Empirical Likelihood Methods for Pretest-Posttest Studies

by

Min Chen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2015

© Min Chen 2015

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Pretest-posttest trials are an important and popular method to assess treatment effects in many scientific fields. In a pretest-posttest study, subjects are randomized into two groups: treatment and control. Before the randomization, the pretest responses and other baseline covariates are recorded. After the randomization and a period of study time, the posttest responses are recorded. Existing methods for analyzing the treatment effect in pretest-posttest designs include the two-sample t-test using only the posttest responses, the paired t-test using the difference of the posttest and the pretest responses, and the analysis of covariance method which assumes a linear model between the posttest and the pretest responses. These methods are summarized and compared by Yang and Tsiatis (2001) under a general semiparametric model which only assumes that the first and second moments of the baseline and the follow-up response variable exist and are finite. Leon et al. (2003) considered a semiparametric model based on counterfactuals, and applied the theory of missing data and causal inference to develop a class of consistent estimator on the treatment effect and identified the most efficient one in the class. Huang et al. (2008) proposed a semiparametric estimation procedure based on empirical likelihood (EL) which incorporates the pretest responses as well as baseline covariates to improve the efficiency.

The EL approach proposed by Huang et al. (2008) (the HQF method), however, dealt with the mean responses of the control group and the treatment group separately, and the confidence intervals were constructed through a bootstrap procedure on the conventional normalized Z-statistic. In this thesis, we first explore alternative EL formulations that directly involve the parameter of interest, i.e., the difference of the mean responses between the treatment group and the control group, using an approach similar to Wu and Yan (2012). Pretest responses and other baseline covariates are incorporated to impute the potential posttest responses. We consider the regression imputation as well as the non-parametric kernel imputation. We develop asymptotic distributions of the empirical likelihood ratio statistic that are shown to be scaled chi-squares. The results are used to construct confidence intervals and to conduct statistical hypothesis tests. We also derive

the explicit asymptotic variance formula of the HQF estimator, and compare it to the asymptotic variance of the estimator based on our proposed method under several scenarios. We find that the estimator based on our proposed method is more efficient than the HQF estimator under a linear model without an intercept that links the posttest responses and the pretest responses. When there is an intercept, our proposed model is as efficient as the HQF method. When there is misspecification of the working models, our proposed method based on kernel imputation is most efficient.

While the treatment effect is of primary interest for the analysis of pretest-posttest sample data, testing the difference of the two distribution functions for the treatment and the control groups is also an important problem. For two independent samples, the nonparametric Mann-Whitney test has been a standard tool for testing the difference of two distribution functions. Owen (2001) presented an EL formulation of the Mann-Whitney test but the computational procedures are heavy due to the use of a U-statistic in the constraints. We develop empirical likelihood based methods for the Mann-Whitney test to incorporate the two unique features of pretest-posttest studies: (i) the availability of baseline information for both groups; and (ii) the missing by design structure of the data. Our proposed methods combine the standard Mann-Whitney test with the empirical likelihood method of Huang, Qin and Follmann (2008), the imputation-based empirical likelihood method of Chen, Wu and Thompson (2014a), and the jackknife empirical likelihood (JEL) method of Jing, Yuan and Zhou (2009). The JEL method provides a major relief on computational burdens with the constrained maximization problems. We also develop bootstrap calibration methods for the proposed EL-based Mann-Whitney test when the corresponding EL ratio statistic does not have a standard asymptotic chi-square distribution. We conduct simulation studies to compare the finite sample performances of the proposed methods. Our results show that the Mann-Whitney test based on the Huang, Qin and Follmann estimators and the test based on the two-sample JEL method perform very well. In addition, incorporating the baseline information for the test makes the test more powerful.

Finally, we consider the EL method for the pretest-posttest studies when the design and data collection involve complex surveys. We consider both stratification and inverse probability weighting via propensity scores to balance the distributions of the baseline covariates between two treatment groups. We use a pseudo empirical likelihood approach to make inference of the treatment effect. The proposed methods are illustrated through an application using data from the International Tobacco Control (ITC) Policy Evaluation Project Four Country (4C) Survey.

Acknowledgements

First and foremost, I would like to express my deep and sincere gratitude to my PhD supervisors, Dr. Mary E. Thompson and Dr. Changbao Wu, for their patience, kindness, enthusiasm, and immense knowledge. Without their guidance, continuous support and encouragement throughout my PhD study, it would not be possible for me to finish this thesis.

I would also like to thank the members of my thesis committee, Dr. Pengfei Li, Dr. Yingli Qin, Dr. Min Tsao (University of Victoria), and Dr. Suzanne Tyas (School of Public Health and Health Systems, University of Waterloo), for their insightful comments and inspiring questions.

Additionally I wish to thank the faculty members and staff of the Department of Statistics and Actuarial Science at the University of Waterloo. I am grateful to Dr. Christian Boudreau for his kind help on SAS programming. I thank Mary Lou Dufton for her willingness to help and her excellent administrative support.

I would like to acknowledge the International Tobacco Control Policy Evaluation Project 4 Country Survey team for permitting the access to the data, which is an important statistical application in my thesis. I particularly thank Grace Li for addressing my many questions on the dataset.

I would like to thank my friends and fellow graduate students at Waterloo who gave me continued encouragement and made my life as a PhD student more joyful.

Last but not least, I want to give my very special thanks to Ying Yan for his love, understanding, and endless support.

Dedication

To my dearest parents who always believe in me, love me and support me unconditionally.

Contents

List of Tables **xviii**

List of Figures **xix**

1 Introduction **1**

1.1 Overview 1

1.2 Empirical Likelihood 4

1.3 Empirical Likelihood for Two-Sample Problems 6

1.4 Outline of the Thesis 9

2 An Imputation Based Empirical Likelihood Approach to Pretest-Posttest Studies **11**

2.1 Introduction 11

2.2 Notations and the HQF Estimator 12

2.3 Linear Regression Imputation-Based Empirical Likelihood Approach 14

2.4 Kernel Regression Imputation-Based Empirical Likelihood Approach 18

2.5 Efficiency Comparisons Among Alternative EL Approaches 20
2.6 Simulation Study 23
2.7 A Real Data Analysis 36
2.8 Concluding Remarks 39
2.9 Proofs and Regularity Conditions 40
2.9.1 Lemmas 40
2.9.2 Proof of Theorem 1 43
2.9.3 Regularity Conditions for Theorem 2 47
2.9.4 Proof of Theorem 2 48

3 Mann-Whitney Test with Empirical Likelihood Methods for Pretest-Posttest Studies 51

3.1 Introduction 51
3.2 Methods for Testing the Difference of Distributions of Two Independent Samples 53
3.2.1 Standard Mann-Whitney Test 53
3.2.2 Two-Sample Empirical Likelihood and Mann-Whitney Test 55
3.2.3 Jackknife Empirical Likelihood for Two-Sample U-Statistics 58
3.2.4 Notations under the Setting of Pretest-Posttest Studies 61
3.3 Adjusted Mann-Whitney Test Based on the HQF Estimators 62
3.4 Empirical Likelihood Based Mann-Whitney Test with Imputation 69
3.5 Two-sample Jackknife EL Method for Mann-Whitney Test 74

3.6 Simulation Studies 78

3.7 Concluding Remarks 82

4 Empirical Likelihood Method for Pretest-Posttest Studies under Complex Survey Design 85

4.1 Introduction 85

4.2 Estimator Based on Propensity Score Stratification and Its Variance Estimation 89

4.2.1 The Propensity Score and Stratification 89

4.2.2 Estimator of θ 91

4.2.3 Variance Estimations 93

4.3 Propensity Score Weighting and Two-Sample Pseudo EL Method 100

4.4 Application to the ITC 4 Country Survey Data 108

4.4.1 Variables and Data Management 108

4.4.2 Propensity Score Model and Balance Diagnosis 110

4.4.3 Data Analysis 111

4.5 Two-Sample Pseudo EL Method with Imputation for Survey Data 116

4.5.1 Notations 116

4.5.2 Two-Sample Pseudo EL Method 117

4.6 Concluding Remarks 120

5 Summary and Future Research 121

5.1 Summary and Future Research Topics 121

5.2 Discussion on Recommendations 123

References **125**

List of Tables

2.1 Inferences on θ under Model (I), $\rho = 0.8$, $\theta_0 = 0.3$ 28

2.2 Inferences on θ under Model (I), $\rho = 0.5$, $\theta_0 = 0.3$ 29

2.3 Inferences on θ under Model (I), $\rho = 0.3$, $\theta_0 = 0.3$ 30

2.4 Inferences on θ under Model (II), $\rho = 0.8$, $\theta_0 = 0.3$ 31

2.5 Inferences on θ under Model (II), $\rho = 0.5$, $\theta_0 = 0.3$ 32

2.6 Inferences on θ under Model (II), $\rho = 0.3$, $\theta_0 = 0.3$ 33

2.7 Comparisons with Model Misspecifications: (I) and (I*) 34

2.8 Comparisons with Model Misspecifications: (II) and (II*) 35

2.9 Treatment effect estimates for 20 ± 5 weeks post randomization CD4 counts for ACTG 175 39

3.1 Scenario (A): Empirical Power of Testing $H_0 : F_1 = F_0$ 82

3.2 Scenario (B): Empirical Power of Testing $H_0 : F_1 = F_0$ 83

3.3 Scenario (C): Empirical Power of Testing $H_0 : F_1 = F_0$ 83

4.1 Email invited group, gDI503, before propensity score stratification 113

4.2 Email invited group, gDI503, after propensity score stratification 113

4.3 Email invited group, gFR245v, before propensity score stratification 113

4.4 Email invited group, gFR245v, after propensity score stratification 114
4.5 Data analyses separately for the email invitation group and the no email invitation group 114
4.6 Pseudo two-sample EL method (the email invitation group and the no email invitation group combined) 114

List of Figures

2.1 Power Function of Testing $H_0 : \theta = 0$ under Model (I), $\rho = 0.8$, $(n_1, n_0) = (50, 50)$ 36

2.2 Power Function of Testing $H_0 : \theta = 0$ under Model (II), $\rho = 0.8$, $(n_1, n_0) = (50, 50)$ 37

2.3 Power Function of Testing $H_0 : \theta = 0$ under Model (I*), $(n_1, n_0) = (200, 200)$ 37

2.4 Power Function of Testing $H_0 : \theta = 0$ under Model (II*), $(n_1, n_0) = (200, 200)$ 38

4.1 Email invited group: gFR245v 115

Chapter 1

Introduction

1.1 Overview

Pretest-posttest studies are an important and popular method for assessing treatment effects or the effectiveness of an intervention in many scientific fields, such as medicine, public health and social sciences. In one type of pretest-posttest studies, a random sample of subjects is selected from the target population, and certain baseline (pretest) information is collected for all subjects in the sample. The subjects are then randomly assigned to either the treatment group or the control group. The responses of interest are recorded after a prespecified follow-up time period (posttest) for both groups. The treatment effect is assessed by the difference of the (mean) responses between the two groups. For more traditional pretest-posttest study designs, the responses are measured to all units in the sample at two different time points, one before the treatment (pretest) and the other after the treatment and a prespecified follow-up time (posttest).

There exist several methods in the literature to evaluate the treatment effects in pretest-posttest studies. These methods include (i) the two-sample t-test which directly compares the posttest measurements of two groups ignoring information from the pretest measurements; (ii) the paired t-test comparing the change between the pretest and posttest measures of responses; (iii) the analysis of covariance procedures which impose a linear model on the posttest measures with the treatment indicator and the pretest responses only (ANCOVA I) or the treatment indicator and the pretest measures and their interaction (ANCOVA II) as covariates. Many researchers, such as Brogan and Kutner (1980), Laird (1983), Crager (1987), Stanek (1988), and Follmann (1991), among others, have discussed these approaches under different scenarios but often with specific model assumptions such as normality or equality of variance of pretest and posttest responses.

Yang and Tsiatis (2001) examined some of the above methods under general conditions. They assumed only that the first and second moments of pretest and posttest responses are finite; the conditional joint distribution of the pretest and posttest responses conditioning on treatment can be arbitrary. They compared the large sample properties of the treatment effect estimators based on these methods. They also proposed a generalized estimating equation (GEE) method which considers the pretest and posttest measures as a multivariate response, and assumes arbitrary mean and covariance matrix. They showed that all these methods yield consistent and asymptotically normal estimators, and the GEE estimator and the ANCOVA II estimator are asymptotically equivalent and most efficient. In Leon et al. (2003), the authors took a semiparametric perspective without any distributional assumptions, and exploited theory of missing data and causal inference to develop a class of consistent treatment effect estimators and identify the most efficient one in the class. Davidian et al. (2005) later considered the situation when there is missing data in the posttest response.

Huang et al. (2008) proposed a semi-parametric procedure based on the empirical likelihood (EL) method to estimate the treatment effect in a pretest-posttest study. Their proposed strategy is to use the baseline information to form constraints when maximizing the EL function but estimation of the mean of the posttest response is handled separately for the treatment group and the control group. The treatment effect is then estimated by taking the difference between the two estimated means. They considered scenarios where posttest responses are subject to missingness, and compared the EL based estimators to the ones in Leon et al. (2003) and to those in Davidian et al. (2005). They found that the EL based estimators achieve the semi-parametric efficiency lower bound under a correctly specified working model which links the posttest response to the pretest response and other baseline covariates; the EL based estimators are more efficient in the semi-parametric sense when a misspecified working model is used to link the posttest response to the pretest response and other baseline covariates.

Although the EL approach proposed in Huang, Qin and Follmann (2008, hereafter referred to as HQF) looks appealing, it seems less natural to estimate the posttest response means for each group separately while the target parameter is actually the difference (i.e., treatment effect). In addition, empirical likelihood ratio confidence intervals or tests for the treatment effect cannot be constructed under the HQF approach. This motivates our proposed EL method for estimating the treatment effect in pretest-posttest studies. There have been considerable research efforts towards the problem of making inference of the treatment effect; however, testing the difference of distributions is rarely studied under the setting of pretest-posttest studies. In this thesis, we also propose the empirical likelihood based methods to test the difference of the distributions of the posttest responses from the treatment group and the control group. Furthermore, we extend our research to the complex survey context. We develop methods for estimating the treatment effect of pretest-posttest studies with observational survey data. Before we present our work, we provide a brief review of the empirical likelihood method in the remainder of this chapter.

The rest of this chapter is organized as follows. In Section 1.2, we briefly review the empirical likelihood (EL) method. In Section 1.3, we summarize the two-sample EL method proposed by Wu and Yan (2012). The outline of the thesis is given in Section 1.4.

1.2 Empirical Likelihood

The method of empirical likelihood (EL) was introduced by Owen (1988), Owen (1990), Owen (2001) for constructing confidence intervals (regions) in nonparametric settings for the mean or other functions of the distribution function. It has become one of the most popular methods in statistical inference over the last 20 years and has applications to many research areas. The empirical likelihood method has many advantages such as data-determined shapes for confidence intervals (regions), ease of incorporation of known constraints on parameters, Bartlett correctability, and a natural method of combining data from multiple sources. The standard empirical likelihood method for the mean can be demonstrated through the following simple example.

