (5.1) \end{aligned}$$

where $\bar{g}_h(\theta) = \sum_i g_{hi}(\theta)/n_h$. Since $\sum_h W_h \hat{g}_h(\theta) = 0$, we have

$$\begin{aligned} 0 &= \sum_h W_h \bar{g}_h(\theta) - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta) \\ &= \bar{g}_{st}(\theta) - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta). \end{aligned} \quad (5.2)$$

or

$$\bar{g}_{st}(\theta) = \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta). \quad (5.3)$$

From (5.1) and (5.3) we get

$$\begin{aligned} &W_h^2 \| \bar{g}_h(\theta) - \hat{g}_h(\theta) \|^2 \\ &= t^\tau(\theta) \left[n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \right]^2 t(\theta) \\ &\leq t^\tau(\theta) \left[\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \right]^2 t(\theta) \\ &= \| \bar{g}_{st}(\theta) \|^2 \rightarrow 0 \text{ (a.s.)} \end{aligned} \quad (5.4)$$

by the strong law of large number. Hence

$$\begin{aligned} \hat{S}_n(\theta) &= \sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\theta) - \bar{g}_h(\theta) + \bar{g}_h(\theta) - \hat{g}_h(\theta)) \\ &\quad \cdot (g_{hi}(\theta) - \bar{g}_h(\theta) + \bar{g}_h(\theta) - \hat{g}_h(\theta))^\tau \\ &= \sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\theta) - \bar{g}_h(\theta))(g_{hi}(\theta) - \bar{g}_h(\theta))^\tau \\ &\quad + \sum_h n W_h^2 n_h^{-1} (\bar{g}_h(\theta) - \hat{g}_h(\theta))(\bar{g}_h(\theta) - \hat{g}_h(\theta))^\tau \\ &\rightarrow S(\theta) \text{ (a.s.)}, \end{aligned} \quad (5.5)$$

where $S(\theta) = \sum_h W_h^2 k_h \sigma_h(\theta)$,

$$\sigma_h(\theta) = E_h(g(x_h, \theta) - E_h g(x_h, \theta))(g(x_h, \theta) - E_h g(x_h, \theta))^\tau.$$

Therefore, we have

$$\hat{S}_n(\theta) = \sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T = O(1) \text{ (a.s.)}. \quad (5.6)$$

Since $\sigma_h(\theta) > \sigma_1$, hence $S(\theta) > \sum_h W_h^2 k_h \sigma_1 > 0$, therefore

$$\hat{S}_n(\theta) > \frac{1}{2} \sum_h W_h^2 k_h \sigma_1 > 0 \text{ (a.s.)}.$$

From (5.3) we have

$$\begin{aligned} \|\bar{g}_{st}(\theta)\|^2 &= t^T(\theta) \left[\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T}{1 + m_h t^T(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \right]^2 t(\theta) \\ &\geq t^T(\theta) \left[\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T}{1 + m_h \|t(\theta)\| \|g_{hi}(\theta) - \hat{g}_h(\theta)\|} \right]^2 t(\theta) \\ &\geq t^T(\theta) \left[\sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T}{1 + m_h \|t(\theta)\| 2u_{h,n_h}^*} \right]^2 t(\theta) \end{aligned} \quad (5.7)$$

where $u_{h,n_h}^* = \max_i \|g_{hi}(\theta)\|$. By the result of Lemma 1. we have $u_{h,n_h}^* = o(n_h^{\frac{1}{3}})$ (a.s.). Hence

$$\max_h (2m_h u_{h,n_h}^*) = \max_h o(n_h^{\frac{1}{3}}) = o(n_{(1)}^{\frac{1}{3}}) = o(n^{\frac{1}{3}}) \text{ (a.s.)}. \quad (5.8)$$

Therefore, from (5.7) and (5.8), we get

$$\begin{aligned} \|\bar{g}_{st}(\theta)\|^2 &\geq t^T(\theta) \left[\frac{1}{1 + \|t(\theta)\| o(n^{\frac{1}{3}})} \cdot \sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T \right]^2 t(\theta) \\ &\geq t^T(\theta) \left[\frac{1}{1 + \|t(\theta)\| o(n^{\frac{1}{3}})} \right]^2 \left[\sum_h k_h W_h^2 \sigma_1 / 2 \right]^2 t(\theta) \text{ (a.s.)} \\ &\geq \left[\frac{\lambda}{1 + \|t(\theta)\| o(n^{\frac{1}{3}})} \right]^2 \|t(\theta)\|^2 \text{ (a.s.)}, \end{aligned}$$

where $\lambda = \sum_h k_h W_h^2 \sigma_{(1)}/2 > 0$ and $\sigma_{(1)}$ is the smallest eigenvalue of σ_1 . Hence

$$\begin{aligned} \|t(\theta)\| &\leq \|\bar{g}_{st}(\theta)\| [1 + \|t(\theta)\| o(n^{\frac{1}{3}})]/\lambda \quad (a.s.) \\ &= \lambda^{-1}(\|\bar{g}_{st}(\theta)\| + \|\bar{g}_{st}(\theta)\| \|t(\theta)\| o(n^{\frac{1}{3}})) \quad (a.s.). \end{aligned} \quad (5.9)$$

Since $\|\bar{g}_h(\theta) - M_h(\theta)\| = o(n^{-\frac{1}{3}}) \quad (a.s.)$, $\|\bar{g}_{st}(\theta)\| = \|\sum_h W_h(\bar{g}_h(\theta) - M_h(\theta))\| = o(n^{-\frac{1}{3}}) \quad (a.s.)$ so that $\|\bar{g}_{st}(\theta)\| o(n^{\frac{1}{3}}) = o(1) \quad (a.s.)$. Therefore, (5.9) becomes $\|t(\theta)\| \leq \lambda^{-1}(\|\bar{g}_{st}(\theta)\| + \|t(\theta)\| o(1)) \quad (a.s.)$, that is, $\|t(\theta)\| = O(n^{-\frac{1}{3}}) \quad (a.s.)$, which establish the last part of (2.12).

From (5.3) we have

$$\begin{aligned} \bar{g}_{st}(\theta) &= \left[\sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T \right] t(\theta) \\ &\quad - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T}{1 + m_h t(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \\ &\quad \cdot m_h t^T(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta)) t(\theta), \end{aligned}$$

and

$$\begin{aligned} &\left\| \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T m_h t^T(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))}{1 + m_h \hat{t}(\theta)^T (g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta) \right\| \\ &\leq \max_h \max_i |m_h t(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))| \\ &\quad \cdot \left\| \sum_h n W_h^2 n_h^{-2} \sum_i \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T}{1 + m_h \hat{t}(\theta)^T (g_{hi}(\theta) - \hat{g}_h(\theta))} \right\| \|t(\theta)\| \\ &= \max_h o(n_h^{\frac{1}{3}}) \|t(\theta)\|^2 O_p(1) \quad (a.s.) \\ &= o(n^{\frac{1}{3}}) O(n^{-\frac{2}{3}}) O(1) \quad (a.s.) \\ &= o(n^{-\frac{1}{3}}) \quad (a.s.). \end{aligned}$$

Therefore,

$$\bar{g}_{st}(\theta) = \left[\sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^T \right] t(\theta) + o(n^{-\frac{1}{3}}) \quad (a.s.),$$

or

$$t(\theta) = \hat{S}_n^{-1}(\theta)\bar{g}_{st}(\theta) + o(n^{-\frac{1}{3}}). \square$$

Proof of lemma 4

We follow Qin and Lawless(1994) in proving this Lemma. Denote $\theta = \theta_0 + un^{-\frac{1}{3}}$ for $\theta \in \{\theta \mid \|\theta - \theta_0\| = n^{-\frac{1}{3}}\}$, where $\|u\| = 1$. First we give a lower bound for $l_E(\theta)$ on the surface of the ball. From Lemmas 1, 2 and 3 we know that when $\|\theta - \theta_0\| \leq n^{-\frac{1}{3}}$,

$$t(\theta) = \hat{S}_n^{-1}(\theta)\bar{g}_{st}(\theta) + o(n^{-\frac{1}{3}}) \text{ (a.s.)} = O(n^{-\frac{1}{3}}) \text{ (a.s.)}$$

uniformly about $\theta \in \{\theta \mid \|\theta - \theta_0\| \leq n^{-\frac{1}{3}}\}$.

By this and Taylor's expansion, we have (uniformly in u)

$$\begin{aligned} l_E(\theta) &= \sum_h \sum_i m_h t^r(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta)) - \frac{1}{2} \sum_h \sum_i [m_h t^r(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))]^2 \\ &\quad + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\ &= nt^r(\theta)\bar{g}_{st}(\theta) - \frac{n}{2} t^r(\theta)\hat{S}_n(\theta)t(\theta) + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\ &= nt^r(\theta)[\bar{g}_{st}(\theta) - \frac{1}{2}\hat{S}_n(\theta)t(\theta)] + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\ &= \frac{n}{2} t^r(\theta)\bar{g}_{st}(\theta) + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\ &= \frac{n}{2} \bar{g}_{st}^r(\theta)\hat{S}_n^{-1}(\theta)\bar{g}_{st}(\theta) + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\ &= \frac{n}{2} \left[\bar{g}_{st}(\theta_0) + \sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta} un^{-\frac{1}{3}} \right]^r \hat{S}_n^{-1}(\theta) \\ &\quad \cdot \left[\bar{g}_{st}(\theta_0) + \sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta} un^{-\frac{1}{3}} \right] + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\ &= \frac{n}{2} \left[\sum_h W_h O(n_h^{-\frac{1}{2}}(\log \log n_h)^{\frac{1}{2}}) + \sum_h W_h E_h \frac{\partial g_{hi}(\theta_0)}{\partial \theta} un^{-\frac{1}{3}} \right]^r S^{-1}(\theta_0) \end{aligned}$$

$$\begin{aligned} & \cdot \left[\sum_h W_h O(n_h^{-\frac{1}{2}} (\log \log n_h)^{\frac{1}{2}}) + \sum_h W_h E_h \frac{\partial g_{hi}(\theta_0)}{\partial \theta} u n^{-\frac{1}{3}} \right] + o(n^{\frac{1}{3}}) \text{ a.s.} \\ & \geq (c - \varepsilon) n^{\frac{1}{3}} \text{ (a.s.)} \end{aligned}$$

where

$$\frac{1}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta} = O(n_h^{-\frac{1}{2}} (\log \log n_h)^{\frac{1}{2}}) \text{ (a.s.)}$$

by the law of iterative logarithm, $\varepsilon > 0$, $c - \varepsilon > 0$ and c is the smallest eigenvalue of

$$\frac{1}{2} \left[\sum_h W_h E_h \frac{\partial g_{hi}(\theta_0)}{\partial \theta} \right]^T S^{-1}(\theta_0) \left[\sum_h W_h E_h \frac{\partial g_{hi}(\theta_0)}{\partial \theta} \right]^T.$$

Similarly,

$$l_E(\theta_0) = \frac{n}{2} \bar{g}_{st}^T(\theta_0) \hat{S}_n^{-1}(\theta_0) \bar{g}_{st}(\theta_0) + o(n^{\frac{1}{3}}) \text{ a.s.} = o(n^{\frac{1}{3}}) \text{ a.s.}$$

Since $l_E(\theta)$ is a continuous function of θ belonging to the ball $\|\theta - \theta_0\| \leq n^{-\frac{1}{3}}$,

$l_E(\theta)$ has minimum value in the interior of this ball, and $\tilde{\theta}$ satisfies (noting (2.5) and (2.6))

$$\begin{aligned} 0 &= \frac{\partial l_E(\theta)}{\partial \theta^\tau} \Big|_{\theta=\tilde{\theta}} \\ &= \sum_h \sum_i \frac{m_h}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \left[\frac{\partial t^\tau(\theta)}{\partial \theta^\tau} (g_{hi}(\theta) - \hat{g}_h(\theta)) \right. \\ &\quad \left. + \left(\frac{\partial g_{hi}^T(\theta)}{\partial \theta^\tau} - \frac{\partial \hat{g}_h^T(\theta)}{\partial \theta^\tau} \right) t(\theta) \right] \Big|_{\theta=\tilde{\theta}} \\ &= \sum_h \sum_i \frac{m_h}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \frac{\partial g_{hi}^T(\theta)}{\partial \theta^\tau} t(\theta) \Big|_{\theta=\tilde{\theta}} \end{aligned}$$

i.e., $Q_{2n}(\tilde{\theta}, \tilde{t}) = 0$. $\square$

Proof of Theorem 1

Taking derivatives with respect to θ^τ , t^τ , we have:

$$\frac{\partial Q_{1n}(\theta_0, 0)}{\partial \theta^\tau} = \sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau}, \quad (5.10)$$

$$\frac{\partial Q_{1n}(\theta_0, 0)}{\partial t^\tau} = - \sum_h \frac{n W_h^2}{n_h^2} \sum_i (g_{hi}(\theta_0) - \bar{g}_h(\theta_0))(g_{hi}(\theta_0) - \bar{g}_h(\theta_0))^\tau = -S_n(\theta_0), \quad (5.11)$$

$$\frac{\partial Q_{2n}(\theta_0, 0)}{\partial \theta^\tau} = 0, \quad \frac{\partial Q_{2n}(\theta_0, 0)}{\partial t^\tau} = \sum_h W_h n_h^{-1} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta}. \quad (5.12)$$

Expanding $Q_{1n}(\tilde{\theta}, \tilde{t})$, $Q_{2n}(\tilde{\theta}, \tilde{t})$ around $(\theta, 0)$, we have

$$\begin{aligned} 0 &= Q_{in}(\tilde{\theta}, \tilde{t}) \\ &= Q_{in}(\theta_0, 0) + \frac{\partial Q_{in}(\theta_0, 0)}{\partial \theta^\tau} (\tilde{\theta} - \theta_0) + \frac{\partial Q_{in}(\theta_0, 0)}{\partial t^\tau} \tilde{t} + o_p(\delta_n), i = 1, 2 \end{aligned} \quad (5.13)$$

where $\delta_n = \|\tilde{\theta} - \theta_0\| + \|\tilde{t}\|$. Noting (2.14) (2.15), (5.10), (5.11), we have

$$\begin{aligned} 0 &= Q_{1n}(\tilde{\theta}, \tilde{t}) \\ &= \bar{g}_{st}(\theta_0) + \left[\sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau} \right] (\tilde{\theta} - \theta_0) - S_n(\theta_0)(\tilde{t} - 0) + o_p(\delta_n) \end{aligned} \quad (5.14)$$

$$0 = Q_{2n}(\tilde{\theta}, \tilde{t}) = \sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta} (\tilde{t} - 0) + o_p(\delta_n). \quad (5.15)$$

The system of equations (5.14) and (5.15) may be written in matrix form as

$$\begin{pmatrix} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix} = \begin{pmatrix} S_n(\theta_0) & M_{n,12} \\ M_{n,21} & 0 \end{pmatrix} \begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix}. \quad (5.16)$$

or

$$\begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} = M_n^{-1} \begin{pmatrix} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix}, \quad (5.17)$$

where M_n is the matrix of (5.16) with $M_{n,12} = - \sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau}$ and $M_{n,21} = M_{n,12}^\tau$. Since

$$S_n(\theta_0) \xrightarrow{p} \sum_h k_h W_h^2 \sigma_h(\theta_0) \triangleq M_{11}, \quad M_{n,12} \xrightarrow{p} - \sum_h W_h E_h \frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau} \triangleq M_{12},$$

it follows that

$$M_n \xrightarrow{p} M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

where $M_{21} = M_{12}^T$ and $M_{22} = 0$. Therefore,

$$\begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} = M^{-1} \begin{pmatrix} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix}. \quad (5.18)$$

From this result and

$$\bar{g}_{st}(\theta_0) = \sum_h \frac{W_h}{n_h} \sum_i g_{hi}(\theta_0) = \sum_h W_h O_p(n_h^{-\frac{1}{2}}) = (\sum_h W_h k_h^{\frac{1}{2}}) O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}),$$

we get $\delta_n = O_p(n^{-\frac{1}{2}})$.

Since

$$M^{-1} = \begin{pmatrix} I & -M_{11}^{-1}M_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & M_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -M_{21}M_{11}^{-1} & I \end{pmatrix}, \quad (5.19)$$

where $M_{22.1} = -M_{21}M_{11}^{-1}M_{12}$, we have

$$\begin{aligned} & \begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} \\ &= M^{-1} \begin{pmatrix} \bar{g}_{st}(\theta_0) \\ 0 \end{pmatrix} + o_p(\delta_n) \\ &= \begin{pmatrix} I & -M_{11}^{-1}M_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & M_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -M_{21}M_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} \bar{g}_{st}(\theta_0) \\ 0 \end{pmatrix} \\ & \quad + o_p(\delta_n) \\ &= \begin{pmatrix} M_{11}^{-1} & -M_{11}^{-1}M_{12}M_{22.1}^{-1} \\ 0 & M_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} \bar{g}_{st}(\theta_0) \\ -M_{21}M_{11}^{-1}\bar{g}_{st}(\theta_0) \end{pmatrix} + o_p(\delta_n) \end{aligned}$$

$$= \begin{pmatrix} U\bar{g}_{st}(\theta_0) \\ -M_{22.1}^{-1}M_{21}M_{11}^{-1}\bar{g}_{st}(\theta_0) \end{pmatrix} + o_p(\delta_n), \quad (5.20)$$

where $U = M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1})$. Hence,

$$\sqrt{n}(\tilde{\theta} - \theta_0) \rightarrow N(0, V),$$

where

$$\begin{aligned} V &= M_{22.1}^{-1}M_{21}M_{11}^{-1}\left(\sum_h W_h^2 k_h \sigma_h(\theta_0)\right)M_{11}^{-1}M_{12}M_{22.1}^{-1} \\ &= M_{22.1}^{-1}M_{21}M_{11}^{-1}M_{12}M_{22.1}^{-1} \\ &= (M_{21}M_{11}^{-1}M_{12})^{-1}, \end{aligned}$$

and

$$\sqrt{n}(\tilde{t} - 0) \rightarrow N(0, U),$$

where

$$\begin{aligned} &M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1})M_{11}M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1}) \\ &= M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1})^2 \\ &= M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1}) \\ &= U. \end{aligned}$$

Also, since

$$\begin{aligned} &-M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1})M_{12}M_{22.1}^{-1} \\ &= -(M_{11}^{-1}M_{12}M_{22.1}^{-1} + M_{11}^{-1}M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1}M_{12}M_{22.1}^{-1}) \\ &= 0, \end{aligned}$$

it follows that $\tilde{\theta}$ and $\tilde{t}$ are asymptotically independent.

Now we turn to the third result. Since

$$\begin{aligned}\tilde{p}_{hi} &= \frac{1}{n_h} \frac{1}{[1 + m_h \tilde{t}^r (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))]} \\ &= \frac{1}{n_h} \left[1 - m_h \tilde{t}^r (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) \right] + o_p(n^{-\frac{3}{2}}),\end{aligned}$$

noting (5.20) we have

$$\begin{aligned}\tilde{F}_n(x) &= \sum_h W_h \sum_i \tilde{p}_{hi} 1(x_{hi} < x) \\ &= \sum_h \frac{1}{n_h} \sum_i 1(x_{hi} < x) - \tilde{t}^r \sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) 1(x_{hi} < x) + o_p(n^{-\frac{1}{2}}) \\ &= \sum_h W_h \frac{1}{n_h} \sum_i 1(x_{hi} < x) - \left[\sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))^r 1(x_{hi} < x) \right] U \bar{g}_{st}(\theta_0) \\ &\quad + o_p(n^{-\frac{1}{2}}) \\ &= \sum_h W_h \frac{1}{n_h} \sum_i 1(x_{hi} < x) - B_n U \bar{g}_{st}(\theta_0) + o_p(n^{-\frac{1}{2}}) \\ &= \sum_h W_h \frac{1}{n_h} \sum_i 1(x_{hi} < x) - B U \bar{g}_{st}(\theta_0) + o_p(n^{-\frac{1}{2}}) \\ &= \sum_h W_h \frac{1}{n_h} \sum_i [1(x_{hi} < x) - B U (g_{hi}(\theta_0) - M_h(\theta_0))] + o_p(n^{-\frac{1}{2}}),\end{aligned}$$

where

$$B_n = \sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))^r 1(x_{hi} < x),$$

$$B = \sum_h W_h^2 k_h E_h[(g(x_h, \theta_0) - E_h g(x_h, \theta_0))^r 1(x_h < x)],$$

and $B_n \xrightarrow{p} B$. Hence

$$\sqrt{n}(\tilde{F}_n(x) - F(x)) \rightarrow N(0, W),$$

where

$$\begin{aligned}
 W &= \sum_h W_h^2 k_h \{ F_h(x) [1 - F_h(x)] - 2BU E_h[(g(x_h, \theta_0) - M_h(\theta_0))1(x_h < x)] \\
 &\quad + BU E_h(g(x_h, \theta_0) - M_h(\theta_0))(g(x_h, \theta_0) - M_h(\theta_0))^T U B^T \} \\
 &= \sum_h W_h^2 k_h F_h(x) [1 - F_h(x)] - BUB^T. \quad \square
 \end{aligned}$$

Proof of Theorem 2

Noting (5.20) and that $\tilde{t}, \tilde{\theta}$ are asymptotically independent, we have

$$\begin{aligned}
 l_E(\tilde{\theta}, \tilde{t}) &= \sum_h \sum_i \log[1 + m_h \tilde{t}^T (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))] \\
 &= \sum_h \sum_i m_h \tilde{t}^T (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) - \frac{1}{2} \sum_h \sum_i m_h^2 [\tilde{t}^T (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))]^2 + o_p(1) \\
 &= \tilde{t}^T \sum_h m_h \sum_i (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) \\
 &\quad - \frac{n}{2} \tilde{t}^T \sum_h n W_h^2 n_h^{-2} \sum_i (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))(g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))^T \tilde{t} + o_p(1) \\
 &= n \tilde{t}^T \bar{g}_{st}(\tilde{\theta}) - \frac{n}{2} \tilde{t}^T \hat{S}_n(\tilde{\theta}) \tilde{t} + o_p(1) \\
 &= n \tilde{t}^T \bar{g}_{st}(\theta_0) + n \tilde{t}^T \left[\sum_h W_h \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau} \right] (\tilde{\theta} - \theta_0) - \frac{n}{2} \tilde{t}^T S(\theta_0) \tilde{t} + o_p(1) \\
 &= n \bar{g}_{st}^T(\theta_0) U \bar{g}_{st}(\theta_0) - \frac{n}{2} \bar{g}_{st}^T(\theta_0) U S(\theta_0) U \bar{g}_{st}(\theta_0) + o_p(1) \\
 &= \frac{n}{2} \bar{g}_{st}^T(\theta_0) U \bar{g}_{st}(\theta_0) + o_p(1).
 \end{aligned}$$

Similarly, we get

$$\begin{aligned}
 l_E(\theta_0, t(\theta_0)) &= \sum_h \sum_i \log[1 + m_h t^T(\theta_0)(g_{hi}(\theta_0) - \hat{g}_h(\theta_0))] \\
 &= \frac{n}{2} \bar{g}_{st}^T(\theta_0) S^{-1}(\theta_0) \bar{g}_{st}(\theta_0) + o_p(1).
 \end{aligned}$$

Hence

$$R = n \bar{g}_{st}^T(\theta_0) (S^{-1}(\theta_0) - U) \bar{g}_{st}(\theta_0) + o_p(1)$$

$$\begin{aligned}
&= -n\bar{g}_{st}^r(\theta_0)(M_{11}^{-1}M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1})\bar{g}_{st}(\theta_0) + o_p(1) \\
&= (\sqrt{n}M_{11}^{-\frac{1}{2}}\bar{g}_{st}(\theta_0))^r(-M_{11}^{-\frac{1}{2}}M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-\frac{1}{2}})(\sqrt{n}M_{11}^{-\frac{1}{2}}\bar{g}_{st}(\theta_0)) + o_p(1).
\end{aligned}$$

Now noting that $\sqrt{n}M_{11}^{-\frac{1}{2}}\bar{g}_{st}(\theta_0) \rightarrow N(0, I)$ and $-M_{11}^{-\frac{1}{2}}M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-\frac{1}{2}}$ is symmetric and idempotent with trace equal to p , it follows that R converges to χ_p^2 . $\square$