Let $\{Y_1, \cdots, Y_n\}$ be independent and identically distributed real valued random variables having a common cumulative distribution function $F(y)$ with mean μ. Let $\{y_1, \cdots, y_n\}$ be a realization of $\{Y_1, \cdots, Y_n\}$. The empirical cumulative distribution function of $Y_1, ..., Y_n$ is defined as

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

It has been shown that F_n uniquely maximizes the nonparametric likelihood $L = \prod_{i=1}^{n} p_i$, where $p_i = F(Y_i) - F(Y_i-)$, subject to the constraints $\sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0$. Moreover, confidence intervals for $\mu = E(Y)$ can be obtained in the following procedure. For any fixed μ, suppose $\hat{\boldsymbol{p}}(\mu) = (\hat{p}_1(\mu), \cdots, \hat{p}_n(\mu))$ maximizes $L = \prod_{i=1}^{n} p_i$ subject to constraints

$$\sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad \sum_{i=1}^{n} p_i y_i = \mu.$$

Using the Lagrange multiplier method, $\hat{p}_i(\mu)$ is given by:

$$\hat{p}_i(\mu) = \frac{1}{n}\frac{1}{1 + \lambda(y_i - \mu)},$$

where the Lagrange multiplier λ is determined by

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

The profile empirical log-likelihood of μ is given by

$$l_p(\mu) = -\sum_{i=1}^{n} \log\{1 + \lambda(y_i - \mu)\} - n \log n.$$

The maximum of $l_p(\mu)$ is attained when $\mu = \hat{\mu} = n^{-1} \sum_{i=1}^{n} y_i$. The profile empirical likelihood ratio function for μ is defined as

$$\begin{aligned} \mathcal{R}(\mu) &= \frac{\max\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i y_i = \mu, p_i \geq 0\}}{\max\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, p_i \geq 0\}} \\ &= \max\{\prod_{i=1}^{n} n p_i : \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \sum_{i=1}^{n} p_i y_i = \mu\}. \end{aligned}$$

Thus,

$$\log \mathcal{R}(\mu) = l_p(\mu) - l_p(\hat{\mu}) = -\sum_{i=1}^{n} \log\{1 + \lambda(y_i - \mu)\}.$$

Owen (2001) proved that $-2 \log \mathcal{R}(\mu)$ has asymptotically a χ^2 distribution with one degree of freedom. This is an important result which is analogous to that for the likelihood ratio statistic under a parametric model, and can be used to test statistical hypotheses and construct confidence intervals for μ.

Since Owen's pioneer work on empirical likelihood, many other researchers have extended and applied EL to various kinds of statistical problems. Qin and Lawless (1994) linked empirical likelihood to estimating equations, especially when the number of unbiased estimating equations may be greater than the number of parameters. They demonstrated that the EL method can effectively combine unbiased estimating equations and lead to the most efficient estimator. In the context of survey sampling, the EL method has been applied to incorporate auxiliary covariate information to improve efficiency, for example, by Chen and Qin (1993), Wu and Sitter (2001) and Wu and Rao (2006). Recently, the empirical likelihood method has become popular in addressing general missing data problems. Some researchers, such as Wang and Rao (2002) and Liang et al. (2007), first imputed the missing data using a kernel regression function of the observed data and then applied an EL method to do the statistical inference. For parameters estimation in estimating equations, Wang and Chen (2009) proposed an EL method with non-parametric imputation of missing data. Qin et al. (2009) explored the use of empirical likelihood to effectively combine unbiased estimating equations by separating the complete data unbiased estimating equations from the incomplete data unbiased estimating equations, and their proposed estimators achieve semi-parametric efficiency lower bound when correctly specifying the missing mechanism. Moreover, attention has also been focused on applying the EL to two-sample problems. Jing (1995) showed that the two-sample empirical likelihood for the difference of two population means is Bartlett correctable. Qin and Zhang (1997) and Qin (1998) considered a calibration-type empirical likelihood method in the context of the estimation of a response mean in case-control studies. Chen et al. (2003) used a two-sample EL method to combine the complete and incomplete observations under missingness completely at random. Cao and van Keilegom (2009) used an EL-based test to examine whether two populations follow the same distribution. Wu and Yan (2012) developed the weighted EL method with great advantage in computation, a pseudo EL method for comparing two population means when the two samples are selected by complex surveys, a two-sample EL method with missing responses, and bootstrap calibration procedures for the proposed EL methods.

1.3 Empirical Likelihood for Two-Sample Problems

In this section, we review some of the main theories in Wu and Yan (2012). Suppose there are two independent and identically distributed samples $\{Y_{11}, \cdots, Y_{1n_1}\}$, and $\{Y_{21}, \cdots, Y_{2n_2}\}$ from Y_1 and Y_2 respectively, with $E(Y_1) = \mu_1$, $Var(Y_1) = \sigma_1^2$, and $E(Y_2) = \mu_2$, $Var(Y_2) = \sigma_2^2$. Let $n = n_1 + n_2$. The parameter of interest is $\theta = \mu_1 - \mu_2$. Wu and Yan (2012) derived the asymptotic distribution of the standard two-sample empirical log-likelihood ratio statistic on θ. They also proposed the weighted two-sample empirical log-likelihood formulation and proved that the weighted two-sample empirical log-likelihood ratio statistic converges to a scaled χ_1^2.

The standard two-sample empirical likelihood function is given by

$$\ell(\boldsymbol{p}_1, \boldsymbol{p}_2) = \sum_{j=1}^{n_1} \log(p_{1j}) + \sum_{j=1}^{n_2} \log(p_{2j}),$$

where $\boldsymbol{p}_1 = (p_{11}, \cdots, p_{1n_1})$ and $\boldsymbol{p}_2 = (p_{21}, \cdots, p_{2n_2})$ are the two sets of probability measure imposed respectively over the two samples. For fixed θ, suppose $\hat{\boldsymbol{p}}_1(\theta) = (\hat{p}_{11}(\theta), \cdots, \hat{p}_{1n_1}(\theta))$ and $\hat{\boldsymbol{p}}_2(\theta) = (\hat{p}_{21}(\theta), \cdots, \hat{p}_{2n_2}(\theta))$ maximize $\ell(\boldsymbol{p}_1, \boldsymbol{p}_2)$ subject to the following constraints:

$$\sum_{j=1}^{n_1} p_{1j} = 1\,, \sum_{j=1}^{n_2} p_{2j} = 1\,, \tag{1.1}$$

$$\sum_{j=1}^{n_1} p_{1j} Y_{1j} - \sum_{j=1}^{n_2} p_{2j} Y_{2j} = \theta\,. \tag{1.2}$$

The standard two-sample empirical log-likelihood ratio statistic on θ is defined as

$$r(\theta) = \sum_{j=1}^{n_1} \log(n_1 \hat{p}_{1j}(\theta)) + \sum_{j=1}^{n_2} \log(n_2 \hat{p}_{2j}(\theta))\,.$$

In Wu and Yan (2012), the authors showed that

$$-2r(\theta) \xrightarrow{d} \chi_1^2 \quad \text{as} \quad n \to \infty\,, \tag{1.3}$$

where "$\xrightarrow{d}$" denotes convergence in distribution. In the proof of this result, they introduced a nuisance parameter $\mu_0 = \mu_2 + O_p(n^{-1/2})$ to facilitate the arguments and rewrote the constraint (1.2) as

$$\sum_{j=1}^{n_1} p_{1j} Y_{1j} = \mu_0 + \theta \quad \text{and} \quad \sum_{j=1}^{n_0} p_{2j} Y_{2j} = \mu_0\,.$$

The nuisance parameter μ_0 serves as a bridge for computing the EL ratio statistic for θ and will eventually be profiled. By (1.3), the $(1-\alpha)$-level confidence interval on θ can be constructed as $\mathcal{C}_1 = \{\theta | -2r(\theta) \le \chi_1^2(\alpha)\}$, where $\chi_1^2(\alpha)$ is the upper $(100\alpha)\%$ quantile from the χ_1^2 distribution. The major computational difficulty comes from solving for the Lagrange multiplier, which needs to be calculated based on two samples with an added nuisance parameter μ_0. Such difficulty can be avoided through the weighted empirical likelihood formulation, for which the computation procedures are much simpler and essentially identical to those for one-sample EL problems.

The weighted empirical log-likelihood function is defined as follows:

$$\ell_w(\boldsymbol{p}_1, \boldsymbol{p}_2) = \frac{w_1}{n_1}\sum_{j=1}^{n_1}\log(p_{1j}) + \frac{w_2}{n_2}\sum_{j=1}^{n_2}\log(p_{2j}),$$

where $w_1 = w_2 = 1/2$. The choice of w_1 and w_2 is to facilitate the reformulation of constraints (1.1) and (1.2) into the following equivalent forms:

$$\sum_{i=1}^{2} w_i \sum_{j=1}^{n_i} p_{ij} = 1\,, \tag{1.4}$$

$$\sum_{i=1}^{2} w_i \sum_{j=1}^{n_i} p_{ij}\boldsymbol{u}_{ij} = \mathbf{0}\,. \tag{1.5}$$

where $\boldsymbol{u}_{ij} = \boldsymbol{Z}_{ij} - \boldsymbol{\eta}$, $\boldsymbol{Z}_{1j} = (1, Y_{1j}/w_1)^T$, $\boldsymbol{Z}_{2j} = (0, -Y_{2j}/w_2)^T$, and $\boldsymbol{\eta} = (w_1, \theta)^T$. Suppose $\hat{p}_{w1j}$ and $\hat{p}_{w2j}$ maximize $\ell_w(\boldsymbol{p}_1, \boldsymbol{p}_2)$ subject to constraints (1.4) and (1.5). Using the standard Lagrange multiplier method, it can be shown:

$$\hat{p}_{wij} = 1/\{n_i(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{ij})\}, \quad i = 1, 2 \text{ and } j = 1, \cdots, n_i\,,$$

and the Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to

$$g(\boldsymbol{\lambda}) = \sum_{i=1}^{2}\frac{w_i}{n_i}\sum_{j=1}^{n_i}\frac{\boldsymbol{u}_{ij}}{1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{ij}} = \mathbf{0}\,. \tag{1.6}$$

The weighted two-sample empirical log-likelihood ratio statistic for θ is defined as:

$$r_w(\theta) = -\sum_{i=1}^{2}\frac{w_i}{n_i}\sum_{j=1}^{n_i}\log(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{ij})\,.$$

Wu and Yan (2012) proved that

$$-2r_w(\theta)/c \xrightarrow{d} \chi^2_1 \ \text{ as } \ n \to \infty\,,$$

where c is a scaling constant. Based on this result, we can construct confidence intervals and conduct hypothesis testing for θ. The weighted two-sample EL formulation is computationally friendly. It does not involve any nuisance parameters, and the equation (1.6) for the Lagrange multiplier can be solved using the one-sample EL algorithm by Wu (2004).

1.4 Outline of the Thesis

As we discussed in Section 1.1, the method proposed by Huang et al. (2008) for estimating the treatment effect handles the mean responses for the treatment group and the control group separately. Empirical likelihood ratio confidence intervals or tests for the treatment effect cannot be constructed under their approach. In Chapter 2, we propose an alternative EL formulation which directly involves the parameter of interest, i.e., the treatment effect, and incorporates baseline information through an imputation approach. Our focus is to derive the empirical likelihood ratio confidence intervals and tests for the treatment effect under the proposed imputation-based framework. Theoretical results are developed, and finite sample performances of the proposed methods with comparison to existing approaches are investigated through simulation studies. An application to a real data set is also presented.

While the treatment effect, measured as the difference between the two mean responses, is of primary interest, testing the difference of the two distribution functions for the treatment and the control groups is also an important problem. The Mann-Whitney test has been a standard tool for testing the difference of distribution functions with two independent samples. In Chapter 3, we develop empirical likelihood based methods for the Mann-Whitney test to incorporate the two unique features of pretest-posttest studies: (i) the availability of baseline information for both groups; and (ii) the structure of the data with the missing by design property. Our proposed methods combine the standard Mann-Whitney test with the empirical likelihood method of Huang et al. (2008), the imputation-based empirical likelihood method we proposed in Chapter 2, and the jackknife empirical likelihood method of Jing et al. (2009). Theoretical results are presented and finite sample performances of proposed methods are evaluated through a simulation study.

In Chapter 4, we investigate the EL methods for estimating the treatment effect in pretest-posttest studies with observational survey data. Methods based on propensity score modelling are very popular for making causal inference with observation data. We develop methods based on propensity score stratification and propensity scores weighting for estimating the treatment effect while accommodating the complex survey design. We also study the theoretic properties of our proposed estimators. The proposed methods are illustrated through an application using the data from the International Tobacco Control Four Country Surveys (ITC 4C).

In Chapter 5, we summarize the thesis and discuss some possible future work for the topics that we have studied.

Chapter 2

An Imputation Based Empirical Likelihood Approach to Pretest-Posttest Studies

2.1 Introduction

In this chapter, we develop the empirical likelihood based method for making inference of the treatment effect in pretest-posttest studies. Our method has two distinct features: (i) The baseline pretest information is used through a direct model-based imputation procedure; and (ii) The EL formulation involves the parameter of interest directly, not the two separate means of responses for the treatment group and the control group. The imputation procedure effectively exploits the key feature of the pretest-posttest studies where the responses are missing by design. The EL estimation theory employs the framework of two-sample EL procedures proposed by Wu and Yan (2012) where the EL ratio statistic is formulated directly for the parameter of interest.

The rest of the chapter is organized as follows. In Section 2.2, we introduce some notation and summarize the EL method by Huang et al. (2008). In Section 2.3, we present our proposed imputation-based two-sample EL estimator for the treatment effect, using a linear model. Our main result is on the asymptotic distribution of the empirical likelihood ratio statistic for the treatment effect. Section 2.4 extends the result when the linear imputation model is replaced by kernel regression. In Section 2.5, we make theoretical comparisons between the efficiencies of the HQF estimator and the imputation-based EL estimators under suitable conditions. Results from a limited simulation study are presented in Section 2.6. An application using a data set from the ACTG 175 study is reported in Section 2.7. Some concluding remarks are given in Section 2.8. Proofs of theoretical results and regularity conditions are given in Section 2.9.

2.2 Notations and the HQF Estimator

Suppose there are n subjects selected from the target population. Measurements on some baseline variables, $\boldsymbol{Z}$, are taken for all n subjects. Each subject is then randomly assigned to either the treatment group or the control group, with probabilities δ and $1-\delta$ respectively. Let n_1 be the number of subjects in the treatment group, and $n_0 = n - n_1$ be the number of subjects in the control group. Let $R_i = 1$ if subject i is assigned to the treatment group and $R_i = 0$ if subject i is assigned to the control group. Because of the randomization, the marginal distribution of $\boldsymbol{Z}$ is assumed to be identical for the two groups. Let Y_1 and Y_0 be the potential posttest responses that a subject would have if assigned to the treatment group and the control group, respectively. Note that Y_1 will not be observed for any subjects in the control group and Y_0 will not be observed for any subjects in the treatment group. Hence, the observed data for the treatment group are $\{(R_i = 1, \boldsymbol{z}_i, y_{1i}) : i = 1, \cdots, n_1\}$, and the observed data for the control group are $\{(R_i = 0, \boldsymbol{z}_i, y_{0i}) : i = n_1 + 1, \cdots, n\}$. Let $\mu_1 = E(Y_1)$ and $\mu_0 = E(Y_0)$. The parameter of interest is the treatment effect $\theta = \mu_1 - \mu_0$.

Huang et al. (2008) proposed to estimate the treatment effect using the empirical likelihood method. However, instead of estimating the treatment effect θ directly, the authors focused on estimating μ_1 and μ_0 separately. The HQF estimator of μ_1 is computed as $\hat{\mu}_{1\text{HQF}} = \sum_{i=1}^{n_1} \hat{p}_i y_{1i}$, where $\hat{p}_i$ are obtained through the following EL method. Let $f(\boldsymbol{z}, y_1)$ be the joint density function of $(\boldsymbol{Z}, Y_1)$ related to the treatment group and $f(\boldsymbol{z})$ be the marginal density function of $\boldsymbol{Z}$. Let $p_i = f(\boldsymbol{z}_i, y_{1i})$ for $i = 1, \cdots, n_1$ and $r_i = f(\boldsymbol{z}_i)$ for $i = n_1 + 1, \cdots, n$. The log empirical likelihood function is given by

$$\ell = \sum_{i=1}^{n_1} \log(p_i) + \sum_{i=n_1+1}^{n} \log(r_i)\,. \tag{2.1}$$

The $\hat{p}_i$ and $\hat{r}_i$ are obtained by maximizing (2.1) subject to $p_i > 0$, $r_i > 0$ and the following constraints:

$$\sum_{i=1}^{n_1} p_i = 1, \quad \sum_{i=n_1+1}^{n} r_i = 1, \tag{2.2}$$

$$\sum_{i=1}^{n_1} p_i a_1(\boldsymbol{z}_i) = \sum_{i=n_1+1}^{n} r_i a_1(\boldsymbol{z}_i)\,, \tag{2.3}$$

where $a_1(\boldsymbol{z}) = E(Y_1|\boldsymbol{Z} = \boldsymbol{z})$. It is assumed that $E[a_1(\boldsymbol{Z})]^2 < \infty$. The constraint (2.3) is the most crucial part for the HQF estimator, since it uses the baseline information from both the treatment group and the control group. The actual form of $a_1(\boldsymbol{z})$ is typically unknown, but one could use a guessed form, with possible loss of efficiency for the final estimator. The solutions to this constrained maximization problem are given by

$$\hat{p}_i = \frac{1}{n_1}\frac{1}{1+\lambda\{a_1(\boldsymbol{z}_i) - b\}}\,, i = 1, \cdots, n_1 \;\text{ and }\; \hat{r}_i = \frac{1}{n_0}\frac{1}{1+\tau\{a_1(\boldsymbol{z}_i) - b\}}\,, i = n_1+1, \cdots, n$$

for a fixed value of $b = \sum_{i=1}^{n_1} p_i a_1(\boldsymbol{z}_i) = \sum_{i=n_1+1}^{n} r_i a_1(\boldsymbol{z}_i)$. The Lagrange multipliers λ and τ are determined by solving

$$\frac{1}{n_1}\sum_{i=1}^{n_1} \frac{a_1(\boldsymbol{z}_i) - b}{1+\lambda\{a_1(\boldsymbol{z}_i) - b\}} = 0 \;\text{ and }\; \frac{1}{n_0}\sum_{i=n_1+1}^{n} \frac{a_1(\boldsymbol{z}_i) - b}{1+\tau\{a_1(\boldsymbol{z}_i) - b\}} = 0\,.$$

The final value of b used for computing the $\hat{p}_i$ can be obtained through profiling over the log empirical likelihood function.

Huang et al. (2008) showed that $\hat{\mu}_{1\text{HQF}}$ has the following asymptotic representation:

$$\begin{aligned}\hat{\mu}_{1\text{HQF}} \;=\; & \frac{1}{n}\sum_{i=1}^{n} \frac{R_i y_{1i}}{\delta} - E\big\{Y_1 \boldsymbol{\psi}_1(\boldsymbol{Z})^T\big\}\big[E\big\{\boldsymbol{\psi}_1(\boldsymbol{Z})\boldsymbol{\psi}_1(\boldsymbol{Z})^T\big\}\big]^{-1}\Big\{\frac{1}{n}\sum_{i=1}^{n} \frac{R_i - \delta}{\delta}\boldsymbol{\psi}_1(\boldsymbol{z}_i)\Big\} \\ & + o_p\big(n^{-1/2}\big)\,,\end{aligned}$$

where $\boldsymbol{\psi}_1(\boldsymbol{z}) = (1, a_1(\boldsymbol{z}))^T$. The authors used this asymptotic representation to prove that, under certain regularity conditions, $\sqrt{n}(\hat{\mu}_{1\text{HQF}} - \mu_1) \to N(0, \sigma_1^2)$, where

$$\sigma_1^2 = \left\{\delta^{-1}E(Y_1^2) - \mu_1^2\right\} - (1-\delta)\delta^{-1}E\left\{Y_1\boldsymbol{\psi}_1(\boldsymbol{Z})^T\right\}\left[E\left\{\boldsymbol{\psi}_1(\boldsymbol{Z})\boldsymbol{\psi}_1(\boldsymbol{Z})^T\right\}\right]^{-1}E\left\{Y_1\boldsymbol{\psi}_1(\boldsymbol{Z})\right\}.$$

Moreover, the authors showed that $\hat{\mu}_{1\text{HQF}}$ is as efficient as the estimator proposed in Leon *et al.* (2003) when $a_1(\boldsymbol{z}) = E(Y_1|\boldsymbol{Z} = \boldsymbol{z})$ is correctly specified for both methods but the HQF estimator is more efficient when a misspecified $a_1(\boldsymbol{z})$ is used for both cases.

Huang *et al.* (2008) proposed to use the same method to estimate μ_0 by $\hat{\mu}_{0\text{HQF}}$, with $\{y_{1i},\, i = 1, \cdots, n_1\}$ replaced by $\{y_{0i},\, i = n_1 + 1, \cdots, n\}$. The same constraint (2.3) is used where $a_1(\boldsymbol{z})$ is replaced by $a_0(\boldsymbol{z}) = E(Y_0|\boldsymbol{Z} = \boldsymbol{z})$. The treatment effect is then estimated as $\hat{\theta}_{\text{HQF}} = \hat{\mu}_{1\text{HQF}} - \hat{\mu}_{0\text{HQF}}$. For confidence intervals or hypothesis tests on θ, the authors proposed to use a nonparametric bootstrap method to estimate the variance of $\hat{\theta}_{\text{HQF}}$. In Section 2.5, we will provide an explicit form of the asymptotic variance of $\hat{\theta}_{\text{HQF}}$. It should be noted that empirical likelihood ratio tests on the treatment effect θ are not available under the EL approach used by HQF.

An interesting and practically useful observation for the HQF estimator is that, if $\boldsymbol{Z}$ is univariate and $a_1(z) = \gamma_0 + \gamma_1 z$ for some unknown γ_0 and γ_1, the constraint (2.3) is equivalent to $\sum_{i=1}^{n_1} p_i z_i = \sum_{i=n_1+1}^{n} r_i z_i$ under the normalization constraints $\sum_{i=1}^{n_1} p_i = 1$ and $\sum_{i=n_1+1}^{n} r_i = 1$.

2.3 Linear Regression Imputation-Based Empirical Likelihood Approach

In this section, we propose an alternative empirical likelihood approach to inferences for pretest-posttest studies. Our method effectively exploits the two distinct features of the problem: (i) availability of baseline information for all subjects in the studies, and (ii) response variables missing by design. Our formulation of the EL function involves directly the parameter of interest, $\theta = \mu_1 - \mu_0$, not the two separate means of the post-test responses. Our primary objective is to develop the empirical likelihood ratio test for the treatment effect θ.

The two features of pretest-posttest sample data can be better summarized through the following table:

i	1	2	$\cdots$	n_1	n_1+1	n_1+2	$\cdots$	n
$\boldsymbol{Z}$	$\boldsymbol{Z}_1$	$\boldsymbol{Z}_2$	$\cdots$	$\boldsymbol{Z}_{n_1}$	$\boldsymbol{Z}_{n_1+1}$	$\boldsymbol{Z}_{n_1+2}$	$\cdots$	$\boldsymbol{Z}_n$
Y_1	Y_{11}	Y_{12}	$\cdots$	Y_{1n_1}	$*$	$*$	$\cdots$	$*$
Y_0	$*$	$*$	$\cdots$	$*$	$Y_{0(n_1+1)}$	$Y_{0(n_1+2)}$	$\cdots$	Y_{0n}

The complete observations of $\boldsymbol{Z}$ on all subjects provide an opportunity to impute the missing values "$*$" of the response variables due to the unique design used for the studies. The imputation-based approach not only uses the baseline information in a more effective way but also produces two samples with enlarged sample sizes. We first consider the following linear regression models for the two response variables Y_1 and Y_0:

$$Y_{1i} = \boldsymbol{Z}_i^T \boldsymbol{\beta}_1 + \epsilon_{1i}, i = 1, \cdots, n\,, \tag{2.4}$$

$$Y_{0i} = \boldsymbol{Z}_i^T \boldsymbol{\beta}_0 + \epsilon_{0i}, i = 1, \cdots, n\,, \tag{2.5}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ are respectively the regression parameters for the treatment and the control, and ϵ_{1i}'s and ϵ_{0i}'s are independent errors with zero mean and variance $\sigma_{\epsilon 1}^2$ and $\sigma_{\epsilon 0}^2$, respectively. It is assumed for simplicity that both models (2.4) and (2.5) include an intercept. The case where there is no intercept, discussed in Sections 2.5 and 2.6, can be handled similarly. The two assumed models imply that the missing responses in one group would follow the same model if the subjects were assigned to the other group.

We consider deterministic regression imputation for the missing responses. Let

$$\hat{\boldsymbol{\beta}}_1 = \Big(\sum_{i=1}^{n} R_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T\Big)^{-1} \sum_{i=1}^{n} R_i \boldsymbol{Z}_i Y_{1i},$$

$$\hat{\boldsymbol{\beta}}_0 = \Big(\sum_{i=1}^{n} (1-R_i) \boldsymbol{Z}_i \boldsymbol{Z}_i^T\Big)^{-1} \sum_{i=1}^{n} (1-R_i) \boldsymbol{Z}_i Y_{0i}$$

be the ordinary least squares estimators for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$. Let

$$Y_{1i}^* = \boldsymbol{Z}_i^T \hat{\boldsymbol{\beta}}_1\,, \quad i = n_1 + 1, \cdots, n \quad \text{and} \quad Y_{0i}^* = \boldsymbol{Z}_i^T \hat{\boldsymbol{\beta}}_0\,, \quad i = 1, \cdots, n_1$$

be respectively the imputed values of Y_1 for the subjects in the control group and the imputed values of Y_0 for the subjects in the treatment group. Note that $E(\mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_1) = E(Y_{1i}) = \mu_1$, and $E(\mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_0) = E(Y_{0i}) = \mu_0$. After the imputation, we obtain two augmented samples for the two posttest response variables given by

$$\{\tilde{Y}_{1i} = R_i Y_{1i} + (1-R_i) Y_{1i}^*, i = 1, \cdots, n\} \;\text{ and }\; \{\tilde{Y}_{0i} = (1-R_i) Y_{0i} + R_i Y_{0i}^*, i = 1, \cdots, n\}.$$

We develop a two-sample empirical likelihood method for the parameter of interest $\theta = \mu_1 - \mu_0 = E(Y_1) - E(Y_0)$, using the formulation described in Wu and Yan (2012). Our primary objective is to construct an EL test on the treatment effect θ using the empirical likelihood ratio statistic. The log empirical likelihood function is given by

$$\ell(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{n} \log(p_i) + \sum_{i=1}^{n} \log(q_i)\,,$$

where $\boldsymbol{p} = (p_1, \cdots, p_n)^T$, $\boldsymbol{q} = (q_1, \cdots, q_n)^T$, $p_i = f(y_{1i})$, $i = 1, \cdots, n$, $q_i = g(y_{0i})$, $i = 1, \cdots, n$, and $f(\cdot)$ and $g(\cdot)$ are the marginal density functions for Y_1 and Y_0. For a fixed value of θ, let $\boldsymbol{p}(\theta) = (p_1(\theta), \cdots, p_n(\theta))^T$ and $\boldsymbol{q}(\theta) = (q_1(\theta), \cdots, q_n(\theta))^T$ be the maximizer of $\ell(\boldsymbol{p}, \boldsymbol{q})$ subject to $p_i > 0$, $q_i > 0$ and the constraints

$$\sum_{i=1}^{n} p_i = 1\,, \quad \sum_{i=1}^{n} q_i = 1\,, \tag{2.6}$$

$$\sum_{i=1}^{n} p_i \tilde{Y}_{1i} - \sum_{i=1}^{n} q_i \tilde{Y}_{0i} = \theta. \tag{2.7}$$

There exists a computational algorithm for finding the solution to this constrained maximization problem for a fixed θ without introducing any additional parameters. See Wu and Yan (2012) for further detail. The maximum EL estimator of θ under the assumed linear models is given by

$$\hat{\theta}_{\text{EL}}^{lin} = \frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_{1i} - \frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_{0i} = \bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 \,. \tag{2.8}$$

We now present one of our major results on the EL ratio statistic on θ. Let $\hat{\boldsymbol{p}}(\theta) = (\hat{p}_1(\theta), \cdots, \hat{p}_n(\theta))^T$ and $\hat{\boldsymbol{q}}(\theta) = (\hat{q}_1(\theta), \cdots, \hat{q}_n(\theta))^T$ be the maximizer of $\ell(\boldsymbol{p}, \boldsymbol{q})$ under the constraints (2.6) and (2.7) for a fixed θ. Let

$$r(\theta) = \sum_{i=1}^{n}\log\left(n\hat{p}_i(\theta)\right) + \sum_{i=1}^{n}\log\left(n\hat{q}_i(\theta)\right)$$

be the EL ratio statistic on θ. We have the following result regarding the asymptotic distribution of $r(\theta)$.

Theorem 1. *Suppose that $E(\|\boldsymbol{Z}\|^2) < \infty$, $\sigma_{\epsilon 1}^2 < \infty$, $\sigma_{\epsilon 0}^2 < \infty$ and $n_1/n \to \delta \in (0,1)$ as $n \to \infty$. Suppose also that models (2.4) and (2.5) hold. Then $-2r(\theta)/c_1$ converges in distribution to a χ^2 random variable with one degree of freedom as $n \to \infty$, where $\theta = E(Y_1) - E(Y_0) = \mu_1 - \mu_0$. The scaling constant c_1 is given by $c_1 = \{(\tilde{V}_1 + \tilde{V}_0)/V\}^{-1}$, where $V = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T\Sigma_Z(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + \delta^{-1}\sigma_{\epsilon 1}^2 + (1-\delta)^{-1}\sigma_{\epsilon 0}^2$, $\tilde{V}_1 = n^{-1}\sum_{i=1}^{n}(\tilde{Y}_{1i} - \mu_1)^2$, $\tilde{V}_0 = n^{-1}\sum_{i=1}^{n}(\tilde{Y}_{0i} - \mu_0)^2$, and Σ_Z is the variance-covariance matrix of $\boldsymbol{Z}$.*

From the proof of Theorem 1 presented in Section 2.9.2 we see that V/n is the asymptotic variance of $\hat{\theta}_{\text{EL}}^{lin}$. Under the same conditions of the theorem, we have that $\sqrt{n}(\hat{\theta}_{\text{EL}}^{lin} - \theta)$ converges in distribution to $N(0, V)$. Results of Theorem 1 can be used to construct the $(1-\alpha)$-level EL ratio confidence interval on θ: $\mathcal{C}_1 = \{\theta \mid -2r(\theta)/\hat{c}_1 \le \chi_1^2(\alpha)\}$, where $\chi_1^2(\alpha)$ is the upper α quantile of the χ_1^2 distribution and $\hat{c}_1$ is a consistent estimator of the scaling constant c_1. It can be shown that if $\hat{c}_1$ is a consistent estimator of c_1 such that $\hat{c}_1 = c_1 + o_p(1)$, then $-2r(\theta)/\hat{c}_1$ also converges in distribution to a χ_1^2 random variable.