Proof of Corollary 1

From the proof of theorem 2.2, we have

$$\begin{aligned}
R_1 &= 2l_E(\tilde{\theta}, \tilde{t}) \\
&= 2 \sum_h \sum_i \log[1 + m_h \tilde{t}^r g_{hi}(\tilde{\theta})] \\
&= n\bar{g}_{st}^r(\theta_0)U\bar{g}_{st}(\theta_0) + o_p(1) \\
&= (\sqrt{n}M_{11}^{-\frac{1}{2}}\bar{g}_{st}(\theta_0))^r M_{11}^{\frac{1}{2}}U M_{11}^{\frac{1}{2}}(\sqrt{n}M_{11}^{-\frac{1}{2}}\bar{g}_{st}(\theta_0)) + o_p(1) \\
&\rightarrow \chi_{r-p}^2.
\end{aligned}$$

Proof of Lemma 6

For fixed θ such that $\|\theta - \theta_0\| < d_n$, consider the function

$$\psi(\lambda) = \sum_h \frac{W_h}{n_h} \sum_i \frac{g_{hi}(\theta) - \hat{g}_h(\theta)}{1 + n^{-\frac{1}{3}-\frac{\delta}{2}}m_h\lambda^r[g_{hi}(\theta) - \hat{g}_h(\theta)]}.$$

From $E\|g(x, \theta)\|^3 < \infty$ we get $\max_h \max_i \|g(x_{hi}, \theta)\| = o(n^{\frac{1}{3}})$ a.s., and $\psi(\lambda)$ is an almost surely continuous function for $\|\lambda\| \leq 1$. When $\|\lambda\| = 1$, we have

$$\begin{aligned}
\lambda^r \psi(\lambda) &= \lambda^r \sum_h \frac{W_h}{n_h} \sum_i (g_{hi}(\theta) - \hat{g}_h(\theta)) \\
&\quad - n^{-\frac{1}{3}-\frac{\delta}{2}} \lambda^r \sum_h n W_h^2 n_h^{-2} \sum_i [g_{hi}(\theta) - \hat{g}_h(\theta)][g_{hi}(\theta) - \hat{g}_h(\theta)]^r \lambda \\
&\quad + O(\lambda^r \sum_h \frac{W_h}{n_h} m_h^2 n^{-\frac{2}{3}-\delta} \sum_i [g_{hi}(\theta) - \hat{g}_h(\theta)] \\
&\quad \cdot \lambda^r [g_{hi}(\theta) - \hat{g}_h(\theta)][g_{hi}(\theta) - \hat{g}_h(\theta)]^r \lambda \text{ a.s.}
\end{aligned}$$

$$\begin{aligned}
&\leq O(n^{-\frac{1}{3}-\delta} - \frac{c}{2}n^{-\frac{1}{3}-\frac{\delta}{2}}) + O(n^{-\frac{1}{3}-\delta}) \text{ a.s.} \\
&< 0, \text{ a.s.}
\end{aligned}$$

where c is the smallest eigenvalue of $\sum_h k_h W_h^2 \sigma_h(\theta_0)$. By Lemma 5 there exists a point $\hat{\lambda}$ such that $\|\hat{\lambda}\| < 1$ and $\psi(\hat{\lambda}) = 0$. Using the implicit-function theorem, we can easily get other theorem's results. $\square$

Proof of Lemma 7

When $\|\theta - \theta_0\| \leq d_n$, from

$$\sum_h \frac{W_h}{n_h} \sum_i \frac{g_{hi}(\theta) - \hat{g}_h(\theta)}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]} = 0,$$

we have

$$\begin{aligned}
0 &= \sum_h \frac{W_h}{n_h} \sum_i \frac{\frac{\partial g_{hi}(\theta)}{\partial \theta^\tau}}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]} \\
&\quad - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{[g_{hi}(\theta) - \hat{g}_h(\theta)][g_{hi}(\theta) - \hat{g}_h(\theta)]^\tau \frac{\partial t}{\partial \theta^\tau}}{[1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))]^2} \\
&\quad - \sum_h n W_h^2 n_h^{-2} \sum_i \frac{g_{hi}(\theta) - \hat{g}_h(\theta)}{[1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))]^2} \left[\frac{\partial g_{hi}(\theta)}{\partial \theta^\tau} - \frac{\partial \hat{g}_h(\theta)}{\partial \theta^\tau} \right]^\tau t \text{ a.s.} \\
&= D(\theta_0) - S(\theta_0) \frac{\partial t}{\partial \theta^\tau} + O(d_n), \text{ a.s.}
\end{aligned}$$

where $D(\theta_0) = \sum_h W_h E_h \frac{\partial g(x_h, \theta_0)}{\partial \theta^\tau}$, $S(\theta_0) = \sum_h k_h W_h^2 \sigma_h(\theta_0)$, hence

$$\frac{\partial t}{\partial \theta^\tau} = S^{-1}(\theta_0) D(\theta_0) + O(d_n). \text{ a.s.}$$

Denote

$$W_n(\theta) = \left[\sum_h \frac{W_h}{n_h} \sum_i \frac{(\partial g(x_{hi}, \theta) / \partial \theta^\tau)^\tau}{1 + m_h t^\tau (\theta) (g(x_{hi}, \theta) - \hat{g}_h(\theta))} \right] t(\theta) = Z_n(\theta) t(\theta),$$

then

$$W_n(\theta) = W_n(\theta_0) + \left(Z_n(\theta_0) \frac{\partial t(\theta_0)}{\partial \theta^\tau} + \frac{\partial Z_n(\theta_0)}{\partial \theta^\tau} t(\theta_0) \right) (\theta - \theta_0) + O(\|\theta - \theta_0\|)^2$$

$$\begin{aligned}
&= O(n^{-\frac{1}{2}} \log \log n) + \left(D^T(\theta_0) S^{-1}(\theta_0) D(\theta_0) + O(n^{-\frac{1}{2}} \log \log n) \right) (\theta - \theta_0) \\
&\quad + O(\|\theta - \theta_0\|)^2 \\
&= D^T(\theta_0) S^{-1}(\theta_0) D(\theta_0) (\theta - \theta_0) + o(\|\theta - \theta_0\|). \text{ a.s.}
\end{aligned}$$

That is, $W_n(\theta)$ can be represented as

$$W_n(\theta) = D^T(\theta_0) S^{-1}(\theta_0) D(\theta_0) (\theta - \theta_0) + V(x, \theta),$$

where $V(x, \theta)$ is almost surely continuous function of θ and $V(x, \theta) = o(\|\theta - \theta_0\|)$ a.s.. Similar to the arguments in Aitchison and Silvey (1958), we can show that the system of equations $Q_{2n}(\theta, t(\theta), \nu) = 0$ and $Q_{3n}(\theta) = 0$ have a solution $\tilde{\theta}, \tilde{\nu}$ such that $\tilde{\theta}, \tilde{\nu} \in \{(\theta, \nu) : \|\theta - \theta_0\| + \|\nu\| \leq d_n\}$.

Now we prove the last part of Lemma 7. Suppose that $\tilde{\theta}$ is a solution in U_{d_n} , θ is a point in U_{d_n} and satisfies $r(\theta) = 0$. Then we have

$$l_E(\theta) - l_E(\tilde{\theta}) = \frac{\partial l_E(\tilde{\theta})}{\partial \theta^\tau} (\theta - \tilde{\theta}) + \frac{1}{2} (\theta - \tilde{\theta})^\tau \frac{\partial^2 l_E(\theta^*)}{\partial \theta \partial \theta^\tau} (\theta - \tilde{\theta}), \quad \theta^* \in U_{d_n}.$$

Note that

$$\begin{aligned}
\frac{\partial l_E(\theta)}{\partial \theta^\tau} &= \sum_h \sum_i \frac{m_h}{1 + m_h t^\tau(\theta) (g_{hi}(\theta) - \hat{g}_h(\theta))} \left(\frac{\partial t(\theta)}{\partial \theta^\tau} \right)^\tau (g_{hi}(\theta) - \hat{g}_h(\theta)) \\
&\quad + \sum_h \sum_i \frac{m_h}{1 + m_h t^\tau(\theta) (g_{hi}(\theta) - \hat{g}_h(\theta))} t^\tau(\theta) \left(\frac{\partial g_{hi}(\theta)}{\partial \theta^\tau} - \frac{\partial \hat{g}_h(\theta)}{\partial \theta^\tau} \right) \\
&= t^\tau(\theta) \sum_h \sum_i \frac{m_h}{1 + m_h t^\tau(\theta) (g_{hi}(\theta) - \hat{g}_h(\theta))} \frac{\partial g_{hi}(\theta)}{\partial \theta^\tau} = n t^\tau(\theta) Z_n^\tau(\theta)
\end{aligned}$$

and $Q_{2n}(\tilde{\theta}, t(\tilde{\theta}), \tilde{\nu}) = 0$, so that $t^\tau(\tilde{\theta}) Z_n^\tau(\tilde{\theta}) = -\tilde{\nu}^\tau R(\tilde{\theta})$ from (3.2). Since the $q \times p$ matrix $R(\theta_0)$ is of rank q , we have

$$\tilde{\nu} = O(t(\tilde{\theta})) = O(\|\tilde{\theta} - \theta_0\|) \text{ a.s.}$$

From

$$0 = r(\theta) - r(\tilde{\theta}) = \frac{\partial r(\tilde{\theta})}{\partial \theta^\tau}(\theta - \tilde{\theta}) + o(\|\theta - \tilde{\theta}\|) \text{ a.s.}$$

we have $R(\tilde{\theta})(\theta - \tilde{\theta}) = o(\|\theta - \tilde{\theta}\|) \text{ a.s.}$, i.e.,

$$\frac{\partial l_E(\tilde{\theta})}{\partial \theta^\tau}(\theta - \tilde{\theta}) = nt^\tau(\tilde{\theta})Z_n^\tau(\tilde{\theta})(\theta - \tilde{\theta}) = -n\tilde{\nu}^\tau R(\tilde{\theta})(\theta - \tilde{\theta}) = no(\|\theta - \tilde{\theta}\|^2) \text{ a.s.}$$

$$\begin{aligned} \frac{\partial^2 l_E(\theta^*)}{\partial \theta \partial \theta^\tau} &= n \frac{\partial}{\partial \theta} [t^\tau(\theta^*)Z_n^\tau(\theta^*)] = n \left[\frac{\partial t^\tau(\theta^*)}{\partial \theta} Z_n^\tau(\theta^*) + t^\tau(\theta^*) \frac{\partial Z_n^\tau(\theta^*)}{\partial \theta} \right] \\ &= n[D(\theta_0)S_n^{-1}(\theta_0)D^\tau(\theta_0)] + O(d_n) \\ &> \frac{c}{2}n + O(d_n), \end{aligned}$$

where c is the smallest eigenvalue of $D(\theta_0)S_n^{-1}(\theta_0)D^\tau(\theta_0)$. Hence

$$l_E(\theta) - l_E(\tilde{\theta}) = no(\|\theta - \tilde{\theta}\|^2) + cn \|\theta - \tilde{\theta}\|^2 > 0 \text{ a.s.}$$

Proof of theorem 3

Taking derivatives about θ^τ , t^τ , ν^τ , we have

$$\begin{aligned} \frac{\partial Q_{1n}(\theta_0, 0, 0)}{\partial \theta^\tau} &= \sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau}, \quad \frac{\partial Q_{1n}(\theta_0, 0, 0)}{\partial t^\tau} = -S_n(\theta_0), \\ \frac{\partial Q_{1n}(\theta_0, 0, 0)}{\partial \nu^\tau} &= 0, \quad \frac{\partial Q_{2n}(\theta_0, 0, 0)}{\partial \theta^\tau} = 0, \quad \frac{\partial Q_{2n}(\theta_0, 0, 0)}{\partial t^\tau} = \sum_h \frac{W_h}{n_h} \sum_i \left(\frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau} \right)^\tau, \\ \frac{\partial Q_{2n}(\theta_0, 0, 0)}{\partial \nu^\tau} &= R^\tau(\theta), \quad \frac{\partial Q_{3n}(\theta_0, 0, 0)}{\partial \theta^\tau} = R(\theta), \quad \frac{\partial Q_{3n}(\theta_0, 0, 0)}{\partial t^\tau} = 0, \\ \frac{\partial Q_{3n}(\theta_0, 0, 0)}{\partial \nu^\tau} &= 0. \end{aligned}$$

Expanding $Q_{1n}(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r)$, $Q_{2n}(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r)$, $Q_{3n}(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r)$ at $(\theta, 0, 0)$, we have

$$\begin{aligned} 0 &= Q_{1n}(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r) \\ &= \bar{g}_{st}(\theta_0) + D_n(\theta_0)(\tilde{\theta}_r - \theta_0) - S_n(\theta_0) \tilde{t}_r + o_p(\delta_n), \end{aligned} \tag{5.21}$$

$$0 = Q_{2n}(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r) = D_n^r(\theta_0) \tilde{t}_r + R^r(\theta_0) \tilde{\nu}_r + o_p(\delta_n), \quad (5.22)$$

$$0 = Q_{3n}(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r) = R(\theta_0) \tilde{\nu}_r + o_p(\delta_n), \quad (5.23)$$

where $\delta_n = \|\tilde{\theta}_r - \theta_0\| + \|\tilde{t}_r\| + \|\tilde{\nu}_r\|$, $D_n(\theta_0) = \sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta^r}$. The system of equations (5.21) (5.22) and (5.23) written in matrix form is

$$\begin{pmatrix} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix} = \begin{pmatrix} S_n(\theta_0) & -D_n(\theta_0) & 0 \\ -D_n^r(\theta_0) & 0 & R^r(\theta_0) \\ 0 & R(\theta_0) & 0 \end{pmatrix} \begin{pmatrix} \tilde{t}_r \\ \tilde{\theta}_r - \theta_0 \\ \tilde{\nu}_r \end{pmatrix}.$$

or

$$\begin{pmatrix} \tilde{t} \\ \tilde{\xi}_r \end{pmatrix} = \begin{pmatrix} M_{11} & -M_{12} \\ -M_{21} & M_{22} \end{pmatrix}^{-1} \begin{pmatrix} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix}, \quad (5.24)$$

where

$$M_{11} = \sum_h k_h W_h^2 \sigma_h(\theta_0), \quad M_{12} = (D(\theta_0), 0), \quad D(\theta_0) = \sum_h W_h E_h \frac{\partial g(x_h, \theta_0)}{\partial \theta^r},$$

$$M_{22} = \begin{pmatrix} 0 & R^r(\theta_0) \\ R(\theta_0) & 0 \end{pmatrix}, \quad M_{21} = M_{12}. \quad \text{From this and}$$

$$\bar{g}_{st}(\theta_0) = \sum_h \frac{W_h}{n_h} \sum_i g_{hi}(\theta_0) = \sum_h W_h O_p(n_h^{-\frac{1}{2}}) = (\sum_h W_h k_h^{\frac{1}{2}}) O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}).$$

we know that $\delta_n = O_p(n^{-\frac{1}{2}})$. Since

$$\tilde{\xi}_r = (0, I) \begin{pmatrix} \tilde{t}_r \\ \tilde{\xi}_r \end{pmatrix} = (0, I) \begin{pmatrix} M_{11} & -M_{12} \\ -M_{21} & M_{22} \end{pmatrix}^{-1} \begin{pmatrix} \bar{g}_{st}(\theta_0) \\ 0 \end{pmatrix} + o_p(\delta_n),$$

we know that

$$\tilde{\xi}_r = M_{22.1}^{-1} M_{21} M_{11}^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n), \quad (5.25)$$

where $M_{22.1} = M_{22} - M_{21}M_{11}^{-1}M_{12}$. Or

$$\begin{aligned}\tilde{\xi}_r &= M_{22.1}^{-1} \begin{pmatrix} D^r(\theta_0) \\ 0 \end{pmatrix} M_{11}^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ &= M_{22.1}^{-1} \begin{pmatrix} D^r(\theta_0) M_{11}^{-1} \bar{g}_{st}(\theta_0) \\ 0 \end{pmatrix} + o_p(\delta_n).\end{aligned}\tag{5.26}$$

Since

$$\begin{aligned}M_{22.1} &= M_{22} - M_{21}M_{11}^{-1}M_{12} \\ &= \begin{pmatrix} 0 & R^r \\ R & 0 \end{pmatrix} - \begin{pmatrix} D^r \\ 0 \end{pmatrix} M_{11}^{-1}(D, 0) \\ &= \begin{pmatrix} -V & R^r \\ R & 0 \end{pmatrix},\end{aligned}$$

where $R = R(\theta_0)$, $D = D(\theta_0)$, $V = D^r M_{11}^{-1} D$, we have

$$\begin{aligned}M_{22.1}^{-1} &= \begin{pmatrix} -V & R^r \\ R & 0 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} I & V^{-1}R^r \\ 0 & I \end{pmatrix} \begin{pmatrix} -V^{-1} & 0 \\ 0 & (RV^{-1}R^r)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ RV^{-1} & I \end{pmatrix} \\ &= \begin{pmatrix} -P & Q^r \\ Q & H \end{pmatrix},\end{aligned}$$

where

$$H = (RV^{-1}R^r)^{-1}, \quad Q = HRV^{-1}, \quad P = V^{-1}(I - R^rQ).\tag{5.27}$$

Hence using the central limit theorem for $\sqrt{n}\bar{g}_{st}(\theta_0)$ and Slutsky's theorem, we can

easily get

$$\begin{aligned}\sqrt{n} \tilde{\xi}_r &= \sqrt{n} M_{22.1}^{-1} M_{21} M_{11}^{-1} \bar{g}_{st}(\theta_0) \rightarrow N \left(0, M_{22.1}^{-1} \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} M_{22.1}^{-1} \right) \\ &= N \left(0, \begin{pmatrix} PVP & -PVQ^r \\ -QVP & QVQ^r \end{pmatrix} \right) = N \left(0, \begin{pmatrix} P & 0 \\ 0 & H \end{pmatrix} \right). \quad \square\end{aligned}$$

Proof of Theorem 4

From (5.26) we know that

$$\tilde{\xi}_r = \begin{pmatrix} -P & Q^r \\ Q & H \end{pmatrix} \begin{pmatrix} D^r M_{11}^{-1} \bar{g}_{st}(\theta_0) \\ 0 \end{pmatrix} + o_p(\delta_n) = \begin{pmatrix} -PD^r M_{11}^{-1} \bar{g}_{st}(\theta_0) \\ QD^r M_{11}^{-1} \bar{g}_{st}(\theta_0) \end{pmatrix} + o_p(\delta_n).$$

Hence,

$$\begin{aligned}\tilde{\nu}_r &= QD^r M_{11}^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n) = HRV^{-1}D^r M_{11}^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ &= HR(D^r M_{11}^{-1} D)^{-1} D^r M_{11}^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ &= HRD^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n).\end{aligned}\tag{5.28}$$

$$\begin{aligned}\tilde{\theta}_r - \theta_0 &= -PD^r M_{11}^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n) = -PVD^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ &= V^{-1}(-I + R^r Q)VD^{-1} \bar{g}_{st}(\theta_0) + o_p(\delta_n) \\ &= -D^{-1} \bar{g}_{st}(\theta_0) + V^{-1}R^r(QVD^{-1} \bar{g}_{st}(\theta_0)) + o_p(\delta_n) \\ &= -D^{-1} \bar{g}_{st}(\theta_0) + V^{-1}R^r(QD^r M_{11}^{-1} \bar{g}_{st}(\theta_0)) + o_p(\delta_n) \\ &= D^{-1} \bar{g}_{st}(\theta_0) + V^{-1}R^r \tilde{\nu}_r + o_p(\delta_n).\end{aligned}\tag{5.29}$$

Therefore,

$$\begin{aligned}\sum_h \frac{W_h}{n_h} \sum_i g(x_{hi}, \tilde{\theta}_r) &= \bar{g}_{st}(\theta_0) + \sum_h \frac{W_h}{n_h} \sum_i \frac{\partial g(x_{hi}, \theta_0)}{\partial \theta^r} (\tilde{\theta}_r - \theta_0) + o_p(\delta_n) \\ &= \bar{g}_{st}(\theta_0) + D(\tilde{\theta}_r - \theta_0) + o_p(\delta_n) \\ &= DV^{-1}R^r \tilde{\nu}_r + o_p(\delta_n).\end{aligned}\tag{5.30}$$

By $Q_{1n}(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r) = 0$, we know that $\sum_h W_h Q_{1n}(\tilde{\theta}_r, \tilde{t}_r, \tilde{\nu}_r) = 0$, i.e.,

$$\sum_h \frac{W_h}{n_h} \sum_i \frac{g(x_{hi}, \tilde{\theta}_r) - \hat{g}_h(\tilde{\theta}_r)}{1 + m_h \tilde{t}_r^\tau [g(x_{hi}, \tilde{\theta}_r) - \hat{g}_h(\tilde{\theta}_r)]} = 0$$

or

$$\sum_h \frac{W_h}{n_h} \sum_i [g(x_{hi}, \tilde{\theta}_r) - \hat{g}_h(\tilde{\theta}_r)] - \hat{S}_n(\tilde{\theta}_r) \tilde{t}_r + o_p(\delta_n) = 0.$$

We have

$$\tilde{t}_r = \hat{S}_n^{-1}(\tilde{\theta}_r) \sum_h W_h n_h^{-1} \sum_i g(x_{hi}, \tilde{\theta}_r) + o_p(\delta_n) = M_{11}^{-1} \bar{g}_{st}(\tilde{\theta}_r) + o_p(\delta_n). \quad (5.31)$$

Therefore, noting (5.30),

$$\begin{aligned} 2l_E(\tilde{\theta}_r) &= 2 \sum_h \sum_i \log \{1 + m_h \tilde{t}_r^\tau [g(x_{hi}, \tilde{\theta}_r) - \hat{g}_h(\tilde{\theta}_r)]\} \\ &= 2 \sum_h \sum_i m_h \tilde{t}_r^\tau [g(x_{hi}, \tilde{\theta}_r) - \hat{g}_h(\tilde{\theta}_r)] - n \tilde{t}_r^\tau \hat{S}_n(\tilde{\theta}_r) \tilde{t}_r + o_p(1) \\ &= 2n \tilde{t}_r^\tau \bar{g}_{st}(\tilde{\theta}_r) - n \tilde{t}_r^\tau \bar{g}_{st}(\tilde{\theta}_r) + o_p(1) \\ &= n \bar{g}_{st}^\tau(\tilde{\theta}_r) M_{11}^{-1} \bar{g}_{st}(\tilde{\theta}_r) + o_p(1) \\ &= n \tilde{\nu}_r^\tau R V^{-1} D^\tau M_{11}^{-1} D V^{-1} R^\tau \tilde{\nu}_r + o_p(1) \\ &= n \tilde{\nu}_r^\tau R V^{-1} R^\tau \tilde{\nu}_r + o_p(1) = n \tilde{\nu}_r^\tau H^{-1} \tilde{\nu}_r + o_p(1) \\ &= n \tilde{\nu}_r^\tau H_{\tilde{\theta}_r}^{-1} \tilde{\nu}_r + o_p(1) = LA + o_p(1) \end{aligned}$$

Since

$$\begin{aligned} 0 &= \sum_h W_h n_h^{-1} \sum_i g(x_{hi}, \hat{\theta}) \\ &= \sum_h W_h n_h^{-1} \sum_i g_{hi}(\theta_0) + D(\hat{\theta} - \theta_0) + o_p(n^{-\frac{1}{2}}) \end{aligned}$$

we know, noting (5.28),

$$\hat{\theta} - \theta_0 = -D^{-1} \bar{g}_{st}(\theta_0) + o_p(n^{-\frac{1}{2}})$$

$$\begin{aligned}
r(\hat{\theta}) &= r(\theta_0) + R(\theta_0)(\hat{\theta} - \theta_0) + o_p(n^{-\frac{1}{2}}) \\
&= -R(\theta_0)D^{-1}\bar{g}_{st}(\theta_0) + o_p(n^{-\frac{1}{2}}) = -H^{-1}\tilde{\nu}_r + o_p(n^{-\frac{1}{2}}).
\end{aligned}$$

Hence

$$WA = n\tilde{\nu}_r^r H^{-1}HH^{-1}\tilde{\nu}_r + o_p(1) = n\tilde{\nu}_r^r H^{-1}\tilde{\nu}_r + o_p(1) = LA + o_p(1). \square$$

Chapter 4

Pseudo Empirical Likelihood and General Estimating Equations Using Stratified Sampling

4.1 Introduction

We applied empirical likelihood(EL) to stratified simple random sampling and established asymptotic properties of EL estimators and likelihood ratio tests in chapters 2 and 3. The purpose of this chapter is to use empirical likelihood and estimating equations in more complex sampling situations. Sample surveys often use a combination of several of the following sampling methods: stratified sampling, cluster sampling, unequal probability sampling and multi-stage sampling. In this chapter, we will consider stratified sampling again, but samples taken from each stratum can be complex.