2.4 Kernel Regression Imputation-Based Empirical Likelihood Approach

The results presented in Section 2.3 require the validity of the assumed linear regression models (2.4) and (2.5). In this section, we consider nonparametric kernel regression models as a robust alternative. Imputation for missing responses based on a kernel regression model was discussed in Cheng (1994). Wang and Rao (2002) considered the one-sample EL method with kernel regression imputation for missing values. Let $m_1(\boldsymbol{z}) = E(Y_1|\boldsymbol{Z} = \boldsymbol{z})$ and $m_0(\boldsymbol{z}) = E(Y_0|\boldsymbol{Z} = \boldsymbol{z})$. We replace the linear regression imputed values $Y_{1i}^* = \boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}}_1$ and $Y_{0i}^* = \boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}}_0$ by kernel regression imputed values $\hat{m}_1(\boldsymbol{Z}_i)$ and $\hat{m}_0(\boldsymbol{Z}_i)$, respectively. Cheng (1994) used the following kernel estimators for $m_1(\boldsymbol{z})$ and $m_0(\boldsymbol{z})$:

$$\hat{m}_1(\boldsymbol{z}) = \sum_{i=1}^n R_i Y_{1i} K((\boldsymbol{z} - \boldsymbol{Z}_i)/h_n) / \sum_{i=1}^n R_i K((\boldsymbol{z} - \boldsymbol{Z}_i)/h_n)\,, \tag{2.9}$$

$$\hat{m}_0(\boldsymbol{z}) = \sum_{i=1}^n (1-R_i) Y_{0i} K((\boldsymbol{z} - \boldsymbol{Z}_i)/h_n) / \sum_{i=1}^n (1-R_i) K((\boldsymbol{z} - \boldsymbol{Z}_i)/h_n)\,, \tag{2.10}$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth sequence which decreases to zero as n goes to infinity. When the sample sizes are not large enough, neighbourhoods of certain values of $\boldsymbol{z}$ might contain very few observations, which might cause $\hat{m}_1(\boldsymbol{z})$ or $\hat{m}_1(\boldsymbol{z})$ to be very unstable. Wang and Rao (2002) proposed to use the following modified versions of the kernel estimators by first defining

$$\hat{g}_1(\boldsymbol{z}) = (nh_n)^{-1} \sum_{i=1}^n R_i K((\boldsymbol{z} - \boldsymbol{Z}_i)/h_n) \quad \text{and} \quad \hat{g}_{1b_n}(\boldsymbol{z}) = \max\{\hat{g}_1(\boldsymbol{z}), b_n\}\,,$$

$$\hat{g}_0(\boldsymbol{z}) = (nh_n)^{-1} \sum_{i=1}^n (1-R_i) K((\boldsymbol{z} - \boldsymbol{Z}_i)/h_n) \quad \text{and} \quad \hat{g}_{0b_n}(\boldsymbol{z}) = \max\{\hat{g}_0(\boldsymbol{z}), b_n\}$$

for a suitably chosen sequence b_n, and then replacing $\hat{m}_1(\boldsymbol{z})$ by $\hat{m}_{1b_n}(\boldsymbol{z}) = \hat{m}_1(\boldsymbol{z})\hat{g}_1(\boldsymbol{z})/\hat{g}_{1b_n}(\boldsymbol{z})$, $\hat{m}_0(\boldsymbol{z})$ by $\hat{m}_{0b_n}(\boldsymbol{z}) = \hat{m}_0(\boldsymbol{z})\hat{g}_0(\boldsymbol{z})/\hat{g}_{0b_n}(\boldsymbol{z})$.

The development presented in Section 2.3 under linear regression imputation can now be imitated under kernel regression imputation if we simply define

$$\begin{aligned}\tilde{Y}_{1i}^{kel} &= R_i Y_{1i} + (1-R_i)\hat{m}_{1b_n}(\boldsymbol{Z}_i)\,, \quad i=1,2,\cdots,n\,,\\ \tilde{Y}_{0i}^{kel} &= (1-R_i)Y_{0i} + R_i\hat{m}_{0b_n}(\boldsymbol{Z}_i)\,, \quad i=1,2,\cdots,n\,.\end{aligned}$$

The maximum EL estimator of θ under the assumed kernel regression models is given by

$$\hat{\theta}_{\text{EL}}^{kel} = \frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_{1i}^{kel} - \frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_{0i}^{kel} = \bar{\tilde{Y}}_1^{kel} - \bar{\tilde{Y}}_0^{kel}\,. \tag{2.11}$$

Let $r^{kel}(\theta)$ be defined in the same way as $r(\theta)$ is computed in Section 2.3, with $\tilde{Y}_{1i}$ and $\tilde{Y}_{0i}$ respectively being replaced by $\tilde{Y}_{1i}^{kel}$ and $\tilde{Y}_{0i}^{kel}$.

Theorem 2. *Under the conditions C1-C6 specified in Section 2.9.3, $-2r^{kel}(\theta)/c_2$ converges in distribution to a χ^2 random variable with one degree of freedom when $n \to \infty$ and $\theta = \mu_1 - \mu_0$. The scaling constant c_2 is given by $c_2 = V^{kel}/(\tilde{V}_1^{kel} + \tilde{V}_0^{kel})$, where*

$$\begin{aligned}\tilde{V}_1^{kel} &= n^{-1}\sum_{i=1}^{n}(\tilde{Y}_{1i}^{kel} - \mu_1)^2\,, \qquad \tilde{V}_0^{kel} = n^{-1}\sum_{i=1}^{n}(\tilde{Y}_{0i}^{kel} - \mu_0)^2\,,\\ V^{kel} &= Var(m_1(\boldsymbol{Z}) - m_0(\boldsymbol{Z})) + \delta^{-1}E(\sigma_1^2(\boldsymbol{Z})) + (1-\delta)^{-1}E(\sigma_0^2(\boldsymbol{Z}))\end{aligned}$$

with $\sigma_j^2(\boldsymbol{z}) = Var(Y_j|\boldsymbol{Z}=\boldsymbol{z})$ for $j=1,0$.

From the proof of Theorem 2 presented in Section 2.9.4 we see that V^{kel}/n is the asymptotic variance of the maximum EL estimator $\hat{\theta}_{\text{EL}}^{kel}$ and $\sqrt{n}(\hat{\theta}_{\text{EL}}^{kel} - \theta)$ converges in distribution to $N(0, V^{kel})$. A $(1-\alpha)$-level EL ratio confidence interval on θ can be constructed as $\mathcal{C}_2 = \{\theta \mid -2r^{kel}(\theta)/\hat{c}_2 \le \chi_1^2(\alpha)\}$, where $\hat{c}_2$ is a consistent estimator of c_2 and $\chi_1^2(\alpha)$ is the upper $\alpha\%$ quantile of the χ_1^2 distribution.

Two of the major issues with kernel regression modelling are the curse of dimensionality and bandwidth selection. The method presented in this section is most helpful when the linear regression models are questionable and the baseline variables $\boldsymbol{Z}$ are of low dimension. In practice, the optimal bandwidth may be difficult to estimate, and the choice of bandwidth can be determined by a data-dependent cross-validation method. The cross validated bandwidth minimizes the sum of squared errors between the data and the estimates from the kernel regression. The procedure of a K-fold cross validation (Hastie et al. (2001)) can be summarized as following: consider a possible set of bandwidths $\{h_1, \cdots, h_p\}$. For each h_i, $i = 1, \cdots, p$, we

1. randomly split the data into (roughly) K equal-sized parts for both treatment group and control group;

2. for the k^{th} part (test), fit kernel regression with bandwidth h_i to the other $K-1$ parts of data (training), and calculate the sum of the squared errors of the fitted values and the true data of the k^{th} part for both groups;

3. repeat number 2 for $k = 1, \cdots, K$, and then calculate the total sum of squared errors.

Repeat the process for $i = 1, \cdots, p$, then the cross-validated bandwidth h^* is the one with the smallest total sum of squared errors.

2.5 Efficiency Comparisons Among Alternative EL Approaches

Our proposed imputation-based EL approaches presented in Sections 2.3 and 2.4 focus on empirical likelihood ratio confidence intervals or tests for the treatment effect, i.e., $\theta = \mu_1 - \mu_0$. The EL approach used by Huang et al. (2008), on the other hand, puts major effort on the point estimation of μ_1 and μ_0 separately. EL ratio confidence intervals on θ are not available in the latter case. In this section, we provide comparisons among the point estimators $\hat{\theta}_{\text{HQF}}$, $\hat{\theta}_{\text{EL}}^{lin}$ and $\hat{\theta}_{\text{EL}}^{kel}$ in terms of asymptotic variances. Some detailed derivations are omitted, since they are similar to those appearing in the proofs of Theorems 1 and 2. We use $AV(\hat{\theta})$ to denote the asymptotic variance of $\hat{\theta}$.

We first derive the asymptotic variance of the HQF estimator of θ. Recall that $\boldsymbol{\psi}_j(\boldsymbol{z}) = (1, a_j(\boldsymbol{z}))^T$, where $a_j(\boldsymbol{z}) = E(Y_j|\boldsymbol{Z} = \boldsymbol{z}),\ \ j = 1, 0$. In Huang et al. (2008), the authors have shown that the estimators $\hat{\mu}_{1\text{HQF}}$ and $\hat{\mu}_{0\text{HQF}}$ have the following asymptotic representations:

$$
\begin{aligned}
\hat{\mu}_{1\text{HQF}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{R_i y_{1i}}{\delta} - E\{Y_1\boldsymbol{\psi}_1(\boldsymbol{Z})^T\}E\{\boldsymbol{\psi}_1(\boldsymbol{Z})\boldsymbol{\psi}_1(\boldsymbol{Z})^T\}^{-1} \\
&\quad \times\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{R_i-\delta}{\delta}\boldsymbol{\psi}_1(\boldsymbol{z}_i)\Big\} + o_p(n^{-1/2}), \\
\hat{\mu}_{0\text{HQF}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{(1-R_i)y_{0i}}{1-\delta} - E\{Y_0\boldsymbol{\psi}_0(\boldsymbol{Z})^T\}E\{\boldsymbol{\psi}_0(\boldsymbol{Z})\boldsymbol{\psi}_0(\boldsymbol{Z})^T\}^{-1} \\
&\quad \times\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{(1-R_i)-(1-\delta)}{1-\delta}\boldsymbol{\psi}_0(\boldsymbol{z}_i)\Big\} + o_p(n^{-1/2}).
\end{aligned}
$$

To make the setting comparable to the kernel regression models used in Section 2.4, we assume that $\sigma_{\epsilon j}^2 = Var(Y_j|\boldsymbol{Z} = \boldsymbol{z})$ is dependent of $\boldsymbol{z}$ for $j = 1, 0$. It follows that $E[a_1(\boldsymbol{Z})] = \mu_1$, $E\{Y_1\boldsymbol{\psi}_1(\boldsymbol{Z})^T\} = (\mu_1, E[\{a_1(\boldsymbol{Z})\}^2])$ and

$$
E\{Y_1\boldsymbol{\psi}_1(\boldsymbol{Z})^T\}E\{\boldsymbol{\psi}_1(\boldsymbol{Z})\boldsymbol{\psi}_1(\boldsymbol{Z})^T\}^{-1} = (0, 1)\,.
$$

The asymptotic representation of $\hat{\mu}_{1\text{HQF}}$ can be rewritten as

$$
\hat{\mu}_{1\text{HQF}} = \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{R_i y_{1i}}{\delta} - \frac{R_i-\delta}{\delta}a_1(\boldsymbol{z}_i)\Big\} + o_p(n^{-1/2})\,.
$$

With parallel development, we also have

$$
\hat{\mu}_{0\text{HQF}} = \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{(1-R_i)y_{0i}}{1-\delta} - \frac{(1-R_i)-(1-\delta)}{1-\delta}a_0(\boldsymbol{z}_i)\Big\} + o_p(n^{-1/2})\,,
$$

Therefore,

$$
\begin{aligned}
\hat{\theta}_{\text{HQF}} &= \hat{\mu}_{1\text{HQF}} - \hat{\mu}_{0\text{HQF}} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{R_i y_{1i}}{\delta} - \frac{R_i-\delta}{\delta}a_1(\boldsymbol{z}_i)\Big\} - \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{(1-R_i)y_{0i}}{1-\delta} - \frac{(1-R_i)-(1-\delta)}{1-\delta}a_0(\boldsymbol{z}_i)\Big\} \\
&\quad + o_p(n^{-1/2}).
\end{aligned}
$$

Let σ_1^2 and σ_0^2 be the asymptotic variances of $\hat{\mu}_{1\text{HQF}}$ and $\hat{\mu}_{0\text{HQF}}$, then from Huang et al. (2008), $\sigma_1^2 = (1/\delta)E(Y_1^2) - \mu_1^2 - [(1-\delta)/\delta]E\{(a_1(\boldsymbol{Z}))^2\}$ and $\sigma_0^2 = [1/(1-\delta)]E(Y_0^2) - \mu_0^2 - [\delta/(1-\delta)]E\{(a_0(\boldsymbol{Z}))^2\}$. Therefore, the asymptotic variance of $\hat{\theta}_{\text{HQF}}$ is given by

$$\begin{aligned} AV(\sqrt{n}\hat{\theta}_{\text{HQF}}) &= AV\{\sqrt{n}(\hat{\mu}_{1\text{HQF}} - \hat{\mu}_{0\text{HQF}})\} \\ &= \Big\{\frac{1}{\delta}E(Y_1^2) - \mu_1^2\Big\} + \Big\{\frac{1}{1-\delta}E(Y_0^2) - \mu_0^2\Big\} - \\ &\quad \frac{1-\delta}{\delta}E\{(a_1(\boldsymbol{Z}))^2\} - \frac{\delta}{1-\delta}E\{(a_0(\boldsymbol{Z}))^2\} - 2E\{a_1(\boldsymbol{Z})a_0(\boldsymbol{Z})\} + 2\mu_1\mu_0 \\ &= Var\{a_1(\boldsymbol{Z}) - a_0(\boldsymbol{Z})\} + \delta^{-1}\sigma_{\epsilon 1}^2 + (1-\delta)^{-1}\sigma_{\epsilon 0}^2 \,. \end{aligned}$$

It is now clear that, if the functions $a_j(\boldsymbol{z}) = E(Y_j|\boldsymbol{Z} = \boldsymbol{z})$, $j = 1, 0$ are correctly specified, the HQF estimator $\hat{\theta}_{\text{HQF}}$ and the kernel imputation-based EL estimator $\hat{\theta}_{\text{EL}}^{kel}$ have the same asymptotic variance, since $AV(\sqrt{n}\hat{\theta}_{\text{EL}}^{kel})$ given in Theorem 2 is identical to $AV(\sqrt{n}\hat{\theta}_{\text{HQF}})$, where $m_j(\boldsymbol{Z}) = a_j(\boldsymbol{Z})$ and $\sigma_j^2(\boldsymbol{Z}) = \sigma_{\epsilon j}^2$. Huang et al. (2008) showed that, with correctly specified $a_j(\boldsymbol{z})$, the estimator $\hat{\theta}_{\text{HQF}}$ achieves the semiparametric efficiency lower bound. It follows that the kernel imputation-based approach presented in Section 2.4 is efficient without the need to specify the mean function $m_j(\boldsymbol{z})$.

Under the two linear models (2.4) and (2.5), we have $a_j(\boldsymbol{Z}) = \boldsymbol{Z}^T\boldsymbol{\beta}_j$, $j = 1, 0$ and

$$AV(\sqrt{n}\hat{\theta}_{\text{HQF}}) = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T\Sigma_Z(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + \delta^{-1}\sigma_{\epsilon 1}^2 + (1-\delta)^{-1}\sigma_{\epsilon 0}^2 \,. \tag{2.12}$$

If the models (2.4) and (2.5) both include an intercept, then $AV(\sqrt{n}\hat{\theta}_{\text{HQF}})$ given by (2.12) is identical to $AV(\sqrt{n}\hat{\theta}_{\text{EL}}^{lin})$ given in Theorem 1. A key result in the proof is that, if an intercept is included in the two linear models, we have $E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z}) = 1$. In this case our imputation-based EL approach has the same efficiency as the EL approach of Huang *et al.* (2008).

If an intercept is not part of the models (2.4) and (2.5), the asymptotic variance formula $AV(\sqrt{n}\hat{\theta}_{\text{HQF}})$ given by (2.12) remains the same. For the imputation-based approach presented in Section 2.3 under the assumed linear regression models, it can be shown that

$$\begin{aligned} AV(\sqrt{n}\hat{\theta}_{\text{EL}}^{lin}) &= (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T\Sigma_Z(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + \delta\sigma_{\epsilon 1}^2 + (1-\delta)\sigma_{\epsilon 0}^2 + \\ &\quad \{[(1-\delta)^2\delta^{-1} + 2(1-\delta)]\sigma_{\epsilon 1}^2 + [\delta^2(1-\delta)^{-1} + 2\delta]\sigma_{\epsilon 0}^2\}K(\boldsymbol{Z}) \,, \end{aligned}$$

where $K(\boldsymbol{Z}) = E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z})$. It follows that $AV(\sqrt{n}\hat{\theta}_{\text{EL}}^{lin}) \leq AV(\sqrt{n}\hat{\theta}_{\text{HQF}})$ if $K(\boldsymbol{Z}) \leq 1$. Suppose A is a $k \times k$ positive definite matrix, and a is a $k \times 1$ vector. It can be shown that

$$(A + aa^T)^{-1} = A^{-1} - \frac{A^{-1}aa^TA^{-1}}{1 + a^TA^{-1}a}.$$

Let $A = Var(\boldsymbol{Z})$ and $a = E(\boldsymbol{Z})$, then we have

$$\begin{aligned} K(\boldsymbol{Z}) &= a^T(A + aa^T)^{-1}a = a^T\left(A^{-1} - \frac{A^{-1}aa^TA^{-1}}{1 + a^TA^{-1}a}\right)a \\ &= a^TA^{-1}a - \frac{a^TA^{-1}aa^TA^{-1}a}{1 + a^TA^{-1}a} = \frac{a^TA^{-1}a}{1 + a^TA^{-1}a} \\ &= 1 - \frac{1}{1 + a^TA^{-1}a} \\ &\leq 1\,. \end{aligned}$$

Therefore, when the models (2.4) and (2.5) don't have an intercept, $\hat{\theta}_{\text{EL}}^{lin}$ is more efficient than $\hat{\theta}_{\text{HQF}}$. A possible explanation for this phenomenon is that the HQF formulation makes an inexplicit assumption that an intercept is always included in the model. For example, if we consider a univariate Z in the models, then the constraint (2.3) reduces to $\sum_{i=1}^{n_1} p_i z_i = \sum_{i=n_1+1}^{n} r_i z_i$ with or without an intercept. Our imputation-based approach, on the other hand, makes explicit use of the model structure and hence has one less model parameter to estimate if the intercept is not part of the models.

2.6 Simulation Study

In this section, we present the results from simulation studies to compare the performances of our proposed methods to existing ones. Point estimators, confidence intervals and hypothesis tests for the treatment effect θ are all considered. We first consider two linear regression models.

Model (I) involves a single pretest baseline variable Z without an intercept:

$$Y_{1i} = \beta_1 Z_{1i} + \varepsilon_{1i}, \quad i = 1, \cdots, n_1, \tag{2.13}$$

$$Y_{0j} = \beta_0 Z_{0j} + \varepsilon_{0j}, \quad j = 1, \cdots, n_0. \tag{2.14}$$

The pretest responses Z_1 and Z_0 are generated independently from a standard exponential distribution with $E(Z) = 1$ and $Var(Z) = 1$. The error terms ε_{1i} and ε_{0j} are generated independently from normal distributions with mean 0, variance σ_{e1}^2 and σ_{e0}^2 respectively. The variances are chosen based on the correlation coefficient ρ between Y and Z, i.e., $\sigma_{e1}^2 = \beta_1^2(1/\rho^2 - 1)$ and $\sigma_{e0}^2 = \beta_0^2(1/\rho^2 - 1)$. The true treatment effect is set as $\theta_0 = \mu_1 - \mu_0 = \beta_1 - \beta_0$.

Model (II) has two baseline variables (X, Z) with an intercept:

$$Y_{1i} = \beta_{10} + \beta_{11} X_{1i} + \beta_{12} Z_{1i} + \varepsilon_{1i}, \quad i = 1, \cdots, n_1, \tag{2.15}$$

$$Y_{0j} = \beta_{00} + \beta_{01} X_{0j} + \beta_{02} Z_{0j} + \varepsilon_{0j}, \quad j = 1, \cdots, n_0. \tag{2.16}$$

The added covariate X follows a Bernoulli distribution with probability $p = 0.5$ (representing "gender" of the subjects). The Z variable and the error terms are similarly generated as in Model (I), with the error variances controlled by the correlation coefficient ρ between Y and the linear predictor $\boldsymbol{Z}_i'\boldsymbol{\beta}$. The true treatment effect is given by $\theta_0 = \mu_1 - \mu_0 = (\beta_{10} + \beta_{11}/2 + \beta_{12}) - (\beta_{00} + \beta_{01}/2 + \beta_{02})$.

For each model, we consider three values of the correlation coefficient ρ at 0.8, 0.5 and 0.3, representing strong, moderate and weak relations between the posttest variable Y and the set of pretest measures. We consider different combinations of sample sizes $(n_1, n_0) = (30, 30), (50, 50), (100, 100)$ and $(50, 100)$. For each simulated sample, we compute three point estimates of θ_0: (i) the naive estimator $\hat{\theta} = \bar{Y}_1 - \bar{Y}_0$ using only the posttest observations Y_{1i} and Y_{0j}; (ii) the imputation-based method under the linear model, $\hat{\theta}_{\text{EL}}^{lin}$; and (iii) the EL method of HQF, $\hat{\theta}_{\text{HQF}}$. For confidence intervals on θ_0, alternative methods are considered for each of the three cases: (i) Confidence interval based on normal approximation to $\hat{\theta}$, denoted as NB; the two-sample EL ratio confidence interval of Wu and Yan

(2012), denoted as WY. (ii) Three confidence intervals for the imputation-based approach, denoted as LM1, LM2 and LM3: The first is the EL ratio confidence interval based on the result of Theorem 1; the second replaces the scaled χ^2 approximation used in LM1 by a bootstrap calibration; the third uses a normal approximation to $\hat{\theta}_{\mathrm{EL}}^{lin}$. (iii) Two versions of the confidence intervals based on $\hat{\theta}_{\mathrm{HQF}}$, denoted as HQF1 and HQF2: The first uses bootstrap method to compute the variance of $\hat{\theta}_{\mathrm{HQF}}$, as suggested by Huang *et al.* (2008); the second uses the asymptotic variance formula provided in Section 2.5. Both HQF1 and HQF2 use normal approximations.

Performances of point estimators are assessed by the simulated bias and mean squared error (MSE). Confidence intervals (CI) on θ are evaluated by the simulated coverage probability (CP), lower and upper tail error rates (L and U) and average length (AL). We only consider 95% confidence intervals on θ. For bootstrap methods, the number of bootstrap samples used is 1000. The total number of simulation runs is 1000.

Simulation results under Model (I) are reported in Tables 2.1-2.3. Here are some key observations: (1) All point estimators have negligible bias, with the imputation-based estimator $\hat{\theta}_{\mathrm{EL}}^{lin}$ having the smallest MSE; (2) The estimator $\hat{\theta}_{\mathrm{EL}}^{lin}$ outperforms $\hat{\theta}_{\mathrm{HQF}}$ in all cases, with the largest gain of efficiency under strong correlation between Y and Z (i.e., $\rho = 0.8$); (3) Both $\hat{\theta}_{\mathrm{EL}}^{lin}$ and $\hat{\theta}_{\mathrm{HQF}}$ perform significantly better than NB and WY for all scenarios considered; (4) All confidence intervals have coverage probabilities very close to the nominal value, including the different versions of LM and HQF and the naive method NB and the EL method WY; (5) The three versions of LM confidence intervals are much shorter than other intervals; (6) All confidence intervals have balanced tail error rates under the simulation settings.

Results under Model (II) are summarized in Tables 2.4-2.6. Note that the model has two baseline variables and an intercept. The most striking observation is that both the LM approach and the HQF method perform well but the difference between the two disappears. This is consistent with the results of the theoretical comparisons discussed in Section 2.5.

The second part of the simulation studies examines the effect of model misspecifications.

We include the kernel regression-based method presented in Section 2.4 as part of the comparisons. We consider two kernel functions: the flat kernel function $K(u) = 1/2$, $|u| \le 1$, denoted as "Flat", and the Epanechnikov kernel function $K(u) = 3/4(1 - u^2)$, $|u| \le 1$, denoted as "Epan", for the case of univariate Z. For the case of two baseline variables, we use $K(u_1, u_2) = K(u_1) * K(u_2)$. The bandwidth h_n for each simulation is chosen by a 10-fold cross validation. In addition to the two linear models (I) and (II), we also consider two nonlinear models. Model (I*) involves a single Z variable:

$$\begin{aligned} Y_{1i} &= \theta_0 + 4\sin(Z_{1i}) + \varepsilon_{1i}\,, \quad i = 1, \cdots, n_1\,, \\ Y_{0j} &= \qquad 4\sin(Z_{0j}) + \varepsilon_{0j}\,, \quad j = 1, \cdots, n_0\,. \end{aligned}$$

Model (II*) involves two baseline variables X and Z:

$$\begin{aligned} Y_{1i} &= \theta_0 + 4\sin(X_{1i} + Z_{1i}) + \varepsilon_{1i}\,, \quad i = 1, \cdots, n_1\,, \\ Y_{0j} &= \qquad 4\sin(X_{0j} + Z_{0j}) + \varepsilon_{0j}\,, \quad j = 1, \cdots, n_0\,. \end{aligned}$$

The baseline variables are generated in the same way as in the two linear models. The error terms ε_{1i} and ε_{0j} are generated from $N(0, 2^2)$. The parameter θ_0 is the true value of the treatment effect $E(Y_1) - E(Y_0)$. We consider larger sample sizes $n_1 = n_0 = 200$ in this case, due to the need for kernel smoothing. The truncation sequence b_n is chosen as 0.0001 for Model (I*) and 0.05 for Model (II*). The point estimator under kernel regression is given by $\hat{\theta}_{\text{EL}}^{ker}$. Two confidence intervals on θ_0 are constructed. The EL ratio confidence interval based on Theorem 2 is denoted as KM1; the interval using the asymptotic variance and normal approximation is denoted as KM2.

The simulation results are reported in Table 2.7 for Models (I) and (I*) and in Table 2.8 for Models (II) and (II*). The value of ρ is 0.80 for both Models (I) and (II). The true value of θ_0 is set at 0.3. The last column of the two tables is the power for testing $H_0 : \theta_0 = 0$. The kernel regression method KM produces acceptable results under Models (I) and (II) but is less efficient than the LM method: bigger MSE, wider confidence intervals, and smaller power of the test. The kernel regression method, however, performs much better under the two nonlinear models (I*) and (II*). The LM method fails completely under the Model (I*).

The last part of the simulation focuses on power functions, $\pi(\theta)$, of testing $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$. The results are summarized by plots of the power functions presented in Figures 2.1-2.4. Those plots reinforce what we have observed from the tables. Under Models (I) and (II), the three tests based on linear regression imputation are more powerful than all other tests. Under the two nonlinear Models (I*) and (II*), the two tests based on kernel regression imputation have much larger power. The comparisons are meaningful since all tests have similar size close to the nominal level 5%.

Table 2.1: Inferences on θ under Model (I), $\rho = 0.8$, $\theta_0 = 0.3$

(n_1, n_0)	Methods	Bias	MSE	(L, CP, U)	AL
(30,30)	NB	0.0794	0.1814	(0.019,0.961,0.020)	1.7014
	WY	0.0794	0.1814	(0.025,0.958,0.017)	1.7208
	LM1	0.0122	0.0490	(0.029,0.941,0.030)	0.8657
	LM2	0.0122	0.0490	(0.026,0.956,0.018)	0.9374
	LM3	0.0122	0.0490	(0.028,0.943,0.029)	0.8596
	HQF1	0.0241	0.0731	(0.033,0.939,0.028)	1.0403
	HQF2	0.0241	0.0731	(0.027,0.941,0.032)	1.0422
(50,50)	NB	0.0369	0.1208	(0.026,0.943,0.031)	1.3285
	WY	0.0369	0.1208	(0.026,0.944,0.030)	1.3461
	LM1	0.0164	0.0267	(0.022,0.958,0.020)	0.6561
	LM2	0.0164	0.0267	(0.019,0.964,0.017)	0.6861
	LM3	0.0164	0.0267	(0.021,0.959,0.020)	0.6526
	HQF1	0.0244	0.0421	(0.031,0.944,0.025)	0.7984
	HQF2	0.0244	0.0421	(0.027,0.946,0.027)	0.8010
(100,100)	NB	0.0002	0.0573	(0.023,0.951,0.026)	0.9421
	WY	0.0002	0.0573	(0.027,0.946,0.027)	0.9511
	LM1	-0.0334	0.0144	(0.021,0.945,0.034)	0.4599
	LM2	-0.0334	0.0144	(0.021,0.947,0.032)	0.4695
	LM3	-0.0334	0.0144	(0.020,0.943,0.037)	0.4583
	HQF1	-0.0421	0.0231	(0.024,0.936,0.040)	0.5694
	HQF2	-0.0421	0.0231	(0.022,0.940,0.038)	0.5699
(50,100)	NB	-0.0007	0.0969	(0.023,0.937,0.040)	1.1845
	WY	-0.0007	0.0969	(0.030,0.937,0.033)	1.1992
	LM1	0.0130	0.0224	(0.029,0.944,0.027)	0.5767
	LM2	0.0130	0.0224	(0.036,0.941,0.023)	0.5984
	LM3	0.0130	0.0224	(0.027,0.946,0.027)	0.5735
	HQF1	0.0142	0.0355	(0.029,0.943,0.028)	0.7194
	HQF2	0.0142	0.0355	(0.025,0.949,0.026)	0.7187