Section 4.2 contains the definition of pseudo empirical likelihood, ratio statistic, estimation equations and the maximum pseudo empirical likelihood ratio (MPELR) estimator. The existence of these estimators and the asymptotic properties of these estimators will be established in Section 6.3. An illustration example will be given in Section 6.4. Proofs are delegated to Section 6.5.

4.2 MPELR Estimators

Let $\{P_\nu, \nu = 1, 2, \dots\}$ be a sequence of finite populations. Throughout the paper ν is used as the index of the finite population, but it may be omitted frequently for simplicity. Each P_ν contains H strata and the h^{th} stratum contains N_h units. Associated with the i^{th} unit of h^{th} stratum is the characteristic x_{hi} , $h = 1, \dots, H$, $i = 1, \dots, N_h$, where x_{hi} is d -dimensional. Note that H , N_h , x_{hi} etc. depend on ν also but the subscript ν is omitted. Suppose P_ν with d -dimension characteristic x_ν has the unknown distribution F_ν and a p -dimension parameter θ associated with it, and the h^{th} stratum of P_ν has d -dimension characteristic $x_{\nu h}$ which has the unknown distribution $F_{\nu h}$. We will omit ν also from F_ν , $F_{\nu h}$, $x_{\nu h}$ in the following discussion when appropriate. We assume that the weight of the h^{th} stratum of P_ν , W_h , is known for every h .

A sample, s_h , with values $x_{h1}, \dots, x_{hN_h}$ from stratum h , is taken according to some sampling scheme and samples are independent across strata. From the context we know that

$$F = \sum_h W_h F_h. \quad (2.1)$$

Suppose $x_{h1}, \dots, x_{hN_h}$, units in h^{th} stratum of P_ν , is a random sample from a super-

population, say $F_{\nu h}^s$. If the entire population P_ν were available, the corresponding likelihood function would be

$$L(F) = \prod_{h=1}^H \prod_{i=1}^{N_h} p_{hi}, \quad (2.2)$$

where $p_{hi} = F_{\nu h}^s(x_{hi}) - F_{\nu h}^s(x_{hi}^-)$. The log-likelihood function then is

$$l(F) = \sum_{h=1}^H \sum_{i=1}^{N_h} \log p_{hi}. \quad (2.3)$$

The nonparametric maximum likelihood estimator of F_ν^s is

$$F_N(x) = N^{-1} \sum_{h=1}^H \sum_{i=1}^{N_h} 1(x_{hi} < x),$$

where $1(x_{hi} < x) = (1(x_{hi,1} < x_1), \dots, 1(x_{hi,d} < x_d))^r$, $x_{hi,j}$ and x_j are j -th component of vector x_{hi} and x respectively, $1(x_{hi,j} < x_j)$ is the indicator of set $(x_{hi,j} < x_j)$. $N = \sum_{h=1}^H N_h$. But we only have a sample, say, s , from the finite population and if we view (2.3) as a finite population total, then we may have available a design unbiased estimator of $l(F)$, namely,

$$\hat{l}(F) = N \sum_h \sum_{i \in s_h} d_{hi}(s) \log p_{hi}, \quad (2.4)$$

where $d_{hi}(s)$ are design weights, with $E[\sum_{i \in s_h} d_{hi}(s) \log p_{hi}] = N^{-1} \sum_{i=1}^{N_h} \log p_{hi}$ for all h , and E refers to expectation with respect to the sampling design. Obviously we have $E(\hat{l}(F)) = l(F)$. We will call (2.4) the ‘‘pseudo empirical likelihood’’. Chen and Sitter(1996) defined pseudo empirical likelihood and we use a similar definition here. Some examples about how to construct weight $d_{hi}(s)$ are given in Section 4.3.

Since $\sum_i p_{hi} = 1$, we know (2.4) is maximized by $F_n(x) = \sum_h W_h F_{n_h}(x)$, where $n = \sum_h n_h$, $F_{n_h}(x) = \sum_{i \in s_h} \omega_{hi}(s) 1(x_{hi} < x)$, $\omega_{hi}(s) = d_{hi}(s) / \sum_{i \in s_h} d_{hi}(s)$. The

log-empirical likelihood ratio is then defined as

$$\begin{aligned}\log \hat{R}(F) &= \log \hat{l}(F) - \log \hat{l}(F_n) \\ &= N \left(\sum_h \sum_{i \in s_h} d_{hi}(s) \log p_{hi} - \sum_h \sum_{i \in s_h} d_{hi}(s) \log \omega_{hi}(s) \right). \quad (2.5)\end{aligned}$$

We also assume that information about θ and F is available in the form of $r \geq p$ functionally independent unbiased estimation functions, that is functions $g_j(x, \theta), j = 1, \dots, r$ such that $E_F\{g_j(x, \theta)\} = 0$. In vector form, that is

$$g(x, \theta) = (g_1(x, \theta), \dots, g_r(x, \theta))^T,$$

where $g(x, \theta)$ satisfies

$$E_F\{g(x, \theta)\} = \sum_h W_h E_{F_h}\{g(x_h, \theta)\} = 0. \quad (2.6)$$

Hence, we define the (negative) log-pseudo empirical likelihood ratio function for θ by

$$l_E(\theta) = -\frac{1}{N} \sup\{\log \hat{R}(F) | p_{hi} \geq 0, \sum_i p_{hi} = 1, \sum_h W_h \sum_i p_{hi} g_{hi}(\theta) = 0\}, \quad (2.7)$$

where $g_{hi}(\theta) = g(x_{hi}, \theta)$ for all h and i . The maximum may be found via Lagrange multiplier method for any given θ . Let

$$G = \sum_h \sum_i d_{hi}(s) \log p_{hi} - \sum_h \lambda_h (\sum_i p_{hi} - 1) - t^T (\sum_h W_h \sum_i p_{hi} g_{hi}(\theta)),$$

where t is a r -dimension vector. Then

$$\frac{\partial G}{\partial p_{hi}} = d_{hi}(s) p_{hi}^{-1} - \lambda_h - W_h t^T g_{hi}(\theta) = 0, \quad i = 1, \dots, n_h; h = 1, \dots, H,$$

$$\sum_i p_{hi} \frac{\partial G}{\partial p_{hi}} = \sum_i d_{hi}(s) - \lambda_h - W_h t^T \sum_i p_{hi} g_{hi}(\theta) = 0$$

$$\Rightarrow \lambda_h = \sum_i d_{hi}(s) - W_h t^\tau \hat{g}_h(\theta),$$

where $\hat{g}_h(\theta) = \sum_i p_{hi} g_{hi}(\theta)$. Hence, letting $m_h = W_h / \sum_i d_{hi}(s)$,

$$p_{hi} = \frac{\omega_{hi}(s)}{1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))}, \quad \omega_{hi}(s) = \frac{d_{hi}(s)}{\sum_i d_{hi}(s)}, \quad (2.8)$$

$$\hat{g}_h(\theta) = \sum_i p_{hi} g_{hi}(\theta), \quad (2.9)$$

with the restriction from $\frac{\partial G}{\partial t} = 0$, i.e.,

$$\sum_h W_h \sum_i p_{hi} g_{hi}(\theta) = 0,$$

or

$$\begin{aligned} Q_{1n}(\theta, t) &= \sum_h W_h \sum_i \frac{\omega_{hi}(s) g_{hi}(\theta)}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]} \\ &= \sum_h W_h \sum_i \frac{\omega_{hi}(s) [g_{hi}(\theta) - \hat{g}_h(\theta)]}{1 + m_h t^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]} = 0. \end{aligned} \quad (2.10)$$

Therefore,

$$l_E(\theta) = \sum_h \sum_i d_{hi}(s) \log[1 + m_h t^\tau (g_{hi}(\theta) - \hat{g}_h(\theta))], \quad (2.11)$$

with t being the solution to (2.10). We may minimize $l_E(\theta)$ to obtain an estimator $\tilde{\theta}$ of the parameter θ (called MPELR estimator). In addition, this yields estimators $\tilde{p}_{hi}$'s from (2.4), and therefore an estimator for distribution function F as

$$\tilde{F}_n(x) = \sum_h W_h \sum_i \tilde{p}_{hi} 1(x_{hi} < x) \quad (2.12)$$

Note that if there is no auxiliary information, i.e., $g(x, \theta) = 0$, this approach yields $\hat{p}_{hi} = d_{hi}(s) / \sum_{i \in s_h} d_{hi}(s)$. Then the estimator of $\bar{X}_N = \sum_h W_h \bar{X}_{N_h}$ will be $\sum_h \sum_i \hat{p}_{hi} x_{hi} = \sum_h W_h (\sum_i d_{hi}(s) x_{hi}) / \sum_i d_{hi}(s)$, a separate ratio estimator.

The existence of solutions to (2.8)-(2.10) and $\tilde{\theta}$ will be discussed in the following section.

4.3 Asymptotic Results

Some conditions are needed to study the existence of solutions and asymptotic properties of MPELR estimators. In the following discussion, we always assume that $n \rightarrow \infty$, $n/n_h \rightarrow k_h > 0$ and

$$\max_{1 \leq i \leq n_h} d_{hi}(s) = O\left(\frac{1}{n}\right), \quad \frac{\max_i d_{hi}(s)}{\min_i d_{hi}(s)} = O(1) \quad (3.1)$$

as $\nu \rightarrow \infty$. The assumption (3.1) means that no survey weight is disproportionately large. We will limit ourself to stratified SRS or unequal probability sampling with replacement. It is of interest to see what (3.1) reduces to in some simple special cases. If the sampling design is stratified SRS sampling, then $d_{hi}(s) = W_h/n_h = N_h/(Nn_h)$, and (3.1) is the same as

$$\frac{N_h}{Nn_h} = O\left(\frac{1}{n}\right) \text{ for each } h,$$

which will be satisfied since $n/n_h \rightarrow k_h > 0$. If the sampling design is a stratified unequal probability sampling design, we suppose that the sample is taken with replacement with selection probability a_{hi} , where a_{hi} are known measure of the i -th unit in stratum h , and the characteristic of interest is positive correlated with a_{hi} . In this case, $d_{hi}(s) = W_h/(N_h n_h a_{hi})$, and (3.1) reduces to

$$\max_i d_{hi}(s) = \frac{W_h}{n_h} \max_i \frac{1}{N_h a_{hi}} = O\left(\frac{1}{n}\right) \text{ for each } h$$

$$\frac{\max_i a_{hi}}{\min_i a_{hi}} = O(1),$$

i.e.,

$$\max_i \frac{1}{N_h a_{hi}} = O(1), \quad \frac{\max_i a_{hi}}{\min_i a_{hi}} = O(1).$$

Actually,

$$\frac{Max_i a_{hi}}{Min_i a_{hi}} = O(1) \text{ imply } Max_i \frac{1}{N_h a_{hi}} = O(1).$$

Hence in this case (3.1) can be reduced to

$$\frac{Max_i a_{hi}}{Min_i a_{hi}} = O(1). \quad (3.2)$$

We will also suppose some other conditions about survey weights. In section 4.2 we required

$$E\left(\sum_{i \in s_h} d_{hi}(s) \log p_{hi}\right) = N^{-1} \sum_{i=1}^{N_h} \log p_{hi}$$

for all h and p_{hi} , we will have

$$E\left(\sum_{i \in s_h} d_{hi}(s)\right) = W_h.$$

Under assumption (3.1), we have

$$Var\left(\sum_{i \in s_h} d_{hi}(s)\right) = O\left(\frac{1}{n}\right),$$

hence, noting that $E(\sum_{i \in s_h} d_{hi}(s)) = W_h$,

$$\sum_{i \in s_h} d_{hi}(s) \xrightarrow{p} W_h,$$

or

$$m_h = \frac{W_h}{\sum_{i \in s_h} d_{hi}(s)} \xrightarrow{p} 1.$$

Lemma 1 shows that the solutions to (2.7)-(2.9) exist for given θ .

Lemma 1 Suppose that $g(x, \theta)$ is continuous and differentiable in the neighbourhood of θ_0 and $\|g(x, \theta)\|^3$ is bounded by some integrable function $G(x)$ in this neighbourhood. Furthermore, assume that $Var(x_h) = \sigma_h(\theta) > \sigma_0 > 0$ for all h and ν in this neighbourhood. Then in the sphere $\{\theta : \|\theta - \theta_0\| \leq d_n\}$, the equation

$Q_{1n}(\theta, t) = 0$ almost surely has roots $t = t(\theta) = O(d_n)$, and $t(\theta)$ is continuous and differentiable when θ belongs to this sphere, where $d_n = n^{-\frac{1}{3}-\delta}$, $\frac{1}{6} > \delta > 0$.

From Lemma 1 we know under some conditions for any θ in the neighbourhood of θ_0 , we will have $t = t(\theta)$ which is the solution to (2.8)-(2.10), and this solution also guarantees $p_{hi} > 0$. We also know $t(\theta) = o(n^{-\frac{1}{3}})$ a.s.

We need some results from Shao (1994) for the following discussion. We summarize them as a Lemma.

Lemma 2 *Under conditions of Lemma 3.1, for any given θ satisfying $\|\theta - \theta_0\| \leq d_n$, we have*

$$\sigma_\nu^{-\frac{1}{2}}(\theta) \bar{g}_{st,\omega}(\theta) \rightarrow_L N(0, I),$$

where $\bar{g}_{st,\omega}(\theta) = \sum_h W_h \sum_i \omega_{hi}(s) g_{hi}(\theta)$, $\sigma_\nu(\theta) = \text{Var}(\bar{g}_{st,\omega}(\theta))$.

Some special examples of Lemma 2 were discussed by a few authors. P.K. Sen discussed multi-stage probability proportional to size(PPS) sampling and its asymptotic normality(Chapter 12, Handbook of statistics 6, edited by Krishnaiah and Rao). The asymptotic results for two-stage and three-stage cluster sampling were given by Rao and Scott(1979, 1981).

Lemma 3 *Under conditions of Lemma 3.1, for any given θ satisfying $\|\theta - \theta_0\| \leq d_n$, we have*

$$t(\theta) = \hat{S}_n^{-1}(\theta) \bar{g}_{st,\omega}(\theta) + o(n^{-\frac{1}{3}}) = O(n^{-\frac{1}{3}}) \text{ (a.s.)}, \quad (3.3)$$

where

$$\hat{S}_n(\theta) = \sum_h W_h m_h \sum_{i \in s_h} \omega_{hi}(s) (g_{hi}(\theta) - \hat{g}_h(\theta)) (g_{hi}(\theta) - \hat{g}_h(\theta))^T, \quad (3.4)$$

$$\hat{g}_h(\theta) = \sum_i \hat{p}_{hi} g_{hi}(\theta), \quad \hat{p}_{hi} = \omega_{hi}(s) / [1 + m_h t^T(\theta) (g_{hi}(\theta) - \hat{g}_h(\theta))],$$

$t(\theta)$ is solution to (2.8)-(2.10).

Lemma 4 *In addition to the conditions of Lemma 1, we further assume that $\partial g(x, \theta) / \partial \theta$ is continuous in a neighborhood of the true value θ_0 ; that $\| \partial g(x, \theta) / \partial \theta \|$ is bounded by some integrable function $G(x)$ in this neighbourhood; and that the rank of $\sum_h W_h E_h[\partial g(x_h, \theta_0) / \partial \theta]$ is p . We also assume $\hat{S}_n(\theta) \rightarrow_p S(\theta)$. Then, as $\nu \rightarrow \infty$, with probability 1, $l_E(\theta)$ attains its minimum value at some point $\tilde{\theta}$ in the interior of the ball $\| \theta - \theta_0 \| \leq n^{-\frac{1}{3}}$, and $\tilde{\theta}$ and $\tilde{t} = \tilde{t}(\tilde{\theta})$ satisfy*

$$Q_{1n}(\tilde{\theta}, \tilde{t}) = 0, \quad Q_{2n}(\tilde{\theta}, \tilde{t}) = 0, \quad (3.5)$$

where

$$Q_{1n}(\theta, t) = \sum_h W_h \sum_i \omega_{hi}(s) \frac{g_{hi}(\theta)}{1 + m_h t^T [g_{hi}(\theta) - \hat{g}_h(\theta)]}, \quad (3.6)$$

$$\hat{g}_h(\theta) = \sum_i \frac{\omega_{hi}(s) g_{hi}(\theta)}{[1 + m_h t^T (g_{hi}(\theta) - \hat{g}_h(\theta))]}, \quad (3.7)$$

$$Q_{2n}(\theta, t) = \sum_h W_h \sum_i \frac{\omega_{hi}(s)}{1 + m_h t^T [g_{hi}(\theta) - \hat{g}_h(\theta)]} \frac{\partial g_{hi}(\theta)}{\partial \theta} t. \quad (3.8)$$

Theorem 1 *In addition to the conditions of Lemma 4 above, we further assume that $\frac{\partial^2 g(x, \theta)}{\partial \theta \partial \theta^T}$ is continuous in θ in a neighbourhood of the true value θ_0 , $\| \frac{\partial^2 g(x, \theta)}{\partial \theta \partial \theta^T} \|$ is bounded by some integrable function $G(x)$ in the neighbourhood. Then*

$$V_\nu^{-\frac{1}{2}}(\tilde{\theta})(\tilde{\theta} - \theta_0) \rightarrow N(0, I),$$

and

$$V_\nu^{-\frac{1}{2}}(\tilde{F})(\tilde{F}_n(x) - F(x)) \rightarrow N(0, 1),$$

where $\tilde{F}_n(x) = \sum_h W_h \sum_i \tilde{p}_{hi} 1(x_{hi} < x)$, $\tilde{p}_{hi} = \omega_{hi}(s) / [1 + m_h \tilde{t}^r(g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))]$,

$$V_\nu(\tilde{\theta}) = M_{22.1}^{-1} M_{21} M_{11}^{-1} \sigma_\nu(\theta_0) M_{11}^{-1} M_{12} M_{22.1}^{-1},$$

$$M_{11} = S(\theta_0), \quad M_{12} = -\sum_h W_h E_h \frac{\partial g_h(\theta_0)}{\partial \theta^r},$$

$$M_{21} = M_{12}^r, \quad M_{22.1} = -M_{21} M_{11}^{-1} M_{12},$$

$$V_\nu(\tilde{F}) = Var \left(\sum_h W_h \sum_i \omega_{hi}(s) [I(x_{hi} < x) - BU(g_{hi}(\theta_0) - M_h(\theta_0))] \right),$$

$$B = \sum_h W_h E_h [g(x_h, \theta_0) - E_h g^r(x_h, \theta_0) I(x_h < x)], \quad U = M_{11}^{-1} (I + M_{12} M_{22.1}^{-1} M_{11}^{-1}).$$

$$M_h(\theta_0) = E g(x_h, \theta_0).$$

Theorem 2 Suppose $n\sigma_\nu(\theta_0) = nVar(\bar{g}_{st,\omega}(\theta_0)) \rightarrow V$ as $\nu \rightarrow \infty$ and the conditions of Theorem 2.1, then in the semi-parametric model (2.6), for testing $H_0 : \theta = \theta_0$ the empirical likelihood ratio test statistic

$$\begin{aligned} R = & 2n \left\{ \sum_h \sum_i d_{hi}(s) \log[1 + m_h t^r(\theta_0)(g_{hi}(\theta_0) - \hat{g}_h(\theta_0))] \right. \\ & \left. - \sum_h \sum_i d_{hi}(s) \log[1 + m_h \tilde{t}^r(\tilde{\theta})(g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))] \right\} \end{aligned}$$

is asymptotically distributed as $\sum_{i=1}^p \lambda_i Z_i^2$ under H_0 , where $Z_1, \dots, Z_p$ are independent $N(0, 1)$ random variables, and λ_i 's are eigenvalues of

$$-V^{\frac{1}{2}} M_{11}^{-1} M_{12} M_{22.1}^{-1} M_{21} M_{11}^{-1} V^{\frac{1}{2}}.$$

Since the asymptotic distribution of R involves unknown parameters λ_i 's, we must have their estimators before any practical testing. Suppose $\hat{V}$ is a consistent estimator of V as discussed previously, then from $-\hat{V}^{\frac{1}{2}} M_{n,11}^{-\frac{1}{2}} M_{n,12} \hat{M}_{22,1}^{-1} M_{n,21} M_{n,11}^{-\frac{1}{2}} \hat{V}^{\frac{1}{2}}$ we can have consistent estimators for λ_i 's, say $\hat{\lambda}_i$, then we can use either exact methods or simulation to calculate the required percentile points and hence do the testing of H_0 . Another way is to use a modified R , denoted by R_c :

$$R_c = R/\hat{\delta},$$

where

$$\hat{\delta} = \text{tr}\{-V^{\frac{1}{2}} M_{11}^{-\frac{1}{2}} M_{12} M_{22,1}^{-1} M_{21} M_{11}^{-\frac{1}{2}} V^{\frac{1}{2}}\}/p\}.$$

We treat R_c approximately distributed as χ_p^2 under H_0 . Rao and Scott(1994) considered such modifications in the content of analysis of categorical survey data.

Corollary 1 *Assume $r > p$. In order to test the model (2.6), we can use the empirical likelihood ratio statistic*

$$R_1 = 2n \sum_h \sum_i d_{hi}(s) \log[1 + m_h \tilde{t}^{\tau}(\tilde{\theta}) g_{hi}(\tilde{\theta})],$$

under the conditions of the Theorem 2, R_1 is asymptotically distributed as

$$\sum_{i=1}^{r-p} \lambda_i Z_i^2$$

if (2.6) is correct, where $Z_1, \dots, Z_{r-p}$ are independent $N(0, 1)$ random variables, and λ_i 's are eigenvalues of $V^{\frac{1}{2}} U V^{\frac{1}{2}}$.

We can use the same method to modify R_1 as for R in the above theorem.

4.4 An example

We consider an illustration of the estimation procedures. We primarily consider large sample aspects, but computational issues are discussed by giving the equations and method to solve them.

Suppose that our sampling scheme is stratified sampling, sample in stratum h is taken according to some method, such as unequal probability sampling with replacement. Samples are independent across strata. We denote $x_{h1}, \dots, x_{hn_h}$ as the sample from stratum h , $h = 1, \dots, H$. We also suppose that $x_{hi} = (z_{hi}, y_{hi})$, where z is auxiliary variable and y is the characteristic variable of interest, that the population mean of z_{hi} 's, $\bar{Z}$, is known, we want to estimate $\theta = \bar{Y}$, the population mean of y_{hi} 's. If we denote

$$g(x, \theta) = (z - \bar{Z}, y - \theta)^T,$$

then the estimating equation (2.6) can be written as

$$\sum_h W_h E_h(z - \bar{Z}) = 0, \quad (4.1)$$

$$\sum_h W_h E_h(y - \theta) = 0. \quad (4.2)$$

Since

$$\frac{\partial g_{hi}(\theta)}{\partial \theta} = (0, -1)^T, \quad t = (t_1, t_2)^T,$$

from (3.8) we have $t_2 = 0$. Then (3.6) becomes

$$Q_{1n}(\theta, t_1) = \sum_h W_h \sum_i \omega_{hi}(s) \frac{z_{hi} - \bar{Z}}{1 + m_h t_1 (z_{hi} - \hat{Z}_h)} = 0, \quad (4.3)$$

$$\theta = \sum_h W_h \sum_i \omega_{hi}(s) \frac{y_{hi}}{1 + m_h t_1 (y_{hi} - \bar{Z} - \hat{Z}_h)}, \quad (4.4)$$

where

$$\hat{\bar{Z}}_h = \sum_h W_h \sum_i \hat{p}_{hi} z_{hi}, \quad \hat{p}_{hi} = \frac{\omega_{hi}(s)}{1 + m_h t_1(y_{hi} - \bar{Z} - \hat{\bar{Z}}_h)}.$$

We have $\hat{\theta} = \sum_h W_h \sum_i \hat{p}_{hi} y_{hi}$. As discussed in chapter 2, solving (4.3) is equivalent to solving the following equations:

$$\sum_i \omega_{hi}(s) \frac{z_{hi} - \bar{Z} - u_h}{1 + m_h \psi(z_{hi} - \bar{Z} - u_h)} = 0, \quad (4.5)$$

$$\sum_h W_h u_h = 0. \quad (4.6)$$

For appropriately selected initial values, e.g.,

$$\psi = 0, \quad u_h = \bar{z}_{h,\omega} = \sum_i \omega_{hi}(s) z_{hi},$$

we can solve (4.5) and (4.6) by using NAG Fortran library function to get a solution $\hat{\psi}, \hat{u}_1, \dots, \hat{u}_H$, and

$$\hat{p}_{hi} = \omega_{hi}(s) [1 + m_h \hat{\psi}(z_{hi} - \bar{Z} - \hat{u}_h)]^{-1}.$$

Hence we can have $\hat{\theta}$ and $\hat{F}_Y(y)$.

Since $M_{12} = (0, 1)^T$, $M_{21} = (0, 1)$, $M_{22.1} = -(0, 1)M_{11}^{-1}(0, 1)^T$, and replacing M_{11} by its estimator

$$\hat{S}_n = \sum_h W_h m_h \sum_i \omega_{hi}(s) (z_{hi} - \bar{z}_{h,\omega}, y_{hi} - \bar{y}_{h,\omega}) (z_{hi} - \bar{z}_{h,\omega}, y_{hi} - \bar{y}_{h,\omega})^T,$$

where $\bar{y}_{h,\omega} = \sum_i \omega_{hi}(s) y_{hi}$, $\bar{z}_{h,\omega} = \sum_i \omega_{hi}(s) z_{hi}$, we have, noting (5.17),

$$\hat{\theta} = \bar{z}_{st,\omega} + \frac{\sum_h W_h m_h \sum_i \omega_{hi}(s) (z_{hi} - \bar{z}_{h,\omega})(y_{hi} - \bar{y}_{h,\omega})}{\sum_h W_h m_h \sum_i \omega_{hi}(s) (z_{hi} - \bar{z}_{h,\omega})^2} (\bar{Z} - \bar{z}_{st,\omega}) + o_p(n^{-\frac{1}{2}}), \quad (4.7)$$

where $\bar{z}_{st,\omega} = \sum_h W_h \bar{z}_{h,\omega}$, $\bar{y}_{st,\omega} = \sum_h W_h \bar{y}_{h,\omega}$.

For stratified simple random sampling, this approximate estimator will be

$$\hat{\theta}' = \bar{y}_{st} + \frac{\sum_h W_h n_h^{-1} \sum_i (z_{hi} - \bar{z}_h)(y_{hi} - \bar{y}_h)}{\sum_h W_h n_h^{-1} \sum_i (z_{hi} - \bar{z}_h)^2} (\bar{Z} - \bar{z}_{st}),$$

which is identical to the estimator obtained by Chen and Sitter(1996). Note that, $\hat{\theta}'$ is not equal to the approximate EL estimator in Chapter 2, The latter estimator is asymptotically equal to the optimal regression estimator.

We will use the above example to show that the asymptotic distribution of pseudo empirical likelihood ratio test statistic is not generally a χ^2 -distribution here. Suppose the sampling scheme is stratified SRS, and $W_h/n_h = c/n$ for all h , that is, proportional allocation, then

$$n\sigma_\nu(\theta_0) = n\text{Var}(\bar{G}_{st}(\theta_0)) = c \sum_h W_h \text{Var}(x_h) = V = cM_{11},$$

$$\hat{S}_n \rightarrow_p \sum_h W_h \text{Var}(x_h) = M_{11}.$$

Denote

$$A = -V^{\frac{1}{2}} M_{11}^{-1} M_{12} M_{22,1}^{-1} M_{21} M_{11}^{-1} V^{\frac{1}{2}},$$

then

$$A^2 = cA,$$

hence

$$R/c \rightarrow_d \chi_1^2.$$

4.5 Proofs

Proof of Lemma 1.

For fixed θ such that $\|\theta - \theta_0\| < d_n$, consider the function

$$\psi(\lambda) = \sum_h W_h \sum_{i \in s_h} \omega_{hi}(s) \frac{g_{hi}(\theta) - \hat{g}_h(\theta)}{1 + n^{-\frac{1}{3}-\frac{\delta}{2}} m_h \lambda^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)]}.$$

From $E \|g(x, \theta)\|^3 < \infty$ we can get $\max_h \max_i \|g(x_{hi}, \theta)\| = o(n^{\frac{1}{3}})$ a.s. for both stratified SRS or PPS with replacement case, and $\psi(\lambda)$ is an almost surely continuous function for $\|\lambda\| \leq 1$. When $\|\lambda\| = 1$, we have

$$\begin{aligned} \lambda^\tau \psi(\lambda) &= \lambda^\tau \sum_h W_h \sum_i \omega_{hi}(s) (g_{hi}(\theta) - \hat{g}_h(\theta)) \\ &\quad - n^{-\frac{1}{3}-\frac{\delta}{2}} \lambda^\tau \sum_h W_h m_h \sum_i \omega_{hi}(s) [g_{hi}(\theta) - \hat{g}_h(\theta)] [g_{hi}(\theta) - \hat{g}_h(\theta)]^\tau \lambda \\ &\quad + O(\lambda^\tau n^{-\frac{2}{3}-\delta} \sum_h W_h m_h^2 \sum_i \omega_{hi}(s) [g_{hi}(\theta) - \hat{g}_h(\theta)] \\ &\quad \cdot \lambda^\tau [g_{hi}(\theta) - \hat{g}_h(\theta)] [g_{hi}(\theta) - \hat{g}_h(\theta)]^\tau \lambda) \text{ a.s.} \\ &= \lambda^\tau \sum_h W_h \sum_i \omega_{hi}(s) g_{hi}(\theta) - n^{-\frac{1}{3}-\frac{\delta}{2}} \lambda^\tau \sum_h W_h m_h \sum_i \omega_{hi}(s) g_{hi}(\theta) g_{hi}^\tau(\theta) \lambda \\ &\quad + n^{-\frac{1}{3}-\frac{\delta}{2}} \lambda^\tau \sum_h W_h m_h \hat{g}_h(\theta) \hat{g}_h^\tau(\theta) \lambda + O(n^{-\frac{1}{3}-\delta}) \text{ a.s.} \\ &\leq O(n^{-\frac{1}{3}-\delta} - \frac{c}{2} n^{-\frac{1}{3}-\frac{\delta}{2}}) + O(n^{-\frac{1}{3}-\delta}) \text{ a.s.} \\ &< 0, \text{ a.s.} \end{aligned}$$

where c is the smallest eigenvalue of σ_0 . By the Lemma above there exists a point $\hat{\lambda}$ such that $\|\hat{\lambda}\| < 1$ and $\psi(\hat{\lambda}) = 0$. Using the implicit-function theorem, we can easily get other Lemma's results. $\square$

Proof of Lemma 3

By using the decomposition

$$\begin{aligned} g_{hi}(\theta) &= g_{hi}(\theta) [1 + m_h t^\tau(\theta) (g_{hi}(\theta) - \hat{g}_h(\theta))] \\ &\quad - (g_{hi}(\theta) - \hat{g}_h(\theta) + \hat{g}_h(\theta)) m_h t^\tau(\theta) (g_{hi}(\theta) - \hat{g}_h(\theta)) \end{aligned}$$

and noting (2.9) we have

$$\begin{aligned}
 \hat{g}_h(\theta) &= \sum_i \hat{p}_{hi} g_{hi}(\theta) \\
 &= \sum_i \frac{\omega_{hi}(s)}{[1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))]} g_{hi}(\theta) \\
 &= \bar{g}_{h,\omega}(\theta) - m_h \sum_i \omega_{hi}(s) \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta), \quad (5.1)
 \end{aligned}$$

where $\bar{g}_{h,\omega}(\theta) = \sum_i \omega_{hi}(s) g_{hi}(\theta)$. Since $\sum_h W_h \hat{g}_h(\theta) = 0$, we have

$$\begin{aligned}
 0 &= \sum_h W_h \bar{g}_{h,\omega}(\theta) - \sum_h W_h m_h \sum_i \omega_{hi}(s) \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta) \\
 &= \bar{g}_{st,\omega}(\theta) - \sum_h W_h m_h \sum_i \omega_{hi}(s) \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta), \quad (5.2)
 \end{aligned}$$

or

$$\bar{g}_{st,\omega}(\theta) = \sum_h W_h m_h \sum_i \omega_{hi}(s) \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} t(\theta). \quad (5.3)$$

From (5.1) and (5.3) we get

$$\begin{aligned}
 W_h^2 \| \bar{g}_{h,\omega}(\theta) - \hat{g}_h(\theta) \|^2 &= t^\tau(\theta) \left[W_h m_h \sum_i \omega_{hi}(s) \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \right]^2 t(\theta) \\
 &\leq t^\tau(\theta) \left[\sum_h W_h m_h \sum_i \omega_{hi}(s) \frac{(g_{hi}(\theta) - \hat{g}_h(\theta))(g_{hi}(\theta) - \hat{g}_h(\theta))^\tau}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \right]^2 t(\theta) \\
 &= \| \bar{g}_{st,\omega}(\theta) \|^2 \rightarrow 0. \quad (a.s.) \quad (5.4)
 \end{aligned}$$

Hence

$$\begin{aligned}
 \hat{S}_n(\theta) &= \sum_h W_h m_h \sum_i \omega_{hi}(s) (g_{hi}(\theta) - \bar{g}_{h,\omega}(\theta) + \bar{g}_{h,\omega}(\theta) - \hat{g}_h(\theta)) \\
 &\quad \cdot (g_{hi}(\theta) - \bar{g}_{h,\omega}(\theta) + \bar{g}_{h,\omega}(\theta) - \hat{g}_h(\theta))^\tau \\
 &= \sum_h W_h m_h \sum_i \omega_{hi}(s) (g_{hi}(\theta) - \bar{g}_{h,\omega}(\theta))(g_{hi}(\theta) - \bar{g}_{h,\omega}(\theta))^\tau \\
 &\quad + \sum_h W_h^2 (\bar{g}_{h,\omega}(\theta) - \hat{g}_h(\theta))(\bar{g}_{h,\omega}(\theta) - \hat{g}_h(\theta))^\tau \quad (a.s.). \quad (5.5)
 \end{aligned}$$

Therefore, we know

$$\hat{S}_n(\theta) = \sum_h W_h m_h \sum_i \omega_{hi}(s) (g_{hi}(\theta) - \hat{g}_h(\theta)) (g_{hi}(\theta) - \hat{g}_h(\theta))^T = O(1) \text{ (a.s.)}. \quad (5.6)$$

From $Q_{1n}(\theta, t(\theta)) = 0$, we have

$$\begin{aligned} 0 &= \sum_h W_h \sum_i \omega_{hi}(s) \frac{g_{hi}(\theta) - \hat{g}_h(\theta)}{1 + m_h t^T(\theta) [g_{hi}(\theta) - \hat{g}_h(\theta)]} \\ &= \sum_h W_h \sum_i \omega_{hi}(s) [g_{hi}(\theta) - \hat{g}_h(\theta)] \\ &\quad - \sum_h W_h m_h \sum_i \omega_{hi}(s) [g_{hi}(\theta) - \hat{g}_h(\theta)] [g_{hi}(\theta) - \hat{g}_h(\theta)]^T t(\theta) + o(n^{-\frac{1}{3}}) \\ &= \sum_h W_h \sum_i \omega_{hi}(s) g_{hi}(\theta) - \hat{S}_n(\theta) t(\theta) + o(n^{-\frac{1}{3}}). \end{aligned}$$

Hence (noting $\hat{S}_n > \sigma_1/2 > 0$ a.s.)

$$t(\theta) = \hat{S}_n^{-1}(\theta) \bar{g}_{st,\omega}(\theta) + o(n^{-\frac{1}{3}}). \square$$

Proof of Lemma 4

Denote $\theta = \theta_0 + un^{-\frac{1}{3}}$ for $\theta \in \{\theta \mid \|\theta - \theta_0\| = n^{-\frac{1}{3}}\}$, where $\|u\| = 1$. First we give a lower bound for $l_E(\theta)$ on the surface of the ball. From Lemma 1, Lemma 3 and Lemma 3 of chapter 3 we know when $\|\theta - \theta_0\| \leq n^{-\frac{1}{3}}$, we have

$$t(\theta) = \hat{S}_n^{-1}(\theta) \bar{g}_{st,\omega}(\theta) + o(n^{-\frac{1}{3}}) \text{ (a.s.)} = O(n^{-\frac{1}{3}}) \text{ (a.s.)}$$

uniformly about $\theta \in \{\theta \mid \|\theta - \theta_0\| \leq n^{-\frac{1}{3}}\}$.

By this and Taylor's expansion, we have (uniformly for u)

$$\begin{aligned} l_E(\theta) &= \sum_h \sum_i m_h t^T(\theta) d_{hi}(s) (g_{hi}(\theta) - \hat{g}_h(\theta)) \\ &\quad - \frac{1}{2} \sum_h \sum_i d_{hi}(s) [m_h t^T(\theta) (g_{hi}(\theta) - \hat{g}_h(\theta))]^2 + o(n^{-\frac{2}{3}}) \text{ (a.s.)} \end{aligned}$$

$$\begin{aligned}
&= t^r(\theta) \bar{g}_{st,\omega}(\theta) - \frac{1}{2} t^r(\theta) \hat{S}_n(\theta) t(\theta) + o(n^{-\frac{2}{3}}) \text{ (a.s.)} \\
&= t^r(\theta) [\bar{g}_{st,\omega}(\theta) - \frac{1}{2} \hat{S}_n(\theta) t(\theta)] + o(n^{-\frac{2}{3}}) \text{ (a.s.)} \\
&= \frac{1}{2} t^r(\theta) \bar{g}_{st,\omega}(\theta) + o(n^{-\frac{2}{3}}) \text{ (a.s.)} \\
&= \frac{1}{2} \bar{g}_{st,\omega}^r(\theta) \hat{S}_n^{-1}(\theta) \bar{g}_{st,\omega}(\theta) + o(n^{-\frac{2}{3}}) \text{ (a.s.)} \\
&= \frac{1}{2} \left[\bar{g}_{st,\omega}(\theta_0) + \sum_h W_h \sum_i d_{hi}(s) \frac{\partial g_{hi}(\theta_0)}{\partial \theta} u n^{-\frac{1}{3}} \right]^r \hat{S}_n^{-1}(\theta) \\
&\quad \cdot \left[\bar{g}_{st,\omega}(\theta_0) + \sum_h W_h \sum_i d_{hi}(s) \frac{\partial g_{hi}(\theta_0)}{\partial \theta} u n^{-\frac{1}{3}} \right] + o(n^{-\frac{2}{3}}) \text{ (a.s.)} \\
&= \frac{1}{2} \left[\sum_h W_h O(n_h^{-\frac{1}{2}} (\log \log n_h)^{\frac{1}{2}}) + \sum_h W_h E_h \frac{\partial g_{hi}(\theta_0)}{\partial \theta} u n^{-\frac{1}{3}} \right]^r S^{-1}(\theta_0) \\
&\quad \cdot \left[\sum_h W_h O(n_h^{-\frac{1}{2}} (\log \log n_h)^{\frac{1}{2}}) + \sum_h W_h E_h \frac{\partial g_{hi}(\theta_0)}{\partial \theta} u n^{-\frac{1}{3}} \right] + o(n^{-\frac{2}{3}}) \text{ a.s.} \\
&\geq (c - \varepsilon) n^{-\frac{2}{3}} \text{ (a.s.)}
\end{aligned}$$

where $c - \varepsilon > 0$ and c is the smallest eigenvalue of

$$\frac{1}{2} \left[\sum_h W_h E_h \frac{\partial g_h(\theta_0)}{\partial \theta} \right]^r S^{-1}(\theta_0) \left[\sum_h W_h E_h \frac{\partial g_h(\theta_0)}{\partial \theta} \right].$$

Similarly

$$l_E(\theta_0) = \frac{1}{2} \bar{g}_{st,\omega}^r(\theta_0) \hat{S}_n^{-1}(\theta_0) \bar{g}_{st,\omega}(\theta_0) + o(n^{-\frac{2}{3}}) \text{ a.s.} = o(n^{-\frac{2}{3}}) \text{ a.s.}$$

Since $l_E(\theta)$ is a continuous function about θ as θ belongs to the ball $\|\theta - \theta_0\| \leq n^{-\frac{1}{3}}$. $l_E(\theta)$ has minimum value in the interior of this ball, and $\tilde{\theta}$ satisfies (noting (2.8) and (2.10))

$$\begin{aligned}
0 &= \frac{\partial l_E(\theta)}{\partial \theta^r} \Big|_{\theta=\tilde{\theta}} \\
&= \sum_h \sum_i \frac{m_h d_{hi}(s)}{1 + m_h t^r(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \left[\frac{\partial t(\theta)}{\partial \theta} (g_{hi}(\theta) - \hat{g}_h(\theta)) \right]
\end{aligned}$$

$$\begin{aligned}
& + \left(\frac{\partial g_{hi}(\theta)}{\partial \theta} - \frac{\partial \hat{g}_h(\theta)}{\partial \theta} \right) t(\theta) \Big|_{\theta=\tilde{\theta}} \\
& = \sum_h \sum_i \frac{m_h d_{hi}(s)}{1 + m_h t^\tau(\theta)(g_{hi}(\theta) - \hat{g}_h(\theta))} \frac{\partial g_{hi}(\theta)}{\partial \theta} t(\theta) \Big|_{\theta=\tilde{\theta}}
\end{aligned}$$

i.e., $Q_{2n}(\tilde{\theta}, \tilde{t}) = 0$. $\square$

Proof of Theorem 1

Taking derivatives about θ^τ , t^τ , we have

$$\frac{\partial Q_{1n}(\theta_0, 0)}{\partial \theta^\tau} = \sum_h W_h \sum_i \omega_{hi}(s) \frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau} = \frac{\partial \bar{g}_{st,\omega}(\theta_0)}{\partial \theta^\tau}. \quad (5.7)$$

$$\frac{\partial Q_{1n}(\theta_0, 0)}{\partial t^\tau} = - \sum_h W_h m_h \sum_i \omega_{hi}(s) (g_{hi}(\theta_0) - \hat{g}_h(\theta_0)) (g_{hi}(\theta_0) - \hat{g}_h(\theta_0))^\tau = -S_n(\theta_0), \quad (5.8)$$

$$\frac{\partial Q_{2n}(\theta_0, 0)}{\partial \theta^\tau} = 0, \quad \frac{\partial Q_{2n}(\theta_0, 0)}{\partial t^\tau} = \sum_h W_h \sum_i \omega_{hi}(s) \frac{\partial g_{hi}(\theta_0)}{\partial \theta} = \frac{\partial \bar{g}_{st,\omega}(\theta_0)}{\partial \theta}. \quad (5.9)$$

Expanding $Q_{1n}(\tilde{\theta}, \tilde{t})$, $Q_{2n}(\tilde{\theta}, \tilde{t})$ at $(\theta, 0)$, we have

$$\begin{aligned}
0 & = Q_{in}(\tilde{\theta}, \tilde{t}) \\
& = Q_{in}(\theta_0, 0) + \frac{\partial Q_{in}(\theta_0, 0)}{\partial \theta^\tau} (\tilde{\theta} - \theta_0) + \frac{\partial Q_{in}(\theta_0, 0)}{\partial t^\tau} \tilde{t} + o_p(\delta_n), i = 1, 2 \quad (5.10)
\end{aligned}$$

where $\delta_n = \|\tilde{\theta} - \theta_0\| + \|\tilde{t}\|$. Hence we have

$$\begin{aligned}
0 & = Q_{1n}(\tilde{\theta}, \tilde{t}) \\
& = \bar{g}_{st,\omega}(\theta_0) + \frac{\partial \bar{g}_{st,\omega}(\theta_0)}{\partial \theta^\tau} (\tilde{\theta} - \theta_0) - S_n(\theta_0) (\tilde{t} - 0) + o_p(\delta_n) \quad (5.11)
\end{aligned}$$

$$0 = Q_{2n}(\tilde{\theta}, \tilde{t}) = \frac{\partial \bar{g}_{st,\omega}(\theta_0)}{\partial \theta} (\tilde{t} - 0) + o_p(\delta_n). \quad (5.12)$$

The system equations of (5.10) (5.11) written in matrix form is

$$\begin{pmatrix} \bar{g}_{st,\omega}(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix} = \begin{pmatrix} S_n(\theta_0) & M_{n,12} \\ M_{n,21} & 0 \end{pmatrix} \begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix}. \quad (5.13)$$

or

$$\begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} = M_n^{-1} \begin{pmatrix} \bar{g}_{st,\omega}(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix}, \quad (5.14)$$

where M_n is the matrix of (5.13), $M_{n,12} = -\frac{\partial \bar{g}_{st,\omega}(\theta_0)}{\partial \theta^\tau}$, $M_{n,21} = M_{n,12}^\tau$. Since

$$S_n(\theta_0) \xrightarrow{p} M_{11}, \quad M_{n,12} \xrightarrow{p} -\sum_h W_h E_h \frac{\partial g_h(\theta_0)}{\partial \theta^\tau} = M_{12},$$

hence

$$M_n \xrightarrow{p} M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}.$$

where $M_{21} = M_{12}^\tau$, $M_{22} = 0$. Therefore

$$\begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} = M^{-1} \begin{pmatrix} \bar{g}_{st,\omega}(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix}. \quad (5.15)$$

From this and $\bar{g}_{st,\omega}(\theta_0) = \sum_h W_h \sum_i \omega_{hi}(s) g_{hi}(\theta_0) = O_p(n^{-\frac{1}{2}})$, we know that $\delta_n = O_p(n^{-\frac{1}{2}})$. Since

$$M^{-1} = \begin{pmatrix} I & -M_{11}^{-1}M_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & M_{22,1}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -M_{21}M_{11}^{-1} & I \end{pmatrix},$$

we have

$$\begin{aligned} & \begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} \\ &= M^{-1} \begin{pmatrix} \bar{g}_{st,\omega}(\theta_0) \\ 0 \end{pmatrix} + o_p(\delta_n) \\ &= \begin{pmatrix} I & -M_{11}^{-1}M_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & M_{22,1}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -M_{21}M_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} \bar{g}_{st,\omega}(\theta_0) \\ 0 \end{pmatrix} \end{aligned} \quad (5.16)$$

$$\begin{aligned}
& + o_p(\delta_n) \\
& = \begin{pmatrix} M_{11}^{-1} & -M_{11}^{-1} M_{12} M_{22.1}^{-1} \\ 0 & M_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} \bar{g}_{st,\omega}(\theta_0) \\ -M_{21} M_{11}^{-1} \bar{g}_{st,\omega}(\theta_0) \end{pmatrix} + o_p(\delta_n) \\
& = \begin{pmatrix} U \bar{g}_{st,\omega}(\theta_0) \\ -M_{22.1}^{-1} M_{21} M_{11}^{-1} \bar{g}_{st,\omega}(\theta_0) \end{pmatrix} + o_p(\delta_n). \tag{5.17}
\end{aligned}$$

Hence

$$V_\nu^{-\frac{1}{2}}(\tilde{\theta})(\tilde{\theta} - \theta_0) \rightarrow N(0, I),$$

where

$$\begin{aligned}
V_\nu(\tilde{\theta}) &= M_{22.1}^{-1} M_{21} M_{11}^{-1} \text{Var}(\bar{g}_{st,\omega}(\theta_0)) M_{11}^{-1} M_{12} M_{22.1}^{-1} \\
&= M_{22.1}^{-1} M_{21} M_{11}^{-1} \sigma_\nu(\theta_0) M_{11}^{-1} M_{12} M_{22.1}^{-1}.
\end{aligned}$$

Now we turn to the second conclusion. Since

$$\begin{aligned}
\tilde{p}_{hi} &= \frac{\omega_{hi}(s)}{[1 + m_h \tilde{t}^\tau (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))]} \\
&= \omega_{hi}(s) \left[1 - m_h \tilde{t}^\tau (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) \right] + o_p(n^{-\frac{3}{2}}),
\end{aligned}$$

noting (5.17) we get

$$\begin{aligned}
& \tilde{F}_n(x) \\
&= \sum_h W_h \sum_i \tilde{p}_{hi} 1(x_{hi} < x) \\
&= \sum_h W_h \sum_i \omega_{hi}(s) 1(x_{hi} < x) \\
&\quad - \tilde{t}^\tau \sum_h W_h m_h \sum_i \omega_{hi}(s) (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) 1(x_{hi} < x) + o_p(n^{-\frac{1}{2}}) \\
&= \sum_h W_h \sum_i \omega_{hi}(s) 1(x_{hi} < x) \\
&\quad - \left[\sum_h W_h m_h \sum_i \omega_{hi}(s) (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))^\tau 1(x_{hi} < x) \right] U \bar{g}_{st,\omega}(\theta_0)
\end{aligned}$$

$$\begin{aligned}
& +o_p(n^{-\frac{1}{2}}) \\
& = \sum_h W_h m_h \sum_i \omega_{hi}(s) 1(x_{hi} < x) - B_n U \bar{g}_{st,\omega}(\theta_0) + o_p(n^{-\frac{1}{2}}) \\
& = \sum_h W_h \sum_i \omega_{hi}(s) 1(x_{hi} < x) - BU \bar{g}_{st,\omega}(\theta_0) + o_p(n^{-\frac{1}{2}}) \\
& = \sum_h W_h \sum_i \omega_{hi}(s) [1(x_{hi} < x) - BU(g_{hi}(\theta_0) - M_h(\theta_0))] + o_p(n^{-\frac{1}{2}}),
\end{aligned}$$

where

$$\begin{aligned}
B_n &= \sum_h W_h m_h \sum_i \omega_{hi}(s) (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))^T 1(x_{hi} < x), \\
B &= \sum_h W_h E_h [(g(x_h, \theta_0) - E_h g(x_h, \theta_0))^T 1(x_h < x)], \\
E_h g(x_h, \theta_0) &= M_h(\theta_0), \quad B_n \xrightarrow{p} B.
\end{aligned}$$

Hence,

$$V_\nu^{-\frac{1}{2}}(\tilde{F})(\tilde{F}_n(x) - F(x)) \rightarrow N(0, 1).$$

We note here since $M_{n,11}$, $M_{n,12}$, $M_{n,21}$ are consistent estimators of M_{11} , M_{12} , M_{21} respectively, if we have a consistent estimator for $\sigma_\nu(\theta_0) = Var(\bar{g}_{st,\omega}(\theta_0))$, say, $\hat{\sigma}(\theta_0)$, then

$$\hat{V}_\nu(\tilde{\theta}) = \hat{M}_{22.1}^{-1} M_{n,21} M_{n,11}^{-1} \hat{\sigma}(\tilde{\theta}) M_{n,11}^{-1} M_{n,12} \hat{M}_{22.1}^{-1}$$

is a consistent estimator of $V_\nu(\tilde{\theta})$, where $\hat{M}_{22.1} = -M_{n,21} M_{n,11}^{-1} M_{n,12}$. There are some methods in literature on how to estimate $\sigma_\nu(\theta_0)$, such as the jackknife estimator. Since B can be estimated by B_n , U by U_n , hence from an estimator of $\sigma_\nu(\theta_0)$, we can also get an estimator of $V_\nu(\tilde{F})$.