Table 2.2: Inferences on θ under Model (I), $\rho = 0.5$, $\theta_0 = 0.3$

(n_1, n_0)	Methods	Bias	MSE	(L, CP, U)	AL
(30,30)	NB	0.1143	0.4777	(0.027,0.951,0.022)	2.7430
	WY	0.1143	0.4777	(0.027,0.952,0.021)	2.7390
	LM1	0.0264	0.2554	(0.030,0.947,0.023)	2.0110
	LM2	0.0264	0.2554	(0.029,0.951,0.020)	2.0903
	LM3	0.0264	0.2554	(0.028,0.948,0.024)	1.9605
	HQF1	0.0644	0.3828	(0.035,0.938,0.027)	2.3839
	HQF2	0.0644	0.3828	(0.030,0.940,0.030)	2.3865
(50,50)	NB	0.0733	0.2983	(0.029,0.940,0.031)	2.1233
	WY	0.0733	0.2983	(0.029,0.942,0.029)	2.1266
	LM1	0.0319	0.1387	(0.023,0.960,0.017)	1.5118
	LM2	0.0319	0.1387	(0.024,0.959,0.017)	1.5475
	LM3	0.0319	0.1387	(0.021,0.963,0.016)	1.4872
	HQF1	0.0559	0.2197	(0.030,0.946,0.024)	1.8297
	HQF2	0.0559	0.2197	(0.026,0.951,0.023)	1.8337
(100,100)	NB	-0.0530	0.1518	(0.023,0.949,0.028)	1.5080
	WY	-0.0530	0.1518	(0.023,0.949,0.028)	1.5101
	LM1	-0.0791	0.0750	(0.021,0.945,0.034)	1.0538
	LM2	-0.0791	0.0750	(0.022,0.946,0.032)	1.0629
	LM3	-0.0791	0.0750	(0.020,0.944,0.036)	1.0446
	HQF1	-0.0979	0.1212	(0.026,0.932,0.042)	1.3031
	HQF2	-0.0979	0.1212	(0.027,0.935,0.038)	1.3051
(50,100)	NB	0.0172	0.2435	(0.026,0.944,0.030)	1.9022
	WY	0.0172	0.2435	(0.028,0.942,0.030)	1.9088
	LM1	0.0275	0.1158	(0.029,0.946,0.025)	1.3287
	LM2	0.0275	0.1158	(0.036,0.937,0.027)	1.3542
	LM3	0.0275	0.1158	(0.028,0.946,0.026)	1.3088
	HQF1	0.0331	0.1848	(0.028,0.943,0.029)	1.6505
	HQF2	0.0331	0.1848	(0.028,0.948,0.024)	1.6476

Table 2.3: Inferences on θ under Model (I), $\rho = 0.3$, $\theta_0 = 0.3$

(n_1, n_0)	Methods	Bias	MSE	(L, CP, U)	AL
(30,30)	NB	0.1656	1.3533	(0.030,0.948,0.022)	4.5855
	WY	0.1656	1.3533	(0.032,0.946,0.022)	4.5670
	LM1	0.0473	0.8583	(0.027,0.955,0.018)	3.7489
	LM2	0.0473	0.8583	(0.031,0.949,0.020)	3.7648
	LM3	0.0473	0.8583	(0.028,0.951,0.021)	3.5920
	HQF1	0.1239	1.2860	(0.033,0.939,0.028)	4.3695
	HQF2	0.1239	1.2860	(0.033,0.939,0.028)	4.3754
(50,50)	NB	0.1269	0.8106	(0.030,0.944,0.026)	3.5340
	WY	0.1269	0.8106	(0.030,0.946,0.024)	3.5311
	LM1	0.0548	0.4652	(0.021,0.962,0.017)	2.8013
	LM2	0.0548	0.4652	(0.025,0.957,0.018)	2.8025
	LM3	0.0548	0.4652	(0.022,0.961,0.017)	2.7245
	HQF1	0.1022	0.7374	(0.033,0.944,0.023)	3.3541
	HQF2	0.1022	0.7374	(0.026,0.953,0.021)	3.3617
(100,100)	NB	-0.1315	0.4321	(0.024,0.939,0.037)	2.5121
	WY	-0.1315	0.4321	(0.026,0.937,0.037)	2.5120
	LM1	-0.1464	0.2519	(0.020,0.948,0.032)	1.9418
	LM2	-0.1464	0.2519	(0.022,0.946,0.032)	1.9350
	LM3	-0.1464	0.2519	(0.020,0.945,0.035)	1.9138
	HQF1	-0.1801	0.4070	(0.027,0.932,0.041)	2.3895
	HQF2	-0.1801	0.4070	(0.026,0.936,0.038)	3.3926
(50,100)	NB	0.0435	0.6695	(0.027,0.950,0.023)	3.1724
	WY	0.0435	0.6695	(0.026,0.952,0.022)	3.1763
	LM1	0.0490	0.3876	(0.028,0.949,0.023)	2.4595
	LM2	0.0490	0.3876	(0.036,0.938,0.026)	2.4524
	LM3	0.0490	0.3876	(0.030,0.944,0.026)	2.3981
	HQF1	0.0610	0.6188	(0.028,0.944,0.028)	3.0266
	HQF2	0.0610	0.6188	(0.028,0.949,0.023)	3.0209

Table 2.4: Inferences on θ under Model (II), $\rho = 0.8$, $\theta_0 = 0.3$

(n_1, n_0)	Methods	Bias	MSE	(L, CP, U)	AL
(30,30)	NB	-0.0387	0.1854	(0.026,0.954,0.020)	1.7186
	WY	-0.0387	0.1854	(0.026,0.952,0.022)	1.7227
	LM1	-0.0410	0.0736	(0.023,0.944,0.033)	1.0514
	LM2	-0.0410	0.0736	(0.020,0.952,0.028)	1.1183
	LM3	-0.0410	0.0736	(0.024,0.945,0.031)	1.0460
	HQF1	-0.0275	0.0749	(0.023,0.946,0.031)	1.0749
	HQF2	-0.0275	0.0749	(0.026,0.942,0.032)	1.0460
(50,50)	NB	0.0399	0.1221	(0.027,0.945,0.028)	1.3295
	WY	0.0399	0.1221	(0.028,0.942,0.030)	1.3356
	LM1	0.0174	0.0441	(0.026,0.948,0.026)	0.8177
	LM2	0.0174	0.0441	(0.018,0.955,0.027)	0.8447
	LM3	0.0174	0.0441	(0.028,0.946,0.026)	0.8142
	HQF1	0.0243	0.0443	(0.028,0.946,0.026)	0.8286
	HQF2	0.0243	0.0443	(0.029,0.945,0.026)	0.8142
(100,100)	NB	-0.0119	0.0566	(0.029,0.950,0.021)	0.9417
	WY	-0.0119	0.0566	(0.029,0.950,0.021)	0.9450
	LM1	-0.0123	0.0226	(0.023,0.951,0.026)	0.5851
	LM2	-0.0123	0.0226	(0.024,0.951,0.025)	0.5923
	LM3	-0.0123	0.0226	(0.024,0.950,0.026)	0.5835
	HQF1	-0.0100	0.0225	(0.024,0.952,0.024)	0.5868
	HQF2	-0.0100	0.0225	(0.025,0.951,0.024)	0.5835
(50,100)	NB	-0.0005	0.0897	(0.029,0.942,0.029)	1.1677
	WY	-0.0005	0.0897	(0.031,0.943,0.026)	1.1715
	LM1	-0.0176	0.0341	(0.022,0.947,0.031)	0.7163
	LM2	-0.0176	0.0341	(0.019,0.952,0.029)	0.7359
	LM3	-0.0176	0.0341	(0.022,0.946,0.032)	0.7134
	HQF1	-0.0145	0.0339	(0.020,0.948,0.032)	0.7252
	HQF2	-0.0145	0.0339	(0.023,0.946,0.031)	0.7134

Table 2.5: Inferences on θ under Model (II), $\rho = 0.5$, $\theta_0 = 0.3$

(n_1, n_0)	Methods	Bias	MSE	(L, CP, U)	AL
(30,30)	NB	-0.0790	0.4720	(0.026,0.947,0.027)	2.7415
	WY	-0.0790	0.4720	(0.029,0.943,0.028)	2.7322
	LM1	-0.0828	0.3758	(0.024,0.945,0.031)	2.4181
	LM2	-0.0828	0.3758	(0.025,0.943,0.032)	2.4721
	LM3	-0.0828	0.3758	(0.029,0.935,0.036)	2.3465
	HQF1	-0.0737	0.3796	(0.026,0.942,0.032)	2.4004
	HQF2	-0.0737	0.3796	(0.030,0.934,0.036)	2.3465
(50,50)	NB	0.0654	0.3096	(0.026,0.950,0.024)	2.1229
	WY	0.0654	0.3096	(0.029,0.946,0.025)	2.1217
	LM1	0.0499	0.2296	(0.028,0.952,0.020)	1.8674
	LM2	0.0499	0.2296	(0.028,0.951,0.021)	1.8808
	LM3	0.0499	0.2296	(0.031,0.946,0.023)	1.8287
	HQF1	0.0539	0.2306	(0.027,0.948,0.025)	1.8576
	HQF2	0.0539	0.2306	(0.033,0.945,0.022)	1.8287
(100,100)	NB	-0.0252	0.1483	(0.029,0.949,0.025)	1.5089
	WY	-0.0252	0.1483	(0.027,0.948,0.025)	1.5092
	LM1	-0.0256	0.1125	(0.020,0.956,0.024)	1.3259
	LM2	-0.0256	0.1125	(0.020,0.958,0.022)	1.3243
	LM3	-0.0256	0.1125	(0.021,0.953,0.026)	1.3105
	HQF1	-0.0238	0.1123	(0.022,0.955,0.023)	1.3157
	HQF2	-0.0238	0.1123	(0.020,0.955,0.025)	1.3105
(50,100)	NB	-0.0339	0.2308	(0.025,0.946,0.029)	1.8662
	WY	-0.0339	0.2308	(0.024,0.947,0.029)	1.8672
	LM1	-0.0475	0.1757	(0.021,0.951,0.028)	1.6350
	LM2	-0.0475	0.1757	(0.020,0.951,0.029)	1.6456
	LM3	-0.0475	0.1757	(0.023,0.947,0.030)	1.6072
	HQF1	-0.0449	0.1751	(0.022,0.944,0.034)	1.6335
	HQF2	-0.0449	0.1751	(0.023,0.948,0.029)	1.6072

Table 2.6: Inferences on θ under Model (II), $\rho = 0.3$, $\theta_0 = 0.3$

(n_1, n_0)	Methods	Bias	MSE	(L, CP, U)	AL
(30,30)	NB	-0.1383	1.3133	(0.023,0.946,0.031)	4.5598
	WY	-0.1383	1.3133	(0.026,0.942,0.032)	4.5385
	LM1	-0.1445	1.2597	(0.022,0.950,0.028)	4.5135
	LM2	-0.1445	1.2597	(0.025,0.942,0.033)	4.4655
	LM3	-0.1445	1.2597	(0.029,0.937,0.034)	4.2862
	HQF1	-0.1404	1.2651	(0.024,0.941,0.035)	4.3756
	HQF2	-0.1404	1.2651	(0.030,0.936,0.034)	4.2862
(50,50)	NB	0.1030	0.8520	(0.026,0.949,0.025)	3.5334
	WY	0.1030	0.8520	(0.027,0.947,0.026)	3.5280
	LM1	0.0978	0.7747	(0.027,0.954,0.019)	3.4713
	LM2	0.0978	0.7747	(0.029,0.951,0.020)	3.4135
	LM3	0.0978	0.7747	(0.034,0.943,0.023)	3.3417
	HQF1	0.0960	0.7771	(0.026,0.948,0.026)	3.3908
	HQF2	0.0960	0.7771	(0.034,0.941,0.025)	3.3417
(100,100)	NB	-0.0449	0.4123	(0.024,0.952,0.024)	2.5162
	WY	-0.0449	0.4123	(0.023,0.955,0.022)	2.5154
	LM1	-0.0453	0.3740	(0.017,0.962,0.021)	2.4472
	LM2	-0.0453	0.3740	(0.019,0.960,0.021)	2.4106
	LM3	-0.0453	0.3740	(0.019,0.955,0.026)	2.3945
	HQF1	-0.0439	0.3739	(0.022,0.954,0.024)	2.4019
	HQF2	-0.0439	0.3739	(0.019,0.956,0.025)	2.3945
(50,100)	NB	-0.0832	0.6392	(0.024,0.942,0.034)	3.1084
	WY	-0.0832	0.6392	(0.024,0.942,0.034)	3.1085
	LM1	-0.0916	0.5902	(0.019,0.953,0.028)	3.0319
	LM2	-0.0916	0.5902	(0.020,0.950,0.030)	2.9898
	LM3	-0.0916	0.5902	(0.023,0.945,0.032)	2.9383
	HQF1	-0.0889	0.5887	(0.021,0.945,0.034)	2.9832
	HQF2	-0.0889	0.5887	(0.022,0.946,0.032)	2.9383

Table 2.7: Comparisons with Model Misspecifications: (I) and (I*)

Model	Kernel	Methods	Bias	MSE	(L, CP, U)	AL	Power
(I)	Flat	NB	-0.0010	0.0312	(0.030,0.940,0.030)	0.6648	0.432
		WY	-0.0010	0.0312	(0.032,0.940,0.028)	0.6684	0.436
		LM1	0.0096	0.0064	(0.021,0.958,0.021)	0.3262	0.953
		LM2	0.0096	0.0064	(0.018,0.961,0.021)	0.3234	0.956
		LM3	0.0096	0.0064	(0.017,0.958,0.025)	0.3228	0.955
		HQF1	0.0122	0.0101	(0.022,0.955,0.023)	0.4026	0.828
		HQF2	0.0122	0.0101	(0.020,0.958,0.022)	0.4033	0.831
		KM1	-0.0126	0.0124	(0.029,0.948,0.023)	0.4239	0.780
		KM2	-0.0126	0.0124	(0.028,0.945,0.027)	0.4217	0.776
	Epan	NB	-0.0010	0.0312	(0.030,0.940,0.030)	0.6648	0.432
		WY	-0.0010	0.0312	(0.032,0.940,0.028)	0.6684	0.436
		LM1	0.0096	0.0064	(0.021,0.958,0.021)	0.3262	0.953
		LM2	0.0096	0.0064	(0.018,0.961,0.021)	0.3234	0.956
		LM3	0.0096	0.0064	(0.017,0.958,0.025)	0.3228	0.955
		HQF1	0.0122	0.0101	(0.022,0.955,0.023)	0.4026	0.828
		HQF2	0.0122	0.0101	(0.020,0.958,0.022)	0.4033	0.831
		KM1	-0.0096	0.0120	(0.029,0.947,0.024)	0.4170	0.794
		KM2	-0.0096	0.0120	(0.027,0.948,0.025)	0.4152	0.788
(I*)	Flat	NB	0.0111	0.0743	(0.026,0.956,0.018)	1.1015	0.182
		WY	0.0111	0.0743	(0.026,0.957,0.017)	1.0987	0.184
		LM1	-0.5028	0.0278	(0.000,0.458,0.542)	0.2841	0.546
		LM2	-0.5028	0.0278	(0.000,0.446,0.554)	0.2812	0.554
		LM3	-0.5028	0.0278	(0.000,0.447,0.553)	0.2815	0.553
		HQF1	-0.0080	0.0198	(0.024,0.944,0.032)	0.5567	0.549
		HQF2	-0.0080	0.0198	(0.025,0.945,0.030)	0.5554	0.566
		KM1	-0.0040	0.0102	(0.028,0.949,0.023)	0.3907	0.840
		KM2	-0.0040	0.0102	(0.028,0.949,0.023)	0.3912	0.839
	Epan	NB	0.0111	0.0743	(0.026,0.956,0.018)	1.1015	0.182
		WY	0.0111	0.0743	(0.026,0.957,0.017)	1.0987	0.184
		LM1	-0.5028	0.0278	(0.000,0.458,0.542)	0.2841	0.546
		LM2	-0.5028	0.0278	(0.000,0.446,0.554)	0.2812	0.554
		LM3	-0.5028	0.0278	(0.000,0.447,0.553)	0.2815	0.553
		HQF1	-0.0080	0.0198	(0.024,0.944,0.032)	0.5567	0.549
		HQF2	-0.0080	0.0198	(0.025,0.945,0.030)	0.5554	0.566
		KM1	-0.0028	0.0101	(0.027,0.950,0.023)	0.3876	0.841
		KM2	-0.0028	0.0101	(0.027,0.950,0.023)	0.3881	0.840

Table 2.8: Comparisons with Model Misspecifications: (II) and (II*)

Model	Kernel	Methods	Bias	MSE	(L, CP, U)	AL	Power
(II)	Flat	NB	0.0068	0.0296	(0.031,0.954,0.015)	0.6652	0.429
		WY	0.0068	0.0296	(0.029,0.956,0.015)	0.6665	0.427
		LM1	-0.0068	0.0108	(0.018,0.955,0.027)	0.4150	0.804
		LM2	-0.0068	0.0108	(0.017,0.958,0.025)	0.4126	0.809
		LM3	-0.0068	0.0108	(0.017,0.959,0.024)	0.4120	0.814
		HQF1	-0.0063	0.0107	(0.016,0.958,0.026)	0.4129	0.808
		HQF2	-0.0063	0.0107	(0.017,0.957,0.026)	0.4120	0.816
		KM1	0.0220	0.0151	(0.032,0.950,0.018)	0.4736	0.711
		KM2	0.0220	0.0151	(0.033,0.950,0.017)	0.4724	0.715
	Epan	NB	0.0068	0.0296	(0.031,0.954,0.015)	0.6652	0.429
		WY	0.0068	0.0296	(0.029,0.956,0.015)	0.6665	0.427
		LM1	-0.0068	0.0108	(0.018,0.955,0.027)	0.4150	0.804
		LM2	-0.0068	0.0108	(0.017,0.958,0.025)	0.4126	0.809
		LM3	-0.0068	0.0108	(0.017,0.959,0.024)	0.4120	0.814
		HQF1	-0.0063	0.0107	(0.016,0.958,0.026)	0.4129	0.808
		HQF2	-0.0063	0.0107	(0.017,0.957,0.026)	0.4120	0.816
		KM1	0.0223	0.0156	(0.032,0.947,0.021)	0.4705	0.713
		KM2	0.0223	0.0156	(0.032,0.948,0.020)	0.4693	0.719
(II*)	Flat	NB	-0.0484	0.0725	(0.028,0.944,0.028)	1.0587	0.169
		WY	-0.0484	0.0725	(0.028,0.945,0.027)	1.0568	0.169
		LM1	-0.0330	0.0596	(0.029,0.951,0.020)	0.9443	0.221
		LM2	-0.0330	0.0596	(0.032,0.945,0.023)	0.9444	0.222
		LM3	-0.0330	0.0596	(0.033,0.943,0.024)	0.9414	0.222
		HQF1	-0.0336	0.0593	(0.035,0.943,0.022)	0.9428	0.220
		HQF2	-0.0336	0.0593	(0.033,0.944,0.023)	0.9414	0.221
		KM1	-0.0604	0.0148	(0.024,0.942,0.034)	0.4598	0.656
		KM2	-0.0604	0.0148	(0.023,0.944,0.033)	0.4603	0.655
	Epan	NB	-0.0484	0.0725	(0.028,0.944,0.028)	1.0587	0.169
		WY	-0.0484	0.0725	(0.028,0.945,0.027)	1.0568	0.169
		LM1	-0.0330	0.0596	(0.029,0.951,0.020)	0.9443	0.221
		LM2	-0.0330	0.0596	(0.032,0.945,0.023)	0.9444	0.222
		LM3	-0.0330	0.0596	(0.033,0.943,0.024)	0.9414	0.222
		HQF1	-0.0336	0.0593	(0.035,0.943,0.022)	0.9428	0.220
		HQF2	-0.0336	0.0593	(0.033,0.944,0.023)	0.9414	0.221
		KM1	-0.0974	0.0168	(0.024,0.953,0.034)	0.4934	0.563
		KM2	-0.0974	0.0168	(0.023,0.954,0.033)	0.4934	0.563

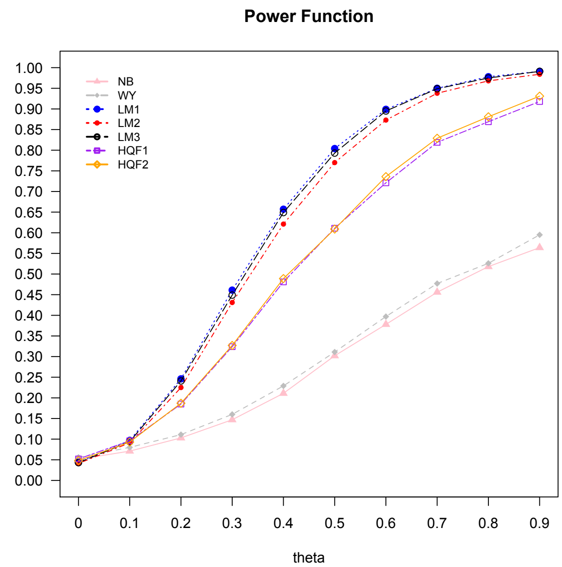

Figure 2.1: Power Function of Testing $H_0 : \theta = 0$ under Model (I), $\rho = 0.8$, $(n_1, n_0) = (50, 50)$

2.7 A Real Data Analysis

The AIDS Clinical Trials Group (ACTG) protocol 175 (Hammer et al. (1996)) was a randomized double-blinded clinical trial comparing monotherapy (with zidovudine or didanosine) and combination therapy (with zidovudine plus either didanosine or zalcitabine) in HIV-I infected subjects whose CD4 cell counts were between 200 to 500 per cubic millimetre. There are 2139 individuals randomly assigned to one of the four regimens: zidovudine monotherapy, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine monotherapy. We are interested in comparing the CD4 counts at 20 ± 5 weeks after randomization between subjects who received zidovudine monotherapy (control) and those who received one of the other three therapies (treatment). The pretest response is the CD4 counts at baseline. Following Leon et al. (2003), the other baseline covariates used in the data analysis are weight, history of intravenous drug use, Karnofsky score (a

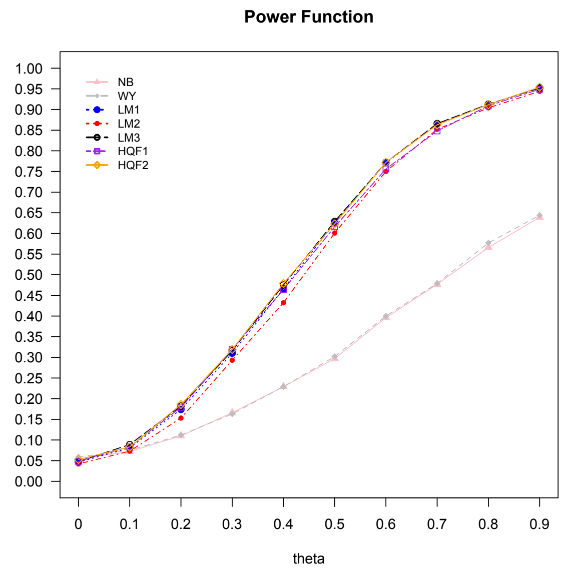

Figure 2.2: Power Function of Testing $H_0 : \theta = 0$ under Model (II), $\rho = 0.8$, $(n_1, n_0) = (50, 50)$

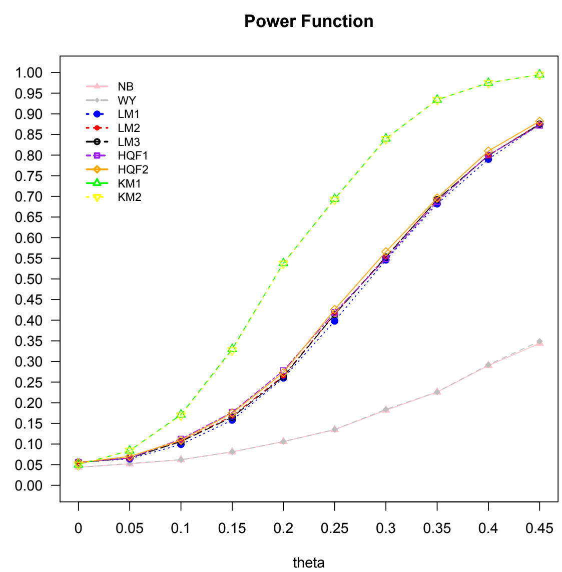

Figure 2.3: Power Function of Testing $H_0 : \theta = 0$ under Model (I*), $(n_1, n_0) = (200, 200)$

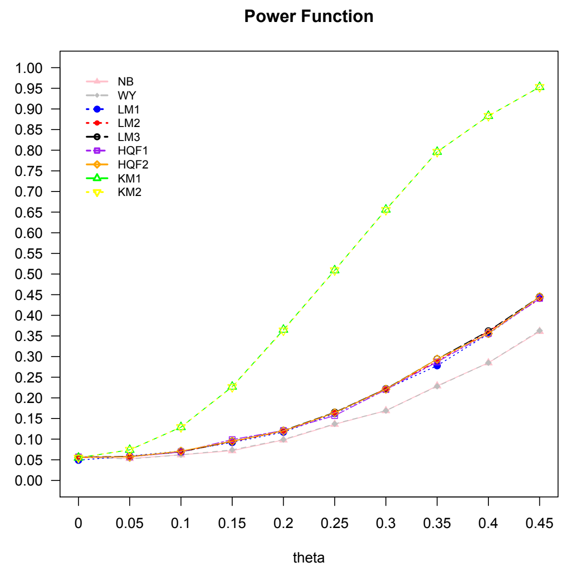

Figure 2.4: Power Function of Testing $H_0 : \theta = 0$ under Model (II*), $(n_1, n_0) = (200, 200)$

categorical variable on a scale of 0-100), number of days of previously received antiretroviral therapy, and the symptomatic indicator. In our analysis, responses of the Karnofsky score are dichotomized into two groups: "Karnofsky = 100" or "not"; and the responses of the number of days of previously received antiretroviral therapy are dichotomized into two groups: "no pre anti therapy" or "some pre anti therapy". All the baseline covariates and the pretest response are standardized.

We applied all the methods that we compared in the simulation studies in the real data analysis. We fit linear models to the observed data in each treatment group by using ordinary least squares for the LM methods and the HQF methods. For the KM method, we considered the product kernel function $K(u_1, u_2, u_3, u_4, u_5, u_6) = \prod_{k=1}^{6} K_1(u_k)$, where $K_1(u)$ is the flat kernel function: $K_1(u) = 1/2$, $|u| \leq 1$. We set the truncation threshold $b_n = 0.00001$ and selected the bandwidth $h_n = 1.49$ using a cross validation method. The results are summarized in Table 2.9. The LM, HQF and KM methods provide similar estimated treatment effect $\hat{\theta}$. Compared to the LM and HQF methods, the KM methods

Table 2.9: Treatment effect estimates for 20 ± 5 weeks post randomization CD4 counts for ACTG 175

Methods	$\hat{\theta}$	95% CI of θ
NB	46.8105	(33.5605, 60.0605)
WY	46.8105	(33.4686, 59.9845)
LM1	49.0076	(38.6798, 59.4406)
LM2	49.0076	(38.9386, 59.1766)
LM3	49.0076	(38.8981, 59.1171)
HQF1	49.0210	(39.2862, 59.8098)
HQF2	49.0210	(38.9115, 59.1305)
KM1	50.6940	(41.1914, 60.3510)
KM2	50.6940	(40.1255, 60.2626)

yield slightly narrower confidence intervals.

2.8 Concluding Remarks

In this chapter, we developed imputation-based empirical likelihood methods for pretest-posttest studies. Our primary goal was to construct confidence intervals or conduct hypothesis tests on the treatment effect using the empirical likelihood ratio statistic. The proposed methods are most efficient when linear regression models adequately describe the relations between the posttest responses and the pretest baseline measures. Kernel regression methods provide a robust alternative approach against possible model misspecifications, and they are practically useful when there are only a few baseline variables with good prediction power. We also derived the explicit asymptotic variance formulas for the imputation-based estimators as well as the estimator proposed by Huang et al. (2008). Our proposed methods can be extended to cover other parameters of interest, such as the distribution functions of the posttest responses. When the subjects are selected from a finite population using a complex survey design, the sampling features need to be taken into account for inferences. These latter problems will be discussed in the following chapters.

2.9 Proofs and Regularity Conditions

2.9.1 Lemmas

Lemma 1. *Under the assumptions of Theorem 1 and models (2.4) and (2.5), we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{Y}_{1i}-\mu_1) \overset{d}{\to} N(0,V_1) \quad \textit{and} \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{Y}_{0i}-\mu_0) \overset{d}{\to} N(0,V_0)\,,$$

where $\mu_1 = E(Y_1)$*,* $\mu_0 = E(Y_0)$*, and*

$$\begin{aligned} V_1 &= \delta\sigma_{\epsilon 1}^2 + \boldsymbol{\beta}_1^T\Sigma_Z\boldsymbol{\beta}_1 + \{(1-\delta)^2\delta^{-1}\sigma_{\epsilon 1}^2 + 2(1-\delta)\sigma_{\epsilon 1}^2\}E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z})\,, \\ V_0 &= (1-\delta)\sigma_{\epsilon 0}^2 + \boldsymbol{\beta}_0^T\Sigma_Z\boldsymbol{\beta}_0 + \{\delta^2(1-\delta)^{-1}\sigma_{\epsilon 0}^2 + 2\delta\sigma_{\epsilon 0}^2\}E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z})\,. \end{aligned}$$

Proof. We prove the result involving μ_1. Write $n^{-1/2}\sum_{i=1}^{n}(\tilde{Y}_{1i}-\mu_1) = T_{n1}+T_{n2}+T_{n3}$, where

$$\begin{aligned} T_{n1} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_i\{Y_{1i}-\boldsymbol{Z}_i^T\boldsymbol{\beta}_1)\}\,, \\ T_{n2} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-R_i)\{\boldsymbol{Z}_i^T(\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1)\}\,, \\ T_{n3} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_1-\mu_1\}. \end{aligned}$$

By the central limit theorem, we have

$$T_{n1} \overset{d}{\to} N(0, \delta E[(Y_{1i}-\boldsymbol{Z}_i^T\boldsymbol{\beta}_1)^2]),$$

and

$$T_{3n} \overset{d}{\to} N(0, Var(\boldsymbol{Z}_i^T\boldsymbol{\beta}_1)).$$

Noting that $\hat{\boldsymbol{\beta}}_1 = \left(\sum_{i=1}^n R_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T\right)^{-1} \sum_{i=1}^n R_i \boldsymbol{Z}_i Y_{1i}$, we have

$$
\begin{aligned}
T_{2n} &= \Big(\frac{1}{n}\sum_{i=1}^n (1-R_i)\boldsymbol{Z}_i^T\Big)\Big\{\frac{1}{n}\sum_{i=1}^n R_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T\Big\}^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n R_i \boldsymbol{Z}_i \epsilon_{1i} \\
&= (1-\delta)\delta^{-1}E(\boldsymbol{Z}^T)\big(E[\boldsymbol{Z}\boldsymbol{Z}^T]\big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n R_i \boldsymbol{Z}_i \epsilon_{1i} + o_p(1) \\
&\stackrel{d}{\to} N(0, (1-\delta)^2\delta^{-1}\sigma_{\epsilon 1}^2 E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z}))\,.
\end{aligned}
$$

since

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n R_i \boldsymbol{Z}_i \epsilon_{1i} &\stackrel{d}{\to} N(0, Var(R_i \boldsymbol{Z}_i \epsilon_{1i})) \\
&= N(0, E(R_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T \epsilon_{1i}^2)) \\
&= N(0, \delta E(\boldsymbol{Z}\boldsymbol{Z}^T)\sigma_{\epsilon 1}^2)\,.
\end{aligned}
$$