Proof of Theorem 2

Noting (5.17) we have

$$nl_E(\tilde{\theta}, \tilde{t})$$

$$\begin{aligned}
&= n \sum_h \sum_i d_{hi}(s) \log[1 + m_h \tilde{t}^\tau (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))] \\
&= n \sum_h m_h \sum_i d_{hi}(s) \tilde{t}^\tau (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) \\
&\quad - \frac{n}{2} \sum_h \sum_i m_h^2 d_{hi}(s) [\tilde{t}^\tau (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))]^2 + o_p(1) \\
&= n \tilde{t}^\tau \sum_h m_h \sum_i d_{hi}(s) (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) \\
&\quad - \frac{n}{2} \tilde{t}^\tau \sum_h m_h^2 \sum_i d_{hi}(s) (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta})) (g_{hi}(\tilde{\theta}) - \hat{g}_h(\tilde{\theta}))^\tau \tilde{t} + o_p(1) \\
&= n \tilde{t}^\tau \bar{g}_{st,\omega}(\tilde{\theta}) - \frac{n}{2} \tilde{t}^\tau \hat{S}_n(\tilde{\theta}) \tilde{t} + o_p(1) \\
&= n \tilde{t}^\tau \bar{g}_{st,\omega}(\theta_0) + n \tilde{t}^\tau \frac{\partial \bar{g}_{st,\omega}(\theta_0)}{\partial \theta^\tau} (\tilde{\theta} - \theta_0) - \frac{n}{2} \tilde{t}^\tau S(\theta_0) \tilde{t} + o_p(1) \\
&= n \bar{g}_{st,\omega}^\tau(\theta_0) U \bar{g}_{st,\omega}(\theta_0) - \frac{n}{2} \bar{g}_{st,\omega}^\tau(\theta_0) U S(\theta_0) U \bar{g}_{st,\omega}(\theta_0) + o_p(1) \\
&= \frac{n}{2} \bar{g}_{st,\omega}^\tau(\theta_0) U \bar{g}_{st,\omega}(\theta_0) + o_p(1).
\end{aligned}$$

Similarly we can get

$$\begin{aligned}
n l_E(\theta_0, t(\theta_0)) &= n \sum_h \sum_i d_{hi}(s) \log[1 + m_h t^\tau(\theta_0) (g_{hi}(\theta_0) - \hat{g}_h(\theta_0))] \\
&= \frac{n}{2} \bar{g}_{st,\omega}^\tau(\theta_0) S^{-1}(\theta_0) \bar{g}_{st,\omega}(\theta_0) + o_p(1).
\end{aligned}$$

Hence

$$\begin{aligned}
R &= n \bar{g}_{st,\omega}^\tau(\theta_0) (S^{-1}(\theta_0) - U) \bar{g}_{st,\omega}(\theta_0) + o_p(1) \\
&= -n \bar{g}_{st,\omega}^\tau(\theta_0) (M_{11}^{-1} M_{12} M_{22.1}^{-1} M_{21} M_{11}^{-1}) \bar{g}_{st,\omega}(\theta_0) + o_p(1).
\end{aligned}$$

Noting that

$$\sqrt{n} V^{-\frac{1}{2}} \bar{g}_{st,\omega}(\theta_0) \rightarrow N(0, I)$$

and

$$-V^{\frac{1}{2}} M_{11}^{-1} M_{12} M_{22.1}^{-1} M_{21} M_{11}^{-1} V^{\frac{1}{2}}$$

is symmetric and nonnegative with rank p , hence we have the theorem's results. $\square$

Proof of Corollary

From the process of proof of theorem 2.2, we have

$$\begin{aligned}
 R_1 &= 2nl_E(\tilde{\theta}, \tilde{t}) \\
 &= 2n \sum_h \sum_i d_{hi}(s) \log[1 + m_h \tilde{t}^T g_{hi}(\tilde{\theta})] \\
 &= n\bar{g}_{st,\omega}^T(\theta_0)U\bar{g}_{st,\omega}(\theta_0) + o_p(1),
 \end{aligned}$$

and the results follows.

Chapter 5

M-estimation in the Presence of Auxiliary Information Using Stratified Random Sampling

5.1 Introduction

Huber (1964) introduced a flexible class of estimators, called ‘M-estimators’, which are generalizations of the usual maximum likelihood estimators and have become a very useful tool in statistical inference. M-estimation plays an important role in robust parameter inference and in non-parametric inference. In this paper, we consider a new class of M-estimators under the following model. Let the population distribution be $F(\cdot)$, and the population be stratified into H strata with known weight W_h of each stratum, $x_{h1}, \dots, x_{hn_h}$ be iid sample from stratum h , $h = 1, \dots, H$. We assume that we have some auxiliary information about the distribution function

F in the sense that there exist $r (\geq 1)$ functions $g_1(x), \dots, g_r(x)$ such that

$$E_F g(x) = \sum_h W_h E_h g(x) = 0, \quad (1.1)$$

where $g(x) = (g_1(x), \dots, g_r(x))^T$, E_h denotes the expectation with respect to the distribution function of stratum h . Here $g(x)$ does not contain any unknown parameters, hence is different from the questions discussed in chapter 3.

Our approach is first to apply empirical likelihood to provide an alternative estimator for the distribution function $F(\cdot)$ by utilizing auxiliary information (1.1). Based on the estimated distribution function which puts unequal weight at each x_{hi} , we propose a new class of M-estimators, and establish some asymptotic results. One advantage of the new M-estimators over those based on the usual empirical distribution function is that they have smaller asymptotic variances, which enables us to improve our statistical inference through M-estimation.

In section 5.2, we describe the profile empirical likelihood under (1.1) and some asymptotic results. In section 5.3, we propose a new class of M-estimators. It is shown that the proposed M-estimators are consistent and asymptotically normally distributed. The proofs are delegated to section 5.4.

5.2 Profile empirical likelihood

Let $x_{h1}, \dots, x_{hn_h}$ be iid sample from stratum h , and samples be independent across strata. The empirical likelihood, as before, given by

$$L = \prod_{h=1}^H \prod_{i=1}^{n_h} p_{hi}, \quad (2.1)$$

where $p_{hi} = Pr(x_h = x_{hi})$. Under the condition that $E_F g(x) = 0$, the profile empirical likelihood is defined to be the maximum of L with respect to p_{hi} subject to

$$\sum_h W_h \sum_i p_{hi} g(x_{hi}) = 0, \quad (2.2)$$

and

$$\sum_i p_{hi} = 1, \quad p_{hi} \geq 0,$$

L can be maximized. The maximum is reached when

$$\hat{p}_{hi} = \frac{1}{n_h} \frac{1}{1 + m_h t^\tau (g(x_{hi}) - \hat{g}_h)},$$

where

$$\hat{g}_h = \sum_{i=1}^{n_h} \hat{p}_{hi} g(x_{hi}), \quad m_h = n W_h n_h^{-1}, \quad n = \sum_{h=1}^H n_h.$$

and t is the solution to

$$\sum_h W_h \hat{g}_h = \sum_h W_h \sum_i \hat{p}_{hi} g(x_{hi}) = 0. \quad (2.3)$$

The numerical method of calculating $\hat{p}_{hi}$ is given in chapter 3.

Now let

$$\hat{F}_n(x) = \sum_h W_h \sum_i \hat{p}_{hi} 1(x_{hi} \leq x),$$

where $1(x_{hi} < x) = (1(x_{hi,1} < x_1), \dots, 1(x_{hi,d} < x_d))^\tau$, $x_{hi,j}$ and x_j are j -th component of vector x_{hi} and x respectively, $1(x_{hi,j} < x_j)$ is the indicator of set $(x_{hi,j} < x_j)$, then $\hat{F}_n(x)$ can be regarded as an alternative estimator of the distribution function of $F(x)$. Note that instead of the usual empirical distribution function giving weight W_h/n_h at each observed point x_{hi} , $\hat{F}_n(x)$ gives the weight $W_h \hat{p}_{hi}$ at each x_{hi} . On the

other hand, if there is no auxiliary information given in (1.1), L attains its maximum at $\hat{p}_{hi} = 1/n_h$, and hence

$$\hat{F}_n(x) = \sum_h W_h \frac{1}{n_h} \sum_i 1(x_{hi} \leq x) = F_n(x),$$

the usual weighted empirical distribution function.

From chapter 3, we have the following results.

Lemma 1 Suppose that as $n \rightarrow \infty$, $n/n_h \rightarrow k_h > 0$ for all h and $E_h[g(x_h) - E_h g(x_h)][g(x_h) - E_h g(x_h)]^r = \sigma_h > \sigma_0$ for all h , where σ_0 is positive definite, E_h denotes the expectation with respect to the distribution of stratum h . Denote $\sigma = \sum_h k_h W_h^2 \sigma_h$, then

$$t = \sigma^{-1} \bar{g}_{st} + o_p(n^{-\frac{1}{2}}). \quad (2.4)$$

Furthermore, if $E \|g(x)\|^4 < \infty$, then

$$t = \sigma^{-1} \bar{g}_{st} + O_p(n^{-1}), \quad (2.5)$$

where $\bar{g}_{st} = \sum_h W_h \frac{1}{n_h} \sum_i g(x_{hi})$. Hence in both cases,

$$\sqrt{nt} \xrightarrow{L} N(0, \sigma^{-1}), \quad \sqrt{n}(\hat{F}_n(x) - F(x)) \xrightarrow{L} N(0, V),$$

where

$$V = \sum_h W_h^2 k_h H_h(x)[1 - F_h(x)] - B^r \sigma^{-1} B,$$

$$B = \sum_h W_h^2 E_h \{(g(x_h) - E_h g(x_h))1(x_h < x)\}.$$

The variance of $\hat{F}_n(x)$ shows clearly that it is generally more efficient than the empirical c.d.f. $F_n(x)$.

5.3 M-estimators

An M-functional $\theta(F)$ associated with distribution function F is defined as a root θ_0 of the equation

$$E_F \psi(x, \theta_0) = \int \psi(x, \theta_0) dF(x) = 0, \quad (3.1)$$

where $x \sim F$ and $x, \theta_0, \psi(x, \theta_0) \in R$. For a stratified sample $x_{11}, \dots, x_{1n_1}; \dots; x_{H1}, \dots, x_{Hn_H}$ from $F(\cdot)$, the usual estimator of θ_0 is given by $\hat{\theta}_n = \theta(F_n)$, where $F_n(\cdot)$ is the natural stratified empirical distribution function, i.e., $F_n(x) = \sum_h \frac{W_h}{n_h} \sum_i 1(x_{hi} < x)$. $\theta(F_n)$ is called the M-estimator corresponding to ψ . We propose a new class of M-estimators by taking the auxiliary information (1.1) into account. Specifically, we propose to estimate θ_0 by $\hat{\theta}_n$, which is defined to be a root, u_0 , of the equation

$$\int \psi(x, u_0) d\hat{F}_n(x) = \sum_h W_h \sum_i \hat{p}_{hi} \psi(x_{hi}, u_0) = 0. \quad (3.2)$$

We will also call $\hat{\theta}_n$ the M-estimator corresponding to ψ . Of course different choices of ψ lead to different estimators. Note that if there is no auxiliary information (1.1), then $\hat{F}_n(\cdot) = F_n(\cdot)$, and hence $\hat{\theta} = \theta(F_n)$, the usual M-estimator. Though we could include $\psi(x, \theta) = 0$ as an estimating equation in the general estimating equations setup of Chapter 3, it may add some difficulty in getting an appropriate solution in practice since $\psi(x, \theta)$ includes the unknown parameter θ . Instead, we first use the empirical likelihood method to get the estimators $\hat{p}_{hi}$ and then solve (3.2) to get an M-estimator for θ . We call it empirical likelihood moment (ELM) estimator.

Next we establish the asymptotic properties of the ELM-estimator $\hat{\theta}_n$. Denote

$$\lambda_F(\theta) = \int \psi(x, \theta) dF(x), \quad (3.3)$$

$$Q_n(\theta) = \sum_h W_h \sum_i \hat{p}_{hi} \psi(x_{hi}, \theta) = \sum_h W_h \frac{1}{n_h} \sum_i \frac{\psi(x_{hi}, \theta)}{1 + m_h t^\tau (g(x_{hi}) - \hat{g}_h)}, \quad (3.4)$$

with t being the solution to (2.3) and

$$\psi'(x, \theta_0) = \frac{\partial \psi(x, \theta)}{\partial \theta} \Big|_{\theta=\theta_0}.$$

We first establish the consistency of the ELM-estimator $\hat{\theta}_n$. Given that the equation $\lambda_F(\theta) = 0$ has a root θ_0 and the equation $Q_n(\theta) = 0$ has a root $\hat{\theta}_n$, we wish to show that $\hat{\theta}_n \rightarrow \theta_0$ in probability under appropriate conditions. In this paper, the norm of a $p \times q$ matrix $A = (a_{ij})_{p \times q}$ is defined by $\|A\| = \left(\sum_{i=1}^p \sum_{j=1}^q a_{ij}^2 \right)^{\frac{1}{2}}$, $p, q \geq 1$.

Theorem 1 establishes the asymptotic consistency of the ELM-estimator $\hat{\theta}_n$.

Theorem 1 *Let θ_0 be an isolated root of $\lambda_F(\theta) = 0$, i.e., θ_0 is the only root in a neighbourhood of θ_0 . Suppose that $Eg(x)g^\tau(x)$ is positive definite, and that in a neighbourhood of θ_0 , $E[|\psi(x, \theta)|]$ is finite. Furthermore, assume that $\psi(x, \theta)$ is continuous in θ and monotone in θ , or $\psi(x, \theta)$ is continuous in θ and bounded, and $\lambda_F(\theta)$ changes sign uniquely in a neighbourhood of θ_0 . Then the equation $Q_n(\theta) = 0$ has a solution $\{\hat{\theta}_n\}$ which converges to θ_0 in probability.*

Next we establish the asymptotic normality of the ELM-estimator $\hat{\theta}_n$. We give the following two theorems.

Theorem 2 *Let θ_0 be an isolated root of $\lambda_F(\theta) = 0$, $\psi(x, \theta)$ be monotone in θ . Suppose that $\lambda_F(\theta)$ is differentiable at $\theta = \theta_0$, with $\lambda'_F(\theta_0) \neq 0$. Furthermore, suppose that $E\psi^2(x, \theta)$ and $E[|g(x)| |\psi(x, \theta)|]$ are finite for θ in a neighborhood of θ_0 , and that $E\psi^2(x, \theta)$ and $E[\psi(x, \theta)g(x)]$ are continuous at $\theta = \theta_0$. Then if $\hat{\theta}_n$ is any solution sequence of the equation $Q_n(\theta) = 0$, we have*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0, \sigma_{r1}^2),$$

where

$$\sigma_{r1}^2 = \frac{\sum_h k_h W_h^2 \text{Var}_h(\psi(x, \theta_0)) - \sigma_{\psi,g}^r \sigma^{-1} \sigma_{\psi,g}}{[\lambda'_F(\theta_0)]^2}, \quad (3.5)$$

with

$$\sigma_{\psi,g} = \sum_h k_h W_h^2 E_h[\psi(x, \theta_0)(g(x) - E_h g(x))].$$

The next theorem does not require monotonicity of $\psi(x, \theta)$ in θ . However, we need to have some smoothness conditions on $\psi(x, \theta)$ and higher moment requirements on $g(x)$.

Theorem 3 *Let θ_0 be an isolated of $\lambda_F(\theta) = 0$. Let $\partial\psi(x, \theta)/\partial\theta$ be continuous at $\theta = \theta_0$ uniformly in x . Suppose that $E[g(x)g(x)^r]$ is positive definite.*

$$E \|g(x)\|^4 < \infty, \quad E[\psi^2(x, \theta_0)] < \infty,$$

$$E[\psi'(x, \theta_0)] = \sum_h k_h W_h E_h \psi'(x, \theta_0).$$

is finite and nonzero, and $E[\|g(x)\|^2 |\psi(x, \theta_0)|^2] < \infty$. Furthermore, suppose that in a neighborhood of θ_0 , $|\psi'(x, \theta)| \leq G(x)$, where $G(\cdot)$ satisfies $E[\|g(x)\|^2 G(x)] < \infty$. Let $\{\hat{\theta}_n\}$ be a consistent solution sequence of the equation $Q_n(\theta) = 0$. then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0, \sigma_{r2}^2),$$

where

$$\sigma_{r2}^2 = \frac{\sum_h k_h W_h^2 \text{Var}_h(\psi(x, \theta_0)) - \sigma_{\psi,g}^r \sigma^{-1} \sigma_{\psi,g}}{[E\psi'(x, \theta_0)]^2}. \quad (3.6)$$

5.4 Proofs

Proof of Theorem 1

Let $\epsilon > 0$ be given. From Lemmas 1 and 2 of Chapter 3, we have

$$|g(x_{hi}) - \hat{g}_h| = o_p(n^{-\frac{1}{2}}), \quad t = O_p(n^{-\frac{1}{2}}).$$

Hence,

$$\max_{1 \leq h \leq H} \max_{1 \leq i \leq n_h} |t^\tau[g(x_{hi}) - \hat{g}_h]| = o_p(1). \quad (4.1)$$

Further,

$$\begin{aligned} Q_n(\theta_0 + \epsilon) &= \sum_h W_h \sum_i \frac{1}{n_h[1 + m_h t^\tau(g(x_{hi}) - \hat{g}_h)]} \psi(x_{hi}, \theta_0 + \epsilon) \\ &= \sum_h W_h \frac{1}{n_h} \sum_i \psi(x_{hi}, \theta_0 + \epsilon) \\ &\quad - \sum_h W_h \frac{m_h}{n_h} \sum_i \frac{t^\tau[g(x_{hi}) - \hat{g}_h]}{1 + m_h t^\tau(g(x_{hi}) - \hat{g}_h)} \psi(x_{hi}, \theta_0 + \epsilon). \end{aligned} \quad (4.2)$$

Using Lemma 1, (4.1) and the strong law of large numbers, the last term in (4.2) has absolute value bounded by

$$\begin{aligned} & \left| \sum_h W_h \frac{m_h}{n_h} \sum_i \frac{t^\tau[g(x_{hi}) - \hat{g}_h]}{1 + m_h t^\tau(g(x_{hi}) - \hat{g}_h)} \psi(x_{hi}, \theta_0 + \epsilon) \right| \\ & \leq (\max_h \max_i |t^\tau[g(x_{hi}) - \hat{g}_h]|) \sum_h W_h \frac{m_h}{n_h} \sum_i \frac{|\psi(x_{hi}, \theta_0 + \epsilon)|}{1 + m_h t^\tau(g(x_{hi}) - \hat{g}_h)} \\ & = o_p(1) O_p(1) = o_p(1). \end{aligned}$$

Therefore, applying the weak law of large numbers gives

$$Q_n(\theta_0 + \epsilon) \xrightarrow{p} E\psi(x, \theta_0 + \epsilon). \quad (4.3)$$

Without loss of generality, assume that $\psi(x, \theta)$ is non-increasing in θ , then $E\psi(x, \theta)$ is a non-increasing function of θ . Since θ_0 is an isolated root of $\lambda_F(\theta) = 0$, we have

$E\psi(x, \theta_0 + \epsilon) < E\psi(x, \theta_0) < E\psi(x, \theta_0 - \epsilon)$ for sufficiently small ϵ . Thus (4.3) implies that

$$\lim_{n \rightarrow \infty} P(Q_n(x, \theta_0 + \epsilon) < 0 < Q_n(x, \theta_0 - \epsilon)) = 1. \quad (4.4)$$

Since $Q_n(\theta)$ is a continuous function of θ for each n , it follows from (4.4) that there exists a sequence $\{\hat{\theta}_n\}$ such that $Q_n(\hat{\theta}_n) = 0$ and $\hat{\theta}_n \xrightarrow{P} \theta_0$.

For the case that $\psi(x, \theta)$ is continuous in θ and is bounded, it follows by the dominated convergence theorem that $\lambda_F(\theta)$ is continuous in θ , we can complete the proof as with part one of the theorem.

Proof of Theorem 2

The consistency of $\hat{\theta}_n$ to θ_0 is guaranteed by Theorem 1. Now we assume that $\psi(x, \theta)$ is non-increasing in θ , so that $Q_n(\theta)$ is non-increasing. Thus for every y ,

$$P(Q_n(y_n) < 0) \leq P\left(\frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{\sigma_{r1}} \leq y\right) = P(\hat{\theta}_n \leq y_n) \leq P(Q_n(y_n) \leq 0),$$

where $y_n = \theta_0 + y\sigma_{r1}/\sqrt{n}$. Thus, to prove the theorem, it suffices to show that for every y ,

$$\lim_{n \rightarrow \infty} P(Q_n(y_n) < 0) = \lim_{n \rightarrow \infty} P(Q_n(y_n) \leq 0) = \Phi(y), \quad (4.5)$$

where $\Phi(\cdot)$ is the standard normal distribution function. Now using Lemma 1. (4.1) and the assumptions given in the theorem, we can show that

$$\begin{aligned} Q_n(y_n) &= \sum_h W_h \sum_i \frac{\psi(x_{hi}, y_n)}{n_h [1 + m_h t^\tau (g(x_{hi}) - \hat{g}_h)]} \\ &= \sum_h \frac{W_h}{n_h} \sum_i \psi(x_{hi}, y_n) - t^\tau \left[\sum_h \frac{W_h m_h}{n_h} \sum_i \psi(x_{hi}, y_n) (g(x_{hi}) - \hat{g}_h) \right] \\ &\quad + \sum_h \frac{W_h m_h^2}{n_h} \sum_i \frac{t^\tau (g(x_{hi}) - \hat{g}_h) (g(x_{hi}) - \hat{g}_h)^\tau t}{1 + m_h t^\tau (g(x_{hi}) - \hat{g}_h)} \psi(x_{hi}, y_n) \end{aligned}$$

$$\begin{aligned}
&= \sum_h \frac{W_h}{n_h} \sum_i \psi(x_{hi}, y_n) - t^\tau \left[\sum_h \frac{W_h m_h}{n_h} \sum_i \psi(x_{hi}, y_n) (g(x_{hi}) - \hat{g}_h) \right] \\
&\quad + o_p(1) \\
&= \sum_h W_h n_h^{-1} \sum_i z_{hi,n} + o_p(1)
\end{aligned}$$

where

$$\begin{aligned}
z_{hi,n} &= \psi(x_{hi}, y_n) - \sigma^*(g(x_{hi}) - \hat{g}_h), \\
\sigma^* &= \sum_h W_h^2 k_h E[\psi(x_{hi}, y_n)(g(x_{hi}) - E_h g(x))^\tau] \sigma^{-1}.
\end{aligned}$$

Thus to prove (4.5) it suffices to show that

$$\lim_{n \rightarrow \infty} P\left(\sum_h \frac{W_h}{n_h} \sum_i z_{hi,n} \leq 0\right) = \Phi(y),$$

or equivalently,

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_h W_h n_h^{-1} \sum_i (z_{hi,n} - E z_{hi,n})}{\sqrt{V_n}} \leq \frac{-\sum_h W_h E \psi(x_h, y_n)}{\sqrt{V_n}}\right) = \Phi(y), \quad (4.6)$$

where V_n is the variance of $\sum_h W_h n_h^{-1} \sum_i (z_{hi,n} - E z_{hi,n})$. Now

$$\lim_{n \rightarrow \infty} \sqrt{n} \left[\sum_h W_h E \psi(x_h, y_n) \right] = y \sigma_{r1} \lambda'_F(\theta_0),$$

$$\begin{aligned}
&\lim_{n \rightarrow \infty} n V_n \\
&= \sum_h k_h W_h^2 \text{Var}_h(\psi(x, \theta_0)) - \sigma^* \sum_h k_h W_h^2 E_h \psi(x_h, \theta_0) [g(x) - E_h g(x)] \\
&= \sum_h k_h W_h^2 \text{Var}_h(\psi(x, \theta_0)) - \left(\sum_h W_h^2 k_h E_h [\psi(x, \theta_0) [g(x) - E_h g(x)]^\tau] \right) \\
&\quad \cdot \sigma^{-1} \left(\sum_h W_h^2 k_h E_h [\psi(x, \theta_0) [g(x) - E_h g(x)]] \right) \\
&= [\lambda'_F(\theta_0)]^2 \sigma_{r1}^2.
\end{aligned}$$

Hence

$$\lim_{n \rightarrow \infty} \sqrt{n V_n} = -\sigma_{r1} \lambda'_F(\theta_0).$$

In order to prove (4.6), it therefore suffices to show that

$$\lim_{n \rightarrow \infty} P \left(\frac{\sum_h W_h n_h^{-1} \sum_i (z_{hi,n} - E z_{hi,n})}{\sqrt{V_n}} \leq y \right) = \Phi(y). \quad (4.7)$$

or equivalently

$$\lim_{n \rightarrow \infty} P \left(\frac{\sum_h W_h n_h^{-1} \sum_i (y_{hi,n} - E y_{hi,n})}{\sqrt{V'_n}} \leq y \right) = \Phi(y). \quad (4.8)$$

where

$$y_{hi,n} = \psi(x_{hi}, y_n) - \sigma^* g(x_{hi}),$$

and V'_n is the variance of $\sum_h W_h n_h^{-1} \sum_i (y_{hi,n} - E y_{hi,n})$.

Since $y_{hi,n} - E y_{hi,n}$, $1 \leq i \leq n$, are independent and identically distributed with mean 0, by appealing to the central limit theorem, it suffices to verify the condition

$$\lim_{n \rightarrow \infty} \sum_h \frac{1}{n_h} \sum_i \int_{|y_{hi,n} - E y_{hi,n}| > \epsilon n_h \sqrt{V_n}} W_h^2 \frac{(y_{hi,n} - E y_{hi,n})^2}{n_h V'_n} dF_h(x) = 0,$$

for every $\epsilon > 0$, or equivalently, for $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n_h V'_n} \int_{|y_{hi,n} - E y_{hi,n}| > \epsilon n_h \sqrt{V_n}} (y_{hi,n} - E y_{hi,n})^2 dF_h(x) = 0.$$

Hence it suffices to prove for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n_h V'_n} \int_{h_n(x) > \epsilon n_h \sqrt{V_n}} h_n^2(x) dF_h(x) = 0,$$

where

$$h_n(x) = |\psi(x_{hi}, y_n)| + \|\sigma^*\| \|g(x_{hi})\|.$$

For every $\alpha > 0$, we have for each x and for n sufficiently large that

$$\psi(x, \theta_0 + \alpha) \leq \psi(x, y_n) \leq \psi(x, \theta_0 - \alpha).$$

Thus, if $u(x) = \max\{|\psi(x, \theta_0 + \alpha)|, |\psi(x, \theta_0 - \alpha)|\}$, then

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n_h V_n'} \int_{h_n(x) > \epsilon n_h \sqrt{V_n}} h_n(x)^2 dF_h(x) \\ & \leq \lim_{n \rightarrow \infty} \frac{1}{n_h V_n'} \int_{h_n(x) > \epsilon n_h \sqrt{V_n}} w(x)^2 dF_h(x) = 0, \end{aligned}$$

where

$$w(x) = u(x) + \|\sigma^*\| \|g(x)\|.$$

The proof is complete.

Proof of Theorem 3

Since $Q_n(\hat{\theta}_n) = 0$ and $Q_n(\theta)$ is differentiable in θ , we have

$$\begin{aligned} 0 &= Q_n(\hat{\theta}_n) = \sum_h W_h \sum_i \frac{\psi(x_{hi}, \hat{\theta}_n)}{n_h [1 + m_h t^r (g(x_{hi}) - \hat{g}_h)]} \\ &= \sum_h \frac{W_h}{n_h} \sum_i \psi(x_{hi}, \hat{\theta}_n) \\ &\quad \cdot \left[1 - m_h t^r (g(x_{hi}) - \hat{g}_h) + \frac{t^r (g(x_{hi}) - \hat{g}_h)(g(x_{hi}) - \hat{g}_h)^r t}{1 + m_h t^r (g(x_{hi}) - \hat{g}_h)} \right] \\ &= \sum_h \frac{W_h}{n_h} \sum_i \psi(x_{hi}, \theta_0) + \sum_h \frac{W_h}{n_h} \sum_i \psi'(x_{hi}, \hat{\theta}_n^*) (\hat{\theta}_n - \theta_0) \\ &\quad - t^r \left[\sum_h \frac{W_h m_h}{n_h} \sum_i \psi(x_{hi}, \theta_0) (g(x_{hi}) - \hat{g}_h) \right] \\ &\quad - t^r \left[\sum_h \frac{W_h m_h}{n_h} \sum_i \psi'(x_{hi}, \hat{\theta}_n^*) (g(x_{hi}) - \hat{g}_h) \right] (\hat{\theta}_n - \theta_0) \\ &\quad + \sum_h \frac{W_h}{n_h} \sum_i \frac{t^r (g(x_{hi}) - \hat{g}_h)(g(x_{hi}) - \hat{g}_h)^r t}{1 + m_h t^r (g(x_{hi}) - \hat{g}_h)} \psi(x_{hi}, \hat{\theta}_n), \end{aligned}$$

where $\hat{\theta}_n^*$ satisfies $|\hat{\theta}_n^* - \theta_0| \leq |\hat{\theta}_n - \theta_0|$. Thus

$$\begin{aligned} \hat{\theta}_n - \theta_0 &= \left(\sum_h \frac{W_h}{n_h} \sum_i \psi'(x_{hi}, \hat{\theta}_n^*) - t^r \left[\sum_h \frac{W_h m_h}{n_h} \sum_i \psi'(x_{hi}, \hat{\theta}_n^*) (g(x_{hi}) - \hat{g}_h) \right] \right)^{-1} \\ &\quad \cdot \left(\sum_h \frac{W_h}{n_h} \sum_i \psi(x_{hi}, \theta_0) - t^r \left[\sum_h \frac{W_h m_h}{n_h} \sum_i \psi(x_{hi}, \theta_0) (g(x_{hi}) - \hat{g}_h) \right] \right) \end{aligned}$$

$$+ \sum_h \frac{W_h}{n_h} \sum_i \frac{t^\tau (g(x_{hi}) - \hat{g}_h)(g(x_{hi}) - \hat{g}_h)^\tau t}{1 + m_h t^\tau (g(x_{hi}) - \hat{g}_h)} \psi(x_{hi}, \hat{\theta}_n).$$

It can be shown that

$$\begin{aligned} & t^\tau \sum_h \frac{W_h m_h}{n_h} \sum_i \psi(x_{hi}, \theta_0)(g(x_{hi}) - \hat{g}_h) \\ &= \bar{g}_{st}^\tau \sigma^{-1} \sum_h n W_h^2 n_h^2 \sum_i \psi(x_{hi}, \theta_0)(g(x_{hi}) - \hat{g}_h) + O_p(n^{-1}) \\ &= \sum_h k_h W_h^2 E_h \psi(x, \theta_0)(g(x) - E_h g(x))^\tau \sigma^{-1} \bar{g}_{st} + O_p(n^{-1}), \\ & \sum_h \frac{W_h}{n_h} \sum_i \frac{t^\tau (g(x_{hi}) - \hat{g}_h)(g(x_{hi}) - \hat{g}_h)^\tau t}{1 + m_h t^\tau (g(x_{hi}) - \hat{g}_h)} \psi(x_{hi}, \hat{\theta}_n) = O_p(n^{-1}), \\ & -t^\tau \left[\sum_h \frac{W_h m_h}{n_h} \sum_i \psi'(x_{hi}, \hat{\theta}_n^*)(g(x_{hi}) - \hat{g}_h) \right] = O_p(n^{-\frac{1}{2}}), \\ & \sum_h \frac{W_h}{n_h} \sum_i \psi'(x_{hi}, \hat{\theta}_n^*) = \sum_h \frac{W_h}{n_h} \sum_i \psi'(x_{hi}, \theta_0) + o_p(1), \end{aligned}$$

hence,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\sum_h \frac{W_h}{n_h} \sum_i \psi(x_{hi}, \theta_0) - \sigma_{\psi, g}^\tau \sigma \bar{g}_{st}}{E \psi'(x, \theta_0)} + o_p(1).$$

Therefore, it follows from the central limit theorem and Slutsky's theorem that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_L N(0, \sigma_\tau^2),$$

where

$$\begin{aligned} \sigma_\tau^2 &= \frac{\sum_h k_h W_h^2 \text{Var}_h(\psi(x, \theta_0) - \sigma_{\psi, g}^\tau \sigma g(x, \theta_0))}{[E \psi'(x, \theta_0)]^2} \\ &= \sigma_{\tau 2}^2. \square \end{aligned}$$

Chapter 6

Empirical Likelihood Inference in the Presence of Measurement Error

6.1 Introduction

Suppose that we want to measure a characteristic of interest of a target population. We have several different “instruments”, one of them is “accurate”, or it has smaller measurement error compared to the other ones. In practice we treat this instrument as “perfect”, or no measurement error. All other instruments will have larger measurement error and we treat them as “imperfect”. Here we use “instrument” to refer the method used to measure the characteristic. It may include factors such as the real instrument used, the personnel involved, working environment, etc. It may be due to the cost, or lack of highly trained personnel for using the perfect instrument.

We use imperfect instruments to measure some samples drawn from the population, and use perfect instrument to measure other samples from the population. We combine all the data, both imperfect and perfect, to make inference on the parameters of the population. Of course, the inference should be more accurate compared to using only the perfect measurement data.

Survey researchers have long been cognizant of the ill effects of measurement error on estimators of means and totals of finite populations. In general, compared with perfect measurements, imperfect but unbiased measurements increase the variance, but do not affect the bias of estimators of means and totals. However, the sample cumulative distribution function(CDF) is no longer an unbiased or consistent estimator of the population CDF even if the measurement error has mean zero. Several authors discussed these problems. Luo, Stokes and Sager gave a good summary of history and current development in this area, though almost all the papers listed there only deal with imperfect sample without calibration sample, i.e., perfect sample. The paper by Luo, etc., may be the first to handle the data collected with measurement error with a calibration sample. In their paper, they used the simulation to compare the performance of several estimates of CDF. The estimators they considered are: the difference, ratio, regression, weighted average of the empirical CDF's of the imperfect and perfect measurements, etc. Their proposed estimators are linear combinations of the empirical CDF's from the imperfect and perfect samples. The weights are chosen so that the resulting estimator has the smallest possible variance subject to an unbiasedness constraint. These weights are also estimated from the data. They did not give any theory about the proposed estimators such as asymptotic variances. They also pointed out that the proposed

weighted linear estimators of CDF may not monotone.

In this chapter we propose to use the empirical likelihood method to make inference on parameters of interest by taking all the data into account, treating the imperfect measurements as auxiliary information. The empirical likelihood estimator of CDF is monotone, asymptotically more efficient than the empirical CDF from perfect sample only, and asymptotically normal distributed. Section 6.2 shows how to use the empirical likelihood to solve the problems above. Section 6.3 discusses the associated asymptotic results., while Section 6.4 discusses likelihood ratio testing. Several examples will be given in section 6.5. Proofs are delegated to Section 6.6.

6.2 Empirical likelihood method in the presence of measurement error

Suppose that there are H different instruments used in measuring a characteristic of interest. The distribution function associated with instrument h ($h = 1, \dots, H$) is $F_h(x) = P(x_h < a)$ with unknown parameter θ , where θ is a p -vector, and the H -th measuring instrument is taken as perfect. Note that we have H different distributions due to possibly different measurement errors in the instruments, although the same characteristics, x_H , are measured by all the instruments. These H different distributions are assumed to be related by the common parameter θ . We also assume that information about θ and F_h is available in the form of $r_h \geq p$ functionally independent unbiased estimation functions, that is functions $g_h(x_h, \theta)$ such that

$$E_{F_h}(g_h(x_h, \theta)) = 0, \quad (2.1)$$

where $g_h(x, \theta)$ is a r_h -dimension vector function. We further suppose that measurements $x_{h1}, \dots, x_{hn_h}$ are *i.i.d.* measurements obtained by instrument h , different measuring by different instruments are independent. The sampling scheme considered here is different from the method of Luo et al (1996, unpublished) and others. They used two-phase sampling with the imperfect sample as the first phase sample and the second phase sample taken from the first phase sample. This means that the second phase sample is dependent on the first phase sample, and usually the second phase sample is a very small portion of the first phase sample. We consider independent samples here for two main reasons. First, in the case different instruments are involved in measuring, the makers may have done a lot of testing and will give the accuracies of different instruments. Or a great number of testing can be made to know the difference between different instruments before practical measuring. The same thing can be said about different degrees of trained personnel. Our method can take this kind of information into account, and make more accurate inference. Second, as pointed by Luo et al (1996, unpublished) and others, when the second phase sample is quite small compared to the first phase sample, the two samples can be treated as independent samples. Actually their recommended estimator for the CDF is based on independent samples.

The empirical likelihood based on the independent samples

$$\{x_{h1}, \dots, x_{hn_h}; h = 1, \dots, H\}$$

is given by

$$L(F_1, \dots, F_H) = \prod_{i=1}^H \prod_{j=1}^{n_h} p_{hi}, \quad (2.2)$$

where $p_{hi} = Pr(x_h = x_{hi})$ and $\sum_{i=1}^{n_h} p_{hi} = 1$ for each h . Only those F_h distributions

which have an atom of probability on each x_{hi} have nonzero likelihood. (2.2) is maximized by the empirical distribution function $F_{h,n_h}(x) = n_h^{-1} \sum_i 1(x_{hi} < x)$. The empirical likelihood ratio is then defined as

$$R(F_1, \dots, F_H) = L(F_1, \dots, F_H) / L(F_{1,n_1}, \dots, F_{H,n_H}),$$

which reduces to

$$R(F_1, \dots, F_H) = \prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi}. \quad (2.3)$$

We remark that formulas here and elsewhere in this chapter also do not require that the x_{hi} 's be distinct. Since we are interested in estimating the parameter θ , and we know the estimating equation (2.1), we define the empirical likelihood ratio function

$$R_E(\theta) = \sup \{ \prod_{h=1}^H \prod_{i=1}^{n_h} n_h p_{hi} \mid p_{hi} \geq 0, \sum_i p_{hi} = 1, \sum_i p_{hi} g_{hi}(\theta) = 0 \}, \quad (2.4)$$

where $g_{hi}(\theta) = g_h(x_{hi}, \theta)$ for all h and i . Here function $g_{hi}(\theta)$ depends on h also, which is different from the one discussed in chapters 2 to 4.

As noted before, for any given θ and h , $\prod_{i=1}^{n_h} p_{hi}$ can be maximized, provided 0 is inside the convex hull of the point $g_{h1}(\theta), \dots, g_{hn_h}(\theta)$. The maximum of $\prod_{i=1}^{n_h} p_{hi}$ subject to the constraints $p_{hi} \geq 0$, $\sum_i p_{hi} = 1$, and $\sum_i p_{hi} g_{hi}(\theta) = 0$ is attained when

$$p_{hi} = p_{hi}(\theta) = n_h^{-1} \{1 + t_h^\tau g_{hi}(\theta)\}^{-1}. \quad (2.5)$$

where $t_h = t_h(\theta)$ is a $r_h \times 1$ vector given as the solution to

$$\frac{1}{n_h} \sum_{i=1}^{n_h} \{1 + t_h^\tau g_{hi}(\theta)\}^{-1} g_{hi}(\theta) = 0. \quad (2.6)$$

Hence the left side of (2.4) is

$$R_E(\theta) = \prod_h \prod_i \{1 + t_h^\tau g_{hi}(\theta)\}^{-1}, \quad (2.7)$$

and the empirical (negative) log-likelihood ratio of θ is

$$l_E(\theta) = \sum_h \sum_i \log[1 + t_h^\tau g_{hi}(\theta)], \quad (2.8)$$

with t_h being the solution to (2.6). We can minimize $l_E(\theta)$ to obtain an estimator $\tilde{\theta}$ of the parameter θ , called the maximum empirical likelihood ratio estimator (MELRE). In addition, this yields estimators $\tilde{p}_{hi}$, from (2.5), and an estimator for the true distribution function F_H as

$$\tilde{F}_{H,n_H}(x) = \sum_{i=1}^{n_H} \tilde{p}_{Hi} 1(x_{Hi} < x). \quad (2.9)$$

$\tilde{\theta}$ is obtained by solving

$$\frac{\partial l_E(\theta)}{\partial \theta} = \sum_h \sum_i \frac{1}{1 + t_h^\tau g_{hi}(\theta)} \left(\frac{\partial g_{hi}(\theta)}{\partial \theta^\tau} \right) t_h = 0 \quad (2.10)$$

and (2.6) together.

In the following section we will prove the existence of $\tilde{\theta}$ and study the asymptotic properties of $\tilde{\theta}$ and $\tilde{F}_{H,n_H}(x)$.

6.3 Asymptotic properties of MELRE

First, we give the conditions for the existence of $\tilde{\theta}$ which will minimize $l_E(\theta)$ defined by (2.8).

In the following, we will use $\|\cdot\|$ to denote Euclidean norm and $n = \sum_h n_h$.

Lemma 1. *Suppose as $n \rightarrow \infty$, $n/n_h \rightarrow k_h > 0$ for all h . And suppose that in a neighborhood of the true value θ_0 , $\text{Var}[g_h(x_h, \theta_0)] = \sigma_h(\theta_0) > 0$ for all h . $\|\partial g_h(x_h, \theta)/\partial \theta\|$ and $\|g_h(x_h, \theta)\|^4$ are bounded by some integrable function $G(x)$*

in this neighbourhood, and the rank of $E[\partial g_h(x_h, \theta)/\partial \theta]$ is p . Then, as $n \rightarrow \infty$, with probability 1 $l_E(\theta)$ attains its minimum value at some point $\tilde{\theta}$ in the interior of the ball $\|\theta - \theta_0\| \leq n^{-\frac{1}{3}}$, and $\tilde{\theta}$ and $\tilde{t}_h = \tilde{t}_h(\tilde{\theta})$ satisfy

$$Q_h(\tilde{\theta}, \tilde{t}_h) = 0, \quad h = 1, \dots, H, \quad Q_{H+1}(\tilde{\theta}, \tilde{t}_1, \dots, \tilde{t}_H) = 0, \quad (3.1)$$

where

$$Q_h(\theta, t_h) = \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{g_{hi}(\theta)}{1 + t_h^T g_{hi}(\theta)}, \quad h = 1, \dots, H, \quad (3.2)$$

$$Q_{H+1}(\theta, t_1, \dots, t_H) = \sum_h \sum_i \frac{1}{1 + t_h^T g_{hi}(\theta)} \left(\frac{\partial g_{hi}(\theta)}{\partial \theta} \right) t_h = 0 \quad (3.3)$$

Theorem 1. In addition to the conditions of Lemma 1, we further assume that $\frac{\partial^2 g_h(x_h, \theta)}{\partial \theta \partial \theta^T}$ is continuous in θ in a neighbourhood of the true value θ_0 , then if $\|\frac{\partial^2 g_h(x_h, \theta)}{\partial \theta \partial \theta^T}\|$ is bounded by some integrable function $G(x)$ in the neighbourhood for all h , then

$$\sqrt{n}(\tilde{\theta} - \theta_0) \rightarrow N(0, V), \quad \sqrt{n}(\tilde{t}_h - 0) \rightarrow N(0, Q),$$

$$\sqrt{n}(\tilde{F}_{H, n_H}(x) - F_H(x)) \rightarrow N(0, W),$$

where $\tilde{F}_{H, n_H}(x) = \sum_i \tilde{p}_{Hi} 1(x_{Hi} < x)$, $\tilde{p}_{Hi} = \frac{1}{n_H} \frac{1}{1 + \tilde{t}_H^T g_{Hi}(\tilde{\theta})}$, and V, Q, W are defined in the proof.

Since usually $n_h \gg n_H$ for $h = 1, \dots, H-1$, we can see the main part of the variance of $\tilde{F}_{H, n_H}(x)$ will come from the first three term of (6.4), which is smaller than the variance of the estimator $F_{H, n_H}(x)$, the cumulative distribution function from the precise data $x_{H1}, \dots, x_{Hn_H}$ only. Some further discussion will be given in Section 6.5.

The estimators for variance of $\tilde{\theta}$ and $\tilde{F}_{H,n_H}(x)$ can be obtained by replacing each component of V and W by their obvious estimators. These variance estimators are asymptotically correct.

6.4 Likelihood Ratio Tests

Several testing issues arise naturally from the application point of view. We first consider how to test hypothesis $H_0 : \theta = \theta_0$.

Theorem 2. *In the semi-parametric model (2.1), under the conditions of theorem 1, for testing $H_0 : \theta = \theta_0$, the empirical likelihood ratio statistic*

$$R = 2 \left(\sum_h \sum_i \log[1 + t_h^T(\theta_0) g_{hi}(\theta_0)] - \sum_h \sum_i \log[1 + \tilde{t}_h^T(\tilde{\theta}) g_{hi}(\tilde{\theta})] \right)$$

is asymptotically distributed as χ_p^2 if $n_1 = \dots = n_H$.

Now we turn to test the model (2.1), i.e.,

$$Eg_h(x_h, \theta) = 0, \quad h = 1, \dots, H. \quad (4.1)$$

Corollary 2.1 *In order to test the model (4.1), we can use the empirical likelihood ratio statistic*

$$R_1 = 2 \sum_h \sum_i \log[1 + \tilde{t}_h^T g_{hi}(\tilde{\theta})].$$

Under the conditions of theorem 2 above, R_1 is asymptotically distributed as χ_{r-p}^2 if (4.1) is correct, and $n_1 = \dots = n_H = n$, where $r = \sum_{h=1}^H r_h$,

6.5 Examples

We present several illustrations of the estimation procedures. Procedures about how to solve equations (3.1) through (3.3) will be given. Large-sample aspects of these estimators will be discussed.

Example 1. Common Mean Model

Suppose we have two different instruments, i.e., $H = 2$, and we only know that they are unbiased, i.e.,

$$Ex_1 = \theta, Ex_2 = \theta,$$

and x_2 refers to the perfect measurement. Suppose that $x_{11}, \dots, x_{1n_1}$ and $x_{21}, \dots, x_{2n_2}$ are independent samples from x_1 and x_2 respectively. Let

$$g_1(x_1, \theta) = x_1 - \theta, g_2(x_2, \theta) = x_2 - \theta,$$

then

$$\frac{\partial g_1}{\partial \theta} = -1 = \frac{\partial g_2}{\partial \theta} = E \frac{\partial g_1}{\partial \theta} = E \frac{\partial g_2}{\partial \theta}.$$

and equation (3.3) becomes

$$n_1 t_1 + n_2 t_2 = 0, \text{ i.e., } t_2 = -\frac{n_1}{n_2} t_1. \quad (5.1)$$

Substituting (5.1) into (3.2), we get

$$\frac{1}{n_1} \sum_{i=1}^{n_1} \frac{x_{1i} - \theta}{1 + t(x_{1i} - \theta)} = 0, \quad (5.2)$$

$$\frac{1}{n_2} \sum_{i=1}^{n_2} \frac{x_{2i} - \theta}{1 - \frac{n_1}{n_2} t(x_{2i} - \theta)} = 0. \quad (5.3)$$

Solving (5.2) and (5.3) by Newton method or using NAG Fortran library function with initial value $(\theta, t) = (\bar{x}, 0)$ often works, where $\bar{x} = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2$.

Let $\sigma_1 = \text{Var}(x_1)$, $\sigma_2 = \text{Var}(x_2)$. Since

$$\sum_h k_h E \frac{\partial g_h}{\partial \theta} \sigma_h^{-1} \frac{\partial g_h}{\partial \theta^\tau} = \sum_h k_h \sigma_h^{-1},$$

replacing k_h by n/n_h we get the asymptotic variance of $\tilde{\theta}$ as

$$\frac{V}{n} = \frac{1}{n} \left(\sum_h \sigma_h^{-1} \right)^{-1} \left(\sum_h \frac{n}{n_h} \sigma_h^{-1} \right)^{-1} \left(\sum_h \sigma_h^{-1} \right)^{-1} = \left(\frac{\sigma_2}{n_1} + \frac{\sigma_1}{n_2} \right) \frac{\sigma_1 \sigma_2}{(\sigma_1 + \sigma_2)^2},$$

which is the same as the variance of the optimal linear combination “estimator”

$$\hat{\theta} = \frac{\sigma_2}{\sigma_1 + \sigma_2} \bar{x}_1 + \frac{\sigma_1}{\sigma_1 + \sigma_2} \bar{x}_2.$$

Note $\hat{\theta}$ is actually not an estimator since we do not know the variances σ_1 and σ_2 , but we use it for comparison here and in the following.

As for the asymptotic variance of $\tilde{F}_{H,n_H}(x)$, i.e., W given by (6.4), after a little calculation we get

$$\begin{aligned} W &= F_2(x)[1 - F_2(x)] + B^2 \sigma_2^{-1} \left[\frac{\sigma_1^{-2} + k_1 k_2^{-1} \sigma_1 \sigma_2^{-3}}{(\sigma_1^{-1} + \sigma_2^{-1})^2} - \frac{2\sigma_1^{-1}}{\sigma_1^{-1} + \sigma_2^{-1}} \right] \\ &= F_2(x)[1 - F_2(x)] \\ &\quad + B^2 \frac{\sigma_2^{-3}}{(\sigma_1^{-1} + \sigma_2^{-1})^2} \left[- \left(\frac{\sigma_2}{\sigma_1} \right)^2 - 2 \left(\frac{\sigma_2}{\sigma_1} \right) + \frac{k_1}{k_2} \left(\frac{\sigma_2}{\sigma_1} \right) \right]. \end{aligned} \quad (5.4)$$

From (5.4) we see that if $\sigma_2 \ll \sigma_1$, then combining the two samples will not improve the efficiency of CDF estimator significantly compared to using only the perfect data. On the other hand, if σ_2/σ_1 is reasonable large, such as greater than 0.5, then for example, $n_2/n_1 = k_1/k_2 < 5/8$, we will have a significantly more efficient estimator for CDF.

Example 2 Additive Model

We consider the following model

$$Ex_1 = \theta_1, \quad Ex_2 = \theta_1, \quad Var(x_1) = \theta_2 + \sigma_0, \quad Var(x_2) = \theta_2,$$

i.e., the variance of imperfect measurement is larger than the variance of the perfect measurement by σ_0 , where $\sigma_0(> 0)$ is known. We note here that if $Ex_1 = \theta_1 + \nu_0$, where ν_0 is known, we can change the data x_1 to $x_1 - \nu_0$, and all the following discussion still applies.

Suppose two independent samples from x_1 and x_2 are $x_{11}, \dots, x_{1n_1}$ and $x_{21}, \dots, x_{2n_2}$ respectively. Denote $z_1 = (x_1, x_1^2)^\tau$, $z_2 = (x_2, x_2^2)^\tau$, $\theta = (\theta_1, \theta_2)^\tau$, $g_1(z_1, \theta) = (x_1 - \theta_1, x_1^2 - \theta_1^2 - \theta_2 - \sigma_0^2)^\tau$, $g_2(z_2, \theta) = (x_2 - \theta_1, x_2^2 - \theta_1^2 - \theta_2)^\tau$, then $Eg_1 = 0$, $Eg_2 = 0$,

$$\frac{\partial g_1}{\partial \theta} = \begin{pmatrix} -1 & -2\theta_1 \\ 0 & -1 \end{pmatrix} = \frac{\partial g_2}{\partial \theta} = E \frac{\partial g_1}{\partial \theta} = E \frac{\partial g_2}{\partial \theta},$$

and equation (3.3) becomes

$$n_1 t_{11} + n_2 t_{21} = 0, \quad n_1 t_{12} + n_2 t_{22} = 0.$$

or

$$t_{21} = -\frac{n_2}{n_1} t_{11}, \quad t_{22} = -\frac{n_2}{n_1} t_{12}.$$

Hence, (3.2) becomes

$$\frac{1}{n_1} \sum_{i=1}^{n_1} \frac{x_{1i} - \theta_1}{1 + t_{11}(x_{1i} - \theta_1) + t_{12}(x_{1i}^2 - \theta_1^2 - \theta_2 - \sigma_0^2)} = 0,$$

$$\frac{1}{n_1} \sum_{i=1}^{n_1} \frac{x_{1i}^2 - \theta_1^2 - \theta_2 - \sigma_0^2}{1 + t_{11}(x_{1i} - \theta_1) + t_{12}(x_{1i}^2 - \theta_1^2 - \theta_2 - \sigma_0^2)} = 0,$$

$$\frac{1}{n_2} \sum_{i=1}^{n_1} \frac{x_{2i} - \theta_1}{1 - n_2 n_1^{-1} t_{11}(x_{2i} - \theta_1) - n_2 n_1^{-1} t_{12}(x_{2i}^2 - \theta_1^2 - \theta_2)} = 0,$$

$$\frac{1}{n_2} \sum_{i=1}^{n_1} \frac{x_{2i}^2 - \theta_1^2 - \theta_2}{1 - n_2 n_1^{-1} t_{11}(x_{2i} - \theta_1) - n_2 n_1^{-1} t_{12}(x_{2i}^2 - \theta_1^2 - \theta_2)} = 0.$$

We can solve the above four equations by letting the initial values

$$\theta_{10} = \frac{n_1}{n_1 + n_2} \bar{x}_1 + \frac{n_2}{n_1 + n_2} \bar{x}_2, \quad \theta_{20} = \frac{n_1}{n_1 + n_2} S_{1n_1}^2 + \frac{n_2}{n_1 + n_2} S_{2n_2}^2$$

$$t_{11,0} = 0, \quad t_{12,0} = 0.$$

Using Newton's method will usually give us a solution to θ_1 , θ_2 , and hence $\tilde{p}_{hi}$, $\tilde{F}_{2n_2}(x)$.

In order to get an idea of the asymptotic properties of these estimators, we suppose x_1 and x_2 have normal distribution. Since

$$E(x_1 - \theta_1)(x_1^2 - \theta_1^2 - \theta_2 - \sigma_0) = 2\theta_1\theta_*, \quad Var(x_1^2) = 2\theta_*^2 + 4\theta_1^2\theta_*,$$

$$E(x_2 - \theta_1)(x_2^2 - \theta_1^2 - \theta_2) = 2\theta_1\theta_2, \quad Var(x_2^2) = 2\theta_2^2 + 4\theta_1^2\theta_2,$$

where $\theta_* = \theta_2 + \sigma_0$, therefore

$$Var(g_1(z_1, \theta)) = \sigma_1 = \begin{pmatrix} \theta_* & 2\theta_1\theta_* \\ 2\theta_1\theta_* & 2\theta_*^2 + 4\theta_1^2\theta_* \end{pmatrix},$$

$$Var(g_2(z_2, \theta)) = \sigma_2 = \begin{pmatrix} \theta_2 & 2\theta_1\theta_2 \\ 2\theta_1\theta_2 & 2\theta_2^2 + 4\theta_1^2\theta_2 \end{pmatrix},$$

$$\frac{\partial g_1}{\partial \theta} \sigma_1^{-1} \frac{\partial g_1}{\partial \theta^T} = \frac{1}{2\theta_*^2} \begin{pmatrix} 2\theta_* & 0 \\ 0 & 1 \end{pmatrix}, \quad \frac{\partial g_2}{\partial \theta} \sigma_2^{-1} \frac{\partial g_2}{\partial \theta^T} = \frac{1}{2\theta_2^2} \begin{pmatrix} 2\theta_2 & 0 \\ 0 & 1 \end{pmatrix},$$

$$M_{22.1} = - \begin{pmatrix} \theta_2^{-1} + \theta_2^{-1} & 0 \\ 0 & (2\theta_2)^{-2} + (2\theta_2)^{-2} \end{pmatrix},$$

hence, by (6.4) we have $V = (k_1\theta_2 + k_2\theta_2)\theta_2\theta_2/(\theta_2 + \theta_2)^2$. If we replace k_h by n/n_h , then $Var(\tilde{\theta}) \approx (n_1^{-1}\theta_2 + n_2^{-1}\theta_2)\theta_2\theta_2/(\theta_2 + \theta_2)^2$, which is the same as the variance of the optimal linear combination estimator

$$\hat{\theta} = \frac{\theta_2}{\theta_2 + \theta_2} \bar{x}_1 + \frac{\theta_2}{\theta_2 + \theta_2} \bar{x}_2.$$

Example 3. Product Model

Suppose

$$Ex_1 = \theta_1, \quad Ex_2 = \theta_1, \quad Var(x_1) = c\theta_2, \quad Var(x_2) = \theta_2,$$

i.e., the ratio of variances of imperfect measurement to perfect measurement is a known constant($c > 0$). The two independent samples from x_1 and x_2 are $x_{11}, \dots, x_{1n_1}$ and $x_{21}, \dots, x_{2n_2}$ respectively. Denote

$$z_1 = (x_1, x_1^2)^T, \quad z_2 = (x_2, x_2^2)^T, \quad \theta = (\theta_1, \theta_2)^T.$$

$$g_1(z_1, \theta) = (x_1 - \theta_1, x_1^2 - \theta_1^2 - c\theta_2)^T, \quad g_2(z_2, \theta) = (x_2 - \theta_1, x_2^2 - \theta_1^2 - \theta_2)^T,$$

then

$$Eg_1 = 0, \quad Eg_2 = 0,$$

$$\frac{\partial g_1}{\partial \theta} = \begin{pmatrix} -1 & -2\theta_1 \\ 0 & -c \end{pmatrix} = E \frac{\partial g_1}{\partial \theta}, \quad \frac{\partial g_2}{\partial \theta} = \begin{pmatrix} -1 & -2\theta_1 \\ 0 & -1 \end{pmatrix} = E \frac{\partial g_2}{\partial \theta}.$$

Hence equations (3.1) through (3.3) become

$$cn_1t_{12} + n_2t_{22} = 0,$$

$$n_1 t_{11} + 2n_1 \theta_1 t_{12} + n_2 t_{21} + 2n_2 \theta_1 t_{22} = 0,$$

$$\frac{1}{n_1} \sum_{i=1}^{n_1} \frac{x_{1i} - \theta_1}{1 + t_{11}(x_{1i} - \theta_1) + t_{12}(x_{1i}^2 - \theta_1^2 - c\theta_2)},$$

$$\frac{1}{n_1} \sum_{i=1}^{n_1} \frac{x_{1i}^2 - \theta_1^2 - c\theta_2}{1 + t_{11}(x_{1i} - \theta_1) + t_{12}(x_{1i}^2 - \theta_1^2 - c\theta_2)} = 0,$$

$$\frac{1}{n_2} \sum_{i=1}^{n_2} \frac{x_{2i} - \theta_1}{1 + t_{21}(x_{2i} - \theta_1) + t_{22}(x_{2i}^2 - \theta_1^2 - \theta_2)} = 0,$$

$$\frac{1}{n_2} \sum_{i=1}^{n_2} \frac{x_{2i}^2 - \theta_1^2 - \theta_2}{1 + t_{21}(x_{2i} - \theta_1) + t_{22}(x_{2i}^2 - \theta_1^2 - \theta_2)} = 0.$$

Solving the above six equations by letting the initial values

$$\theta_{10} = \frac{n_1}{n_1 + n_2} \bar{x}_1 + \frac{n_2}{n_1 + n_2} \bar{x}_2, \quad \theta_{20} = \frac{n_1}{n_1 + n_2} S_{1n_1}^2 / c + \frac{n_2}{n_1 + n_2} S_{2n_2}^2$$

$$t_{1,0} = 0, \quad t_{2,0} = 0$$

and using Newton's method, give us a solution to θ_1 , θ_2 and hence $\tilde{p}_{hi}$, $\tilde{F}_{2n_2}(x)$.

Usually the solution $\tilde{p}_{hi}$ is non-negative.

In the following discussion, we suppose x_1 and x_2 are normally distributed. Then

$$Var(x_1) = c\theta_2, \quad E(x_1 - \theta_1)(x_1^2 - \theta_1^2 - c\theta_2) = 2c\theta_1\theta_2, \quad Var(x_1^2) = 2c^2\theta_2^2 + 4c\theta_1^2\theta_2,$$

$$E(x_2 - \theta_1)(x_2^2 - \theta_1^2 - \theta_2) = 2\theta_1\theta_2, \quad Var(x_2^2) = 2\theta_2^2 + 4\theta_1^2\theta_2.$$

Therefore

$$Var(g_1(z_1, \theta)) = \sigma_1 = \begin{pmatrix} c\theta_2 & 2c\theta_1\theta_2 \\ 2c\theta_1\theta_2 & 2c^2\theta_2^2 + 4c\theta_1^2\theta_2 \end{pmatrix},$$

$$\begin{aligned} \text{Var}(g_2(z_2, \theta)) &= \sigma_2 = \begin{pmatrix} \theta_2 & 2\theta_1\theta_2 \\ 2\theta_1\theta_2 & 2\theta_2^2 + 4\theta_1^2\theta_2 \end{pmatrix}, \\ \frac{\partial g_1}{\partial \theta} \sigma_1^{-1} \frac{\partial g_1}{\partial \theta^T} &= \frac{1}{2c\theta_2^2} \begin{pmatrix} 2\theta_2 & 0 \\ 0 & c \end{pmatrix}, \quad \frac{\partial g_2}{\partial \theta} \sigma_2^{-1} \frac{\partial g_2}{\partial \theta^T} = \frac{1}{2\theta_2^2} \begin{pmatrix} 2\theta_2 & 0 \\ 0 & 1 \end{pmatrix}, \\ M_{22.1} &= -\theta_2^{-1} \begin{pmatrix} 1 + c^{-1} & 0 \\ 0 & (2\theta_2)^{-1} \end{pmatrix}. \end{aligned}$$

Hence, by (6.4) we have $\text{Var}(\tilde{\theta}) \approx \left(\frac{c}{1+c}\right)^2 \left(\frac{1}{n_1 c} + \frac{1}{n_2}\right) \theta_2$, which is the same as the variance of the optimal linear combination estimator

$$\hat{\theta} = \frac{\theta_2}{c\theta_2 + \theta_2} \bar{x}_1 + \frac{c\theta_2}{c\theta_2 + \theta_2} \bar{x}_2.$$

6.6 Proofs

Proof of Lemma 1

Denote $\theta = \theta_0 + un^{-\frac{1}{3}}$ for $\theta \in \{\theta \mid \|\theta - \theta_0\| = n^{-\frac{1}{3}}\}$, where $\|u\| = 1$. First we give a lower bound for $l_E(\theta)$ on the surface of the ball. From Qin and Lawless (1994), when $\|\theta - \theta_0\| \leq n^{-\frac{1}{3}}$, we have

$$t_h(\theta) = S_{hn_h}^{-1}(\theta) \bar{g}_h(\theta) + o(n^{-\frac{1}{3}}) \text{ (a.s.)} = O(n^{-\frac{1}{3}}) \text{ (a.s.)}$$

uniformly about $\theta \in \{\theta \mid \|\theta - \theta_0\| \leq n^{-\frac{1}{3}}\}$, where

$$S_{hn_h}(\theta) = \frac{1}{n_h} \sum_{i=1}^{n_h} g_{hi}(\theta) g_{hi}^T(\theta), \quad \bar{g}_h(\theta) = \frac{1}{n_h} \sum_{i=1}^{n_h} g_{hi}(\theta).$$

By this and Taylor's expansion, we have (uniformly for u)

$$l_E(\theta) = \sum_h \sum_i t_h^T(\theta) g_{hi}(\theta) - \frac{1}{2} \sum_h \sum_i [t_h^T(\theta) g_{hi}(\theta)]^2 + o(n^{\frac{1}{3}}) \text{ (a.s.)}$$

$$\begin{aligned}
&= \sum_h n_h t_h^\tau(\theta) \bar{g}_h(\theta) - \sum_h \frac{n_h}{2} t_h^\tau(\theta) S_{h n_h}(\theta) t_h(\theta) + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\
&= \sum_h n_h [t_h^\tau(\theta) \bar{g}_h(\theta) - \frac{1}{2} t_h^\tau(\theta) S_{h n_h}(\theta) t_h(\theta)] + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\
&= \frac{1}{2} \sum_h n_h t_h^\tau(\theta)_{h n_h}(\theta) t_h(\theta) + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\
&= \frac{1}{2} \sum_h n_h \bar{g}_h^\tau(\theta) S_{h n_h}^{-1}(\theta) \bar{g}_h(\theta) + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\
&= \frac{1}{2} \sum_h n_h \left[\bar{g}_h(\theta_0) + \frac{1}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta} u n^{-\frac{1}{3}} \right]^\tau S_{h n_h}^{-1}(\theta) \\
&\quad \cdot \left[\bar{g}_h(\theta_0) + \frac{1}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta} u n^{-\frac{1}{3}} \right] + o(n^{\frac{1}{3}}) \text{ (a.s.)} \\
&= \frac{1}{2} \sum_h n_h \left[O(n_h^{-\frac{1}{2}} (\log \log n_h)^{\frac{1}{2}}) + E_h \frac{\partial g_h(x_h, \theta)}{\partial \theta} u n^{-\frac{1}{3}} \right]^\tau S_h^{-1}(\theta_0) \\
&\quad \cdot \left[O(n_h^{-\frac{1}{2}} (\log \log n_h)^{\frac{1}{2}}) + E_h \frac{\partial g_h(x_h, \theta)}{\partial \theta} u n^{-\frac{1}{3}} \right] + o(n^{\frac{1}{3}}) \text{ a.s.} \\
&\geq (c - \varepsilon) n^{\frac{1}{3}} \text{ (a.s.)}
\end{aligned}$$

where $c - \varepsilon > 0$ and c is the smallest eigenvalue of

$$\frac{1}{2} \sum_h k_h^{-1} \left[E_h \frac{\partial g_h(x_h, \theta)}{\partial \theta} \right]^\tau S_h^{-1}(\theta_0) \left[E_h \frac{\partial g_h(x_h, \theta)}{\partial \theta} \right].$$

Similarly

$$l_E(\theta_0) = \frac{1}{2} \sum_h n_h \bar{g}_h^\tau(\theta_0) S_{h n_h}^{-1}(\theta_0) \bar{g}_h(\theta_0) + o(n^{\frac{1}{3}}) \text{ a.s.} = o(n^{\frac{1}{3}}) \text{ a.s.}$$

Since $l_E(\theta)$ is a continuous function about θ when θ belongs to the ball $\|\theta - \theta_0\| \leq n^{-\frac{1}{3}}$, $l_E(\theta)$ has minimum value in the interior of this ball, and $\tilde{\theta}$ satisfies (noting (3.2))

$$\begin{aligned}
0 &= \frac{\partial l_E(\theta)}{\partial \theta^\tau} \Big|_{\theta=\tilde{\theta}} \\
&= \sum_h \sum_i \frac{1}{1 + t_h^\tau(\theta) g_{hi}(\theta)} \left[\frac{\partial t_h(\theta)}{\partial \theta} g_{hi}(\theta) + \frac{\partial g_{hi}(\theta)}{\partial \theta} t_h(\theta) \right] \Big|_{\theta=\tilde{\theta}}
\end{aligned}$$

$$= \sum_h \sum_i \frac{1}{1 + t_h^T(\theta) g_{hi}(\theta)} \frac{\partial g_{hi}(\theta)}{\partial \theta} t_h(\theta) \Big|_{\theta=\tilde{\theta}}$$

i.e., $Q_{H+1}(\tilde{\theta}, \tilde{t}_1, \dots, \tilde{t}_H) = 0$. $\square$

Proof of Theorem 1

Taking derivatives about θ^τ , t_h^τ , we have

$$\frac{\partial Q_h(\theta_0, 0)}{\partial \theta^\tau} = \frac{1}{n_h} \sum_i \frac{\partial g_{hi}(\theta_0)}{\partial \theta^\tau} = \frac{\partial \bar{g}_h(\theta_0)}{\partial \theta^\tau},$$

$$\frac{\partial Q_h(\theta_0, 0)}{\partial t_h^\tau} = -\frac{1}{n_h} \sum_i g_{hi}(\theta_0) g_{hi}^\tau(\theta_0) = -S_{h,n_h}(\theta_0).$$

$$\frac{\partial Q_{H+1}(\theta_0, 0, \dots, 0)}{\partial \theta^\tau} = 0, \quad \frac{\partial Q_{H+1}(\theta_0, 0, \dots, 0)}{\partial t_h^\tau} = \frac{\partial \bar{g}_h(\theta_0)}{\partial \theta}.$$

Expanding $Q_h(\tilde{\theta}, \tilde{t}_h)$ at $(\theta_0, 0)$ and $Q_{H+1}(\tilde{\theta}, \tilde{t}_1, \dots, \tilde{t}_H)$ at $(\theta_0, 0, \dots, 0)$, we have

$$\begin{aligned} 0 &= Q_h(\tilde{\theta}, \tilde{t}_h) \\ &= Q_h(\theta_0, 0) + \frac{\partial Q_h(\theta_0, 0)}{\partial \theta^\tau} (\tilde{\theta} - \theta_0) + \frac{\partial Q_h(\theta_0, 0)}{\partial t_h^\tau} \tilde{t}_h + o_p(\delta_n) \\ &= \bar{g}_h(\theta_0) + \frac{\partial \bar{g}_h(\theta_0)}{\partial \theta^\tau} (\tilde{\theta} - \theta_0) - S_{h,n_h}(\theta_0) \tilde{t}_h + o_p(\delta_n), \quad h = 1, \dots, H \end{aligned} \quad (6.1)$$

$$0 = Q_{H+1}(\tilde{\theta}, \tilde{t}_1, \dots, \tilde{t}_H) = \sum_h \frac{\partial \bar{g}_h(\theta_0)}{\partial \theta} \tilde{t}_h + o_p(\delta_n). \quad (6.2)$$

where $\delta_n = \|\tilde{\theta} - \theta_0\| + \sum_h \|\tilde{t}_h\|$. Denoting

$$M_{n,11} = \text{diag}\{S_{1,n_1}(\theta_0), \dots, S_{H,n_H}(\theta_0)\} = \text{diag}\{S_{1,n_1}, \dots, S_{H,n_H}\},$$

$$M_{11} = \text{diag}\{\sigma_1(\theta_0), \dots, \sigma_H(\theta_0)\} = \text{diag}\{\sigma_1, \dots, \sigma_H\},$$

$$\tilde{t} = (\tilde{t}_1^\tau, \dots, \tilde{t}_H^\tau)^\tau, \quad \bar{g}(\theta_0) = (\bar{g}_1^\tau(\theta_0), \dots, \bar{g}_H^\tau(\theta_0))^\tau = (\bar{g}_1^\tau, \dots, \bar{g}_H^\tau)^\tau, \quad \mathbf{1} = (1, \dots, 1)^\tau,$$

$$M_{n,12} = \left(-\frac{\partial \bar{g}_1(\theta_0)}{\partial \theta}, \dots, -\frac{\partial \bar{g}_H(\theta_0)}{\partial \theta} \right)^\tau, \quad M_{12} = \left(-E \frac{\partial g_1(\theta_0)}{\partial \theta}, \dots, -E \frac{\partial g_H(\theta_0)}{\partial \theta} \right)^\tau,$$

$$M_{n,21} = M_{n,12}^T, \quad M_{21} = M_{12}^T, \quad \begin{pmatrix} M_{n,11} & M_{n,12} \\ M_{n,21} & 0 \end{pmatrix} = M_n, \quad M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & 0 \end{pmatrix},$$

equations of (6.1) (6.2) can be written in matrix form as

$$\begin{pmatrix} \bar{g}(\theta_0) + o_p(\delta_n)\mathbf{1} \\ o_p(\delta_n)\mathbf{1} \end{pmatrix} = \begin{pmatrix} M_{n,11} & M_{n,12} \\ M_{n,21} & 0 \end{pmatrix} \begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix}.$$

Since $M_{n,11} \xrightarrow{p} M_{11}$, $M_{n,12} \xrightarrow{p} M_{12}$, $M_{n,21} \xrightarrow{p} M_{21}$, we get

$$\begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} = M^{-1} \begin{pmatrix} \bar{g}(\theta_0) + o_p(\delta_n)\mathbf{1} \\ o_p(\delta_n)\mathbf{1} \end{pmatrix}.$$

From this and $\bar{g}_h(\theta_0) = \frac{1}{n_h} \sum_i g_{hi}(\theta_0) = O_p(n_h^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}})$, we know that $\delta_n = O_p(n^{-\frac{1}{2}})$. Since

$$M^{-1} = \begin{pmatrix} I & -M_{11}^{-1}M_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & M_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -M_{21}M_{11}^{-1} & I \end{pmatrix},$$

where $M_{22.1} = -M_{21}M_{11}^{-1}M_{12}$, we have

$$\begin{aligned} \begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} &= M^{-1} \begin{pmatrix} \bar{g}(\theta_0) \\ 0 \end{pmatrix} + o_p(\delta_n) \\ &= \begin{pmatrix} U\bar{g}(\theta_0) \\ -M_{22.1}^{-1}M_{21}M_{11}^{-1}\bar{g}(\theta_0) \end{pmatrix} + o_p(\delta_n), \end{aligned} \quad (6.3)$$

where $U = M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1})$. Hence

$$\sqrt{n}(\tilde{\theta} - \theta_0) \rightarrow N(0, V),$$

where

$$V = M_{22.1}^{-1}M_{21}M_{11}^{-1} \text{diag}\{k_h\sigma_h\}M_{11}^{-1}M_{12}M_{22.1}^{-1}$$

$$\begin{aligned}
&= M_{22.1}^{-1} M_{21} M_{11}^{-1} \text{diag}\{k_h I_h\} M_{12} M_{22.1}^{-1} \\
&= M_{22.1}^{-1} \left[\sum_h k_h E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1}(\theta_0) E \frac{\partial g_h(\theta_0)}{\partial \theta^r} \right] M_{22.1}^{-1}.
\end{aligned}$$

Further

$$\tilde{t} \rightarrow N(0, Q),$$

where

$$\begin{aligned}
Q &= M_{11}^{-1} (I + M_{12} M_{22.1}^{-1} M_{21} M_{11}^{-1}) \text{diag}\{k_h \sigma_h\} M_{11}^{-1} (I + M_{12} M_{22.1}^{-1} M_{21} M_{11}^{-1}) \\
&= M_{11}^{-1} (I + M_{12} M_{22.1}^{-1} M_{21} M_{11}^{-1}) \text{diag}\{k_h I_{r_h}\} (I + M_{12} M_{22.1}^{-1} M_{21} M_{11}^{-1}).
\end{aligned}$$

Now we turn to the third conclusion. Since

$$\begin{aligned}
\tilde{p}_{Hi} &= \frac{1}{n_H} \frac{1}{1 + \tilde{t}_H^r g_{Hi}(\tilde{\theta})} \\
&= \frac{1}{n_H} (1 - \tilde{t}_H^r g_{Hi}(\tilde{\theta})) + o_p(n^{-\frac{3}{2}}),
\end{aligned}$$

we know

$$\begin{aligned}
&\tilde{F}_{H,n_H}(x) \\
&= \sum_i \tilde{p}_{Hi} 1(x_{Hi} < x) \\
&= \frac{1}{n_H} \sum_i 1(x_{Hi} < x) - \tilde{t}_H^r \frac{1}{n_H} \sum_i g_{Hi}(\tilde{\theta}) 1(x_{Hi} < x) + o_p(n^{-\frac{1}{2}}) \\
&= F_{H,n_H}(x) - \frac{1}{n_H} \sum_i g_{Hi}^r 1(x_{Hi} < x) \tilde{t}_H + o_p(n^{-\frac{1}{2}}) \\
&= F_{H,n_H}(x) - B \tilde{t}_H + o_p(n^{-\frac{1}{2}}),
\end{aligned}$$

where

$$\frac{1}{n_H} \sum_i g_{Hi}^r 1(x_{Hi} < x) \xrightarrow{p} E g_H^r(x_H, \theta_0) 1(x_H < x) \triangleq B.$$

By the definition of U , we have

$$\begin{aligned}
 U &= M_{11}^{-1}(I + M_{12}M_{22.1}^{-1}M_{21}M_{11}^{-1}) \\
 &= M_{11}^{-1} \left[I + \left(E \frac{\partial g_1(\theta_0)}{\partial \theta}, \dots, E \frac{\partial g_H(\theta_0)}{\partial \theta} \right)^{\tau} M_{22.1}^{-1} \left(E \frac{\partial g_1(\theta_0)}{\partial \theta}, \dots, E \frac{\partial g_H(\theta_0)}{\partial \theta} \right) M_{11}^{-1} \right] \\
 &= M_{11}^{-1} \begin{pmatrix} I + E \frac{\partial g_1(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_1(\theta_0)}{\partial \theta} \sigma_1^{-1} & \dots & I + E \frac{\partial g_1(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_H(\theta_0)}{\partial \theta} \sigma_H^{-1} \\ \dots & \dots & \dots \\ I + E \frac{\partial g_H(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_1(\theta_0)}{\partial \theta} \sigma_1^{-1} & \dots & I + E \frac{\partial g_H(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_H(\theta_0)}{\partial \theta} \sigma_H^{-1} \end{pmatrix},
 \end{aligned}$$

and

$$\tilde{t} = \begin{pmatrix} \tilde{t}_1 \\ \dots \\ \tilde{t}_H \end{pmatrix} = \begin{pmatrix} \sigma_1^{-1} \sum_h \left[I + E \frac{\partial g_1(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1} \right] \bar{g}_h(\theta_0) \\ \dots \\ \sigma_H^{-1} \sum_h \left[I + E \frac{\partial g_H(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1} \right] \bar{g}_h(\theta_0) \end{pmatrix}.$$

Hence

$$\begin{aligned}
 \tilde{F}_{H,n_H}(x) &= F_{H,n_H}(x) - B\sigma_H^{-1} \sum_h \left[I + E \frac{\partial g_H(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1} \right] \bar{g}_h(\theta_0) + o_p(n^{-\frac{1}{2}}) \\
 &= \frac{1}{n_H} \sum_i \left[1(x_{Hi} < x) - B\sigma_H^{-1} \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_H(\theta_0)}{\partial \theta} \sigma_H^{-1} \right) g_{Hi}(\theta_0) \right] \\
 &\quad - B\sigma_H^{-1} \sum_{h=1}^{H-1} \left[I + E \frac{\partial g_H(\theta_0)}{\partial \theta^{\tau}} M_{22.1}^{-1} E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1} \right] \bar{g}_h(\theta_0) + o_p(n^{-\frac{1}{2}}),
 \end{aligned}$$

and

$$\sqrt{n_H}(\tilde{F}_{H,n_H}(x) - F_H(x)) \rightarrow N(0, W),$$

where

$$W = F_H(x)[1 - F_H(x)]$$

$$\begin{aligned}
& -2B\sigma_H^{-1} \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^\tau} M_{22.1}^{-1} E \frac{\partial g_H(\theta_0)}{\partial \theta} \sigma_H^{-1} \right) B^\tau \\
& + B\sigma_H^{-1} \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^\tau} M_{22.1}^{-1} E \frac{\partial g_H(\theta_0)}{\partial \theta} \sigma_H^{-1} \right) \\
& \cdot \sigma_H \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^\tau} M_{22.1}^{-1} E \frac{\partial g_H(\theta_0)}{\partial \theta} \sigma_H^{-1} \right)^\tau \sigma_H^{-1} B^\tau \\
& + B\sigma_H^{-1} \sum_{h=1}^{H-1} \left(\frac{k_h}{k_H} \right) \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^\tau} M_{22.1}^{-1} E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1} \right) \\
& \cdot \sigma_h \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^\tau} M_{22.1}^{-1} E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1} \right)^\tau \sigma_h^{-1} B^\tau.
\end{aligned}$$

Noting that $UM_{11}U = U$, we have

$$\begin{aligned}
W &= F_H(x)[1 - F_H(x)] \\
& - B\sigma_H^{-1} \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^\tau} M_{22.1}^{-1} E \frac{\partial g_H(\theta_0)}{\partial \theta} \sigma_H^{-1} \right) B^\tau \\
& + B\sigma_H^{-1} \sum_{h=1}^{H-1} \left(\frac{k_h}{k_H} \right) \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^\tau} M_{22.1}^{-1} E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1} \right) \\
& \cdot \sigma_h \left(I + E \frac{\partial g_H(\theta_0)}{\partial \theta^\tau} M_{22.1}^{-1} E \frac{\partial g_h(\theta_0)}{\partial \theta} \sigma_h^{-1} \right)^\tau \sigma_h^{-1} B^\tau. \square
\end{aligned} \tag{6.4}$$

Proof of Theorem 2

Noting (6.3), we have

$$\begin{aligned}
l_E(\tilde{\theta}, \tilde{t}) &= \sum_h \sum_i \log[1 + \tilde{t}_h^\tau g_{hi}(\tilde{\theta})] \\
&= \sum_h \sum_i \tilde{t}_h^\tau g_{hi}(\tilde{\theta}) - \frac{1}{2} \sum_h \sum_i [\tilde{t}_h^\tau g_{hi}(\tilde{\theta})]^2 + o_p(1) \\
&= \sum_h n_h \tilde{t}_h^\tau \bar{g}_h(\tilde{\theta}) - \frac{1}{2} \sum_h n_h \tilde{t}_h^\tau S_{h,n_h} \tilde{t}_h + o_p(1) \\
&= \sum_h n_h \tilde{t}_h^\tau \bar{g}_h(\tilde{\theta}) - \frac{1}{2} \sum_h n_h \tilde{t}_h^\tau \sigma_h \tilde{t}_h + o_p(1) \\
&= \sum_h n_h \tilde{t}_h^\tau \left[\bar{g}_h(\theta_0) + \frac{\partial \bar{g}_h(\theta_0)}{\partial \theta^\tau} (\tilde{\theta} - \theta_0) \right] - \frac{1}{2} \sum_h n_h \tilde{t}_h^\tau \sigma_h \tilde{t}_h + o_p(1)
\end{aligned}$$

$$\begin{aligned}
&= \sum_h n_h \tilde{t}_h^\tau \bar{g}_h(\theta_0) - \frac{1}{2} \sum_h n_h \tilde{t}_h^\tau \sigma_h \tilde{t}_h + o_p(1) \\
&= \bar{g}^\tau(\theta_0) [UN - \frac{1}{2} UNM_{11}U] \bar{g}(\theta_0) + o_p(1),
\end{aligned}$$

where U , M_{11} are defined in section 4, $N = \text{diag}\{n_1 I_{r_1}, \dots, n_H I_{r_H}\}$. Similarly, we get

$$\begin{aligned}
l_E(\theta_0, t(\theta_0)) &= \sum_h \sum_i \log[1 + t_h^\tau(\theta_0) g_{hi}(\theta_0)] \\
&= \sum_h n_h \bar{g}_h(\theta_0) \sigma_h^{-1} \bar{g}_h(\theta_0) + o_p(1) \\
&= \frac{1}{2} \bar{g}^\tau(\theta_0) N M_{11}^{-1} \bar{g}(\theta_0) + o_p(1).
\end{aligned}$$

Hence,

$$R = \bar{g}^\tau(\theta_0) (N M_{11}^{-1} + UNM_{11}U - 2UN) \bar{g}(\theta_0) + o_p(1).$$

Letting

$$P = N M_{11}^{-1} + UNM_{11}U - 2UN.$$

we can see only when $n_1 = \dots = n_H = n$, then $UN = NU$, and P is symmetric; otherwise it is asymmetric, and the distribution of R is of an unknown form. Since we suppose that $n_1 = \dots = n_H$, therefore

$$N M_{11}^{-1} + UNM_{11}U - 2UN = n(M_{11}^{-1} + UM_{11}U - 2U) = n(M_{11}^{-1} - U)$$

By the definition of U , we have

$$\begin{aligned}
P &= M_{11}^{\frac{1}{2}} (M_{11}^{-1} - U) M_{11}^{\frac{1}{2}} \\
&= M_{11}^{\frac{1}{2}} (M_{11}^{-1} M_{12} M_{22.1}^{-1} M_{21} M_{11}^{-1}) M_{11}^{\frac{1}{2}}.
\end{aligned}$$

Hence P is idempotent, with trace equal to p , the theorem follows.

Proof of Corollary 2

From the proof of theorem 2, we have

$$\begin{aligned}
 R_1 &= 2l_E(\tilde{\theta}, \tilde{t}) \\
 &= 2 \sum_h \sum_i \log[1 + \tilde{t}_h^\tau(\tilde{\theta}) g_{hi}(\tilde{\theta})] \\
 &= \bar{g}^\tau(\theta_0)(2UN - UNM_{11}U)\bar{g}(\theta_0) + o_p(1) \\
 &= n \left(M_{11}^{-\frac{1}{2}} \bar{g}(\theta_0) \right)^\tau P_1 \left(M_{11}^{-\frac{1}{2}} \bar{g}(\theta_0) \right) + o_p(1) \\
 &\xrightarrow{p} \chi_{r-p}^2,
 \end{aligned}$$

where $P_1 = M_{11}^{\frac{1}{2}} U M_{11}^{\frac{1}{2}}$.

Chapter 7

Conclusions and Further Research

7.1 Conclusions

An empirical likelihood approach to the use of auxiliary information in stratified survey is introduced. We first considered how to make inference on the population mean $\bar{Y}$ and the distribution function $F_Y(y)$ of characteristic y of a population when the population mean of an auxiliary variable x is known. Then we extended the form of auxiliary information to more general format, namely, the general estimating equations. In the above two cases, only stratified simple random sampling (with or without replacement) is considered. We proceeded to consider more complex sampling designs, i.e., in each stratum, other sampling methods such as probability proportional to size sampling are used. Chapter 5 is trying to solve the calculation problem which may rise from Chapter 3, that is, when we solve the estimating equations involving unknown parameters, sometime the solution $\hat{p}_{hi}$ may be negative. Chapter 6 is an immediate application of empirical likelihood, namely, using

empirical likelihood method in the presence of measurement error.

The empirical likelihood estimators of parameters such as $\bar{Y}$, the mean of the interested characteristic y of population, and the distribution function of y , are shown in Chapter 2 to be asymptotically equivalent to optimal estimators given by Rao and Liu(1992). However, the EL estimator of $F_Y(y)$, is guaranteed to be monotone, non-negative, unlike the optimal estimator. Empirical likelihood ratio test is also considered, though our simulations show that its finite sample properties are not as good as expected.

The above results are generalized to deal with the case of auxiliary information summarised in the form of general estimating equations, hence enlarging the application areas of the empirical likelihood method. Methods to deal with cases where parameters are subject to constraints are also included. Empirical likelihood ratio test is also discussed. The results show that empirical likelihood ratio test, Lagrange multiplier test, and Wald test lead to the same asymptotic distribution.

In practice, a complex survey design is often employed. How to incorporate such designs into empirical likelihood inference is presented in Chapter 4. Here the empirical likelihood had been changed to pseudo-likelihood. All the inferences are based on the pseudo-likelihood. The EL ratio test is no longer asymptotically χ^2 -distributed. It is a weighted sum of independent χ_1^2 variable.

The method of combining EL, general estimating equations and M-estimation is presented in Chapter 5. The key point here is that the general estimation equations and EL are used to estimate distribution function first, where the auxiliary information summarised through estimating equations does not include any unknown parameters, then use this estimated distribution function in M-estimation. In some

complex case, this will reduce the difficulty in calculation, but the auxiliary information should be summarised without unknown parameters.

In the presence of measurement error, EL can be used to combine information from different instruments to make more accurate inference. In other words, we can borrow strength from other populations to make inference on the population of interest. Some form of relation among these populations must be known before we can borrow strength from other populations.

7.2 Further Research

While it is hoped that the results of this research will be of use in making more accurate inference when auxiliary information is available, many areas remain in which further research might be profitably undertaken. Three areas are now discussed.

The simulations presented in Chapter 2 show that the empirical likelihood (ratio) estimator for parameters of interest are very good. However, the empirical likelihood ratio tests are not as good as expected. Higher order correction to the tests may improve the performance of EL ratio testing.

The EL estimators for the distribution function from Chapters 2 to 6 are nonnegative, monotone, hence can be used to estimate quantiles. Large sample properties of these estimators need to be studied.

Throughout the paper, the weights of all strata are assumed known. This will limit the application of EL method to more complex sampling such as multi-stage sampling. Typically, in multiple-stage sampling, the weights of second stage sampling are not known. How to combine the auxiliary information in estimating weights

and the parameter of interest will be more challenging.

•

References

- Aitchison, J., & Silvey, S.D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.*, 29, 813-828.
- Bickel, P.J. & Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* 12, 470-82.
- Chaudhary M.A. & Sen, P.K. (1995). Asymptotic Distribution of Estimators from Unequal Probability Sampling, *Proceedings of the American Statistical Association*, Section on Survey Research Methods, 345-349.
- Chen, J. & Qin, J. (1993). Empirical likelihood estimation for finite population and the effective usage of auxiliary information. *Biometrika* 80 107-116.
- Chen, J. & Sitter, R.R. A Pseudo Empirical Likelihood Approach to the Effective Use of Auxiliary Information in Complex Surveys. *Research Report*, No. 96-01.
- Cochran, W.G. (1977). *Sampling Techniques*, 3rd. New York: Wiley.
- DiCiccio, T.J., P. Hall & J.P. Romano (1989). Comparison of parametric and empirical likelihood functions. *Biometrika*, 76, 465-476.
- DiCiccio, T.J. & Romano, J. (1990). Nonparametric confidence limits by resampling methods and least favorable families. *Internat. Statist. Rev.* , 58, 59-76
- Fuller, W.A. (1995). Estimation in the presence of measurement error. *Int. Statist. Rev.*, 63, 121-141.
- Ghosh M. & Rao, J.N.K. (1994). Small Area Estimation: An Appraisal, *Statistical*

- Science*, 9, 55-93.
- Hall, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.*, 18, 121-140
- Hartley, H.O. & Rao, J.N.K. (1968). A new estimation theory for sample surveys, *Biometrika*, 55, 547-557.
- Hochberg, Y. (1977), On the Use of Double Sampling Schemes in Analyzing Categorical Data with Misclassification Errors, *JASA*, 72, 914-921.
- Holl, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.*, 18, 121-140.
- Huber, P. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, 35, 73-101
- Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimates, *Ann. Math. Statist.*, 40, 633-643.
- Luo, M. & Stokes, L. & Sager, T. Estimation of the CDF of a Finite Population in the Presence of a Calibration Sample.
- Ohlsson, E. (1986). Asymptotic Normality of the Rao-Hartley-Cochran Estimator: An Application of the Martingale CLT, *Scand J. Statist.*, 13, 17-28.
- Overton, W. S. (1989). Effects of measurement and other extraneous errors on estimated distribution functions in the national surface water surveys. *Technical report 129, Dept. of Statistics, Oregon State University.*
- 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.
- Qin, J. & Lawless, J.F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* 22 300-325
- Qin, J. & Lawless, J.F. (1995). Estimating equations, empirical likelihood and constraints on parameters, *The Canadian Journal of Statistics*, 23, 145-159.
- Rao, J.N.K. (1994). Estimating Totals and Distribution Functions Using Auxiliary Information at the Estimation Stage, *Journal of Official Statistics*, 10, 153-165
- Rao, J.N.K. & Kovar, J.G. & Mantel, H.J. (1990). On estimating Distribution Functions and Quantiles from Survey Data using Auxiliary Information, *Biometrika*. 77, 365-375
- Rao, J.N.K. & Liu, J. (1992). On estimating distribution function from sample survey data using supplementary information at the estimation stage. *Nonparametric Statistics and Related Topics (A.K.Md.E. Saleh, Ed.)*, New York: Elswior. 399-407
- Rao, J.N.K. & Scott, A.J. (1979) Chi-Squared Tests for Analysis of Categorical Data From Complex Surveys, *Proceedings of the American Statistical Association*, Section on Survey Research Methods, 58-66.
- Rao, J.N.K. & Scott, A.J. (1981). The analysis of Categorical Data From Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables, *JASA*, 76, 221-230.
- Rosen, B. (1972), Asymptotic Theory for Successive Sampling with Varying Probabilities, I and II, *Annals of Mathematical Statistics*, 43, 373-397; 748-776.
- Särndal, C.E., Swensson, B. & Wretman, J. (1991). Model assisted survey sampling,

Springer.

- Sen, P.K. (1988), Asymptotics in Finite Population Sampling, *Hand Book of Statistics 6*, P.R. Krishnaiah and C.R. Rao Eds., Elsevier Science Publishers B.V, 291-331.
- Shao, J. (1994). L-statistics in complex survey problems, *Ann. Statist.*, 22, 946-967.
- Tenenbein, A. (1970), A Double sampling Scheme for Estimating from Binomial Data with Misclassifications, *JASA*, 65, 1350-1361.
- Tenenbein, A. (1971), A Double sampling Scheme for Estimating from Binomial Data with Misclassifications: Sample Size Determination, *Biometrics*, 27, 935-944.
- Tenenbein, A. (1972), A Double sampling Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection, *Technometrics*, 14, 187-202.
- Zhang, B. (1995). M-estimation and Quantile Estimation in the Presence of Auxiliary Information, *Journal of Statistical Planning and Inference*, 44, 77-94.
- Zhang, B. (1996), Estimating a Population Variance with Known Mean, *International Statistical Review*, 64, 215-229.
- Zhong, C. & Rao, JNK (1996). Empirical likelihood inference using stratified sampling using auxiliary information, *The proceeding of JSM 96*.

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

© 1993, Applied Image, Inc.. All Rights Reserved