Noting that $V_1 = AV(T_{n1} + T_{n2} + T_{n3})$, $Cov(T_{n1}, T_{n3}) = 0$, $Cov(T_{n2}, T_{n3}) = 0$ and

$$
Cov(T_{n1}, T_{n2}) = (1-\delta)\sigma_{\epsilon 1}^2 E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z})\,,
$$

therefore,

$$
\begin{aligned}
V_1 &= \delta E[(Y_{1i} - \boldsymbol{Z}_i^T\boldsymbol{\beta}_1)^2] + Var(\boldsymbol{Z}_i^T\boldsymbol{\beta}_1) + (1-\delta)^2\delta^{-1}E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z})\sigma_{\epsilon 1}^2 + \\
&\quad 2(1-\delta)E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z})\sigma_{\epsilon 1}^2 \\
&= \delta\sigma_{\epsilon 1}^2 + \boldsymbol{\beta}_1^T\Sigma_Z\boldsymbol{\beta}_1 + \{(1-\delta)^2\delta^{-1}\sigma_{\epsilon 1}^2 + 2(1-\delta)\sigma_{\epsilon 1}^2\}E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z})\,.
\end{aligned}
$$

□

Lemma 2. *Under the conditions of Theorem 1 and models (2.4) and (2.5), we have*

$$
\begin{aligned}
\tilde{V}_1 &= \frac{1}{n}\sum_{i=1}^n (\tilde{Y}_{1i} - \mu_1)^2 = \delta\sigma_{\epsilon 1}^2 + \boldsymbol{\beta}_1^T\Sigma_Z\boldsymbol{\beta}_1 + o_p(1), \\
\tilde{V}_0 &= \frac{1}{n}\sum_{i=1}^n (\tilde{Y}_{0i} - \mu_0)^2 = (1-\delta)\sigma_{\epsilon 0}^2 + \boldsymbol{\beta}_0^T\Sigma_Z\boldsymbol{\beta}_0 + o_p(1).
\end{aligned}
$$

Proof. We prove the result for $\tilde{V}_1$. Noting that $\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 = o_p(1)$, we can write

$$\tilde{Y}_{1i} - \mu_1 = R_i(Y_{1i} - \boldsymbol{Z}_i^T \boldsymbol{\beta}_1) + (1 - R_i)\boldsymbol{Z}_i^T(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + (\boldsymbol{Z}_i^T \boldsymbol{\beta}_1 - \mu_1)\,,$$

which leads to $\tilde{V}_1 = R_{n1} + R_{n2} + R_{n3} + o_p(1)$ where

$$\begin{aligned} R_{n1} &= \frac{1}{n}\sum_{i=1}^{n} R_i(Y_{1i} - \boldsymbol{Z}_i^T \boldsymbol{\beta}_1)^2\,, \\ R_{n2} &= \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{Z}_i^T \boldsymbol{\beta}_1 - \mu_1)^2\,, \\ R_{n3} &= \frac{2}{n}\sum_{i=1}^{n} R_i(Y_{1i} - \boldsymbol{Z}_i^T \boldsymbol{\beta}_1)(\boldsymbol{Z}_i^T \boldsymbol{\beta}_1 - \mu_1)\,. \end{aligned}$$

By the law of large numbers, we have $R_{n1} \xrightarrow{P} \delta E\{(Y_{1i} - \boldsymbol{Z}_i^T \boldsymbol{\beta}_1)^2\}$, $R_{n2} \xrightarrow{P} Var(\boldsymbol{Z}_i^T \boldsymbol{\beta}_1)$, and $R_{n3} \xrightarrow{P} 2E\{R_i(Y_{1i} - \boldsymbol{Z}_i^T \boldsymbol{\beta}_1)(\boldsymbol{Z}_i^T \boldsymbol{\beta}_1 - \mu_1)\}$, where "$\xrightarrow{P}$" means convergence in probability. Therefore,

$$\begin{aligned} R_{n1} &= \delta\sigma_{\epsilon 1}^2 + o_p(1)\,, \\ R_{n2} &= \boldsymbol{\beta}_1^T \Sigma_Z \boldsymbol{\beta}_1 + o_p(1)\,, \quad \text{and} \\ R_{n3} &= o_p(1)\,. \end{aligned}$$

It follows that $\tilde{V}_1 = \delta\sigma_{\epsilon 1}^2 + \boldsymbol{\beta}_1^T \Sigma_Z \boldsymbol{\beta}_1 + o_p(1)$. □

Lemma 3. *Under the assumptions of Theorem 1 and models (2.4) and (2.5),* $\max_{1 \le i \le n} |\tilde{Y}_{ji}| = o_p(\sqrt{n})$, $j = 0, 1$.

Proof. We have $\max_{1 \le i \le n} |Y_{ji}| = o_p(\sqrt{n})$ and $\max_{1 \le i \le n} \|\boldsymbol{Z}_i\| = o_p(\sqrt{n})$ under the assumed conditions (Lemma 11.2 of Owen, 2001). The results follow from

$$\max_{1 \le i \le n} |\tilde{Y}_{ji}| \le \max_{1 \le i \le n} |Y_{ji}| + \max_{1 \le i \le n} \|\boldsymbol{Z}_i\|^T \hat{\boldsymbol{\beta}}_j$$

and the fact that $\hat{\boldsymbol{\beta}}_j = O_p(1)$ for $j = 0, 1$. □

2.9.2 Proof of Theorem 1

Proof. To facilitate asymptotic derivations, we introduce a nuisance parameter $\mu = \sum_{i=1}^n q_i \tilde{Y}_{0i}$ and rewrite constraint (2.7) as

$$\sum_{i=1}^n p_i \tilde{Y}_{1i} = \mu + \theta \quad \text{and} \quad \sum_{i=1}^n q_i \tilde{Y}_{0i} = \mu\,.$$

For fixed values of θ and μ, the solutions to the constrained maximization problem are given by

$$\hat{p}_i = \frac{1}{n[1+\lambda_1(\tilde{Y}_{1i} - \mu - \theta)]} \quad \text{and} \quad \hat{q}_i = \frac{1}{n[1+\lambda_0(\tilde{Y}_{0i} - \mu)]},$$

where the Lagrange multipliers λ_1 and λ_0 are the solutions to

$$\frac{1}{n}\sum_{i=1}^n \frac{\tilde{Y}_{1i} - \mu - \theta}{1+\lambda_1(\tilde{Y}_{1i} - \mu - \theta)} = 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \frac{\tilde{Y}_{0i} - \mu}{1+\lambda_0(\tilde{Y}_{0i} - \mu)} = 0\,.$$

Let $r(\theta, \mu)$ be the empirical log-likelihood ratio statistic on (θ, μ). We have

$$\begin{aligned} r(\theta,\mu) &= \sum_{i=1}^n \log(n\hat{p}_i) + \sum_{i=1}^n \log(n\hat{q}_i) \\ &= -\sum_{i=1}^n \log[1+\lambda_1(\tilde{Y}_{1i} - \mu - \theta)] - \sum_{i=1}^n \log[1+\lambda_0(\tilde{Y}_{0i} - \mu)]. \end{aligned}$$

Let $\hat{\mu} = \hat{\mu}(\theta)$ be the maximizer of $r(\theta, \mu)$ for a given θ, which can be obtained through profiling. The solution is obtained by setting

$$\frac{\partial r(\theta,\mu)}{\partial \mu} = \sum_{i=1}^n \frac{\lambda_1}{1+\lambda_1(\tilde{Y}_{1i} - \mu - \theta)]} + \sum_{i=1}^n \frac{\lambda_0}{1+\lambda_0(\tilde{Y}_{0i} - \mu)]} = 0\,. \tag{2.17}$$

Note that λ_1 and λ_0 both depend on μ but (2.17) holds due to the fact that

$$\sum_{i=1}^n \frac{(\partial\lambda_1/\partial\mu)(\tilde{Y}_{1i} - \mu - \theta)}{1+\lambda_1(\tilde{Y}_{1i} - \mu - \theta)]} + \sum_{i=1}^n \frac{(\partial\lambda_0/\partial\mu)(\tilde{Y}_{0i} - \mu - \theta)}{1+\lambda_0(\tilde{Y}_{0i} - \mu)]} = 0\,.$$

It follows from (2.17) that $n(\lambda_1 + \lambda_0) = 0$, which leads to $\lambda_1 = -\lambda_0$. Without loss of generality, we only need to consider those θ and μ such that $\theta = E(Y_1) - E(Y_0) + O(n^{-1/2})$

and $\mu = E(Y_0) + O(n^{-1/2})$. By the Taylor series expansion, the empirical loglikelihood ratio statistic can be written as

$$\begin{aligned} r(\theta, \mu) &= -\sum_{i=1}^{n} \log\left[1 + \lambda_1(\tilde{Y}_{1i} - \mu - \theta)\right] - \sum_{i=1}^{n} \log\left[1 + \lambda_0(\tilde{Y}_{0i} - \mu)\right] \\ &= -\sum_{i=1}^{n} \left\{ \lambda_1(\tilde{Y}_{1i} - \mu - \theta) - \frac{1}{2}(\lambda_1(\tilde{Y}_{1i} - \mu - \theta))^2 \right\} + \gamma_{1n} \\ &\quad - \sum_{i=1}^{n} \left\{ \lambda_0(\tilde{Y}_{0i} - \mu) - \frac{1}{2}(\lambda_0(\tilde{Y}_{0i} - \mu))^2 \right\} + \gamma_{0n}, \end{aligned}$$

with

$$|\gamma_{1n}| \le d_1 \sum_{i=1}^{n} |\lambda_1(\tilde{Y}_{1i} - \mu - \theta)|^3 \ \text{ in probability}$$

$$|\gamma_{0n}| \le d_2 \sum_{i=1}^{n} |\lambda_0(\tilde{Y}_{0i} - \mu)|^3 \ \text{ in probability,}$$

where d_1 and d_2 are positive constants. It follows from Lemmas 3 and standard arguments from Owen (2001), that $|\gamma_{1n}| \le o_p(1)$ and $|\gamma_{0n}| \le o_p(1)$.

Meanwhile, since

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^{n} \frac{\tilde{Y}_{1i} - \mu - \theta}{1 + \lambda_1(\tilde{Y}_{1i} - \mu - \theta)} \\ &= \frac{1}{n} \sum_{i=1}^{n} (\tilde{Y}_{1i} - \mu - \theta) \left[1 - \lambda_1(\tilde{Y}_{1i} - \mu - \theta) + \frac{\lambda_1^2(\tilde{Y}_{1i} - \mu - \theta)^2}{1 + \lambda_1(\tilde{Y}_{1i} - \mu - \theta)} \right], \end{aligned}$$

and

$$n^{-1} \sum_{i=1}^{n} |\tilde{Y}_{1i} - \mu - \theta|^3 |\lambda_1|^2 |1 + (\tilde{Y}_{1i} - \mu - \theta)\lambda_1|^{-1} = o_p(n^{1/2}) O_p(n^{-1}) O_p(1) = o_p(n^{-1/2}),$$

then we have

$$\lambda_1 = \tilde{V}_1^{-1} \frac{1}{n} \sum_{i=1}^{n} (\tilde{Y}_{1i} - \mu - \theta) + o_p(n^{-1/2}), \tag{2.18}$$

where $\tilde{V}_1 = n^{-1}\sum_{i=1}^n(\tilde{Y}_{1i} - \mu_1)^2$.

Similarly,

$$\lambda_0 \quad = \quad \tilde{V}_0^{-1}\frac{1}{n}\sum_{i=1}^n(\tilde{Y}_{0i} - \mu) + o_p(n^{-1/2})\,, \tag{2.19}$$

where $\tilde{V}_0 = n^{-1}\sum_{i=1}^n(\tilde{Y}_{0i} - \mu_0)^2$. The profile solution $\hat{\mu} = \hat{\mu}(\theta)$, which satisfies $\lambda_1 = -\lambda_0$, has the following asymptotic representation:

$$\hat{\mu} = \nu(\bar{\tilde{Y}}_1 - \theta) + (1-\nu)\bar{\tilde{Y}}_0 + o_p(n^{-1/2})\,, \tag{2.20}$$

where $\nu = \tilde{V}_1^{-1}[\tilde{V}_0^{-1} + \tilde{V}_1^{-1}]^{-1}$, $\bar{\tilde{Y}}_1 = n^{-1}\sum_{i=1}^n \tilde{Y}_{1i}$ and $\bar{\tilde{Y}}_0 = n^{-1}\sum_{i=1}^n \tilde{Y}_{0i}$.

Moreover, note that since

$$\begin{aligned} 0 \quad &= \quad \frac{1}{n}\sum_{i=1}^n \frac{\lambda_1(\tilde{Y}_{1i} - \mu - \theta)}{1 + \lambda_1(\tilde{Y}_{1i} - \mu - \theta)} \\ &= \quad \frac{1}{n}\sum_{i=1}^n \lambda_1(\tilde{Y}_{1i} - \mu - \theta)\left[1 - \lambda_1(\tilde{Y}_{1i} - \mu - \theta) + \frac{\lambda_1^2(\tilde{Y}_{1i} - \mu - \theta)^2}{1 + \lambda_1(\tilde{Y}_{1i} - \mu - \theta)}\right], \end{aligned}$$

and

$$n^{-1}\sum_{i=1}^n |\tilde{Y}_{1i} - \mu - \theta|^3|\lambda_1|^3|1 + (\tilde{Y}_{1i} - \mu - \theta)\lambda_1|^{-1} = o_p(n^{1/2})O_p(n^{-3/2})O_p(1) = o_p(n^{-1}),$$

i.e.,

$$\sum_{i=1}^n |\lambda_1|^3|\tilde{Y}_{1i} - \mu - \theta|^3|1 + (\tilde{Y}_{1i} - \mu - \theta)\lambda_1|^{-1} = o_p(1),$$

then,

$$\sum_{i=1}^n \lambda_1(\tilde{Y}_{1i} - \mu - \theta) = \sum_{i=1}^n \lambda_1^2(\tilde{Y}_{1i} - \mu - \theta)^2 + o_p(1). \tag{2.21}$$

Likewise,

$$\sum_{i=1}^n \lambda_0(\tilde{Y}_{0i} - \mu) = \sum_{i=1}^n \lambda_0^2(\tilde{Y}_{0i} - \mu)^2 + o_p(1). \tag{2.22}$$

The empirical likelihood ratio statistic on the parameter of interest, θ, is given by $r(\theta) = r(\theta, \hat{\mu}(\theta))$. Using the asymptotic representations (2.18), (2.19), (2.20), (2.21), and (2.22), we have

$$\begin{aligned}
-2r(\theta) &= 2\sum_{i=1}^{n}\Big\{\lambda_1(\tilde{Y}_{1i} - \hat{\mu} - \theta) - \frac{1}{2}(\lambda_1(\tilde{Y}_{1i} - \hat{\mu} - \theta))^2\Big\} + \\
&\quad 2\sum_{i=1}^{n}\Big\{\lambda_0(\tilde{Y}_{0i} - \hat{\mu}) - \frac{1}{2}(\lambda_0(\tilde{Y}_{0i} - \hat{\mu}))^2\Big\} + o_p(1) \\
&= \sum_{i=1}^{n}\lambda_1(\tilde{Y}_{1i} - \hat{\mu} - \theta) + \sum_{i=1}^{n}\lambda_0(\tilde{Y}_{0i} - \hat{\mu}) + o_p(1) \\
&= \tilde{V}_1^{-1}n^{-1}\Big\{\sum_{i=1}^{n}(\tilde{Y}_{1i} - \hat{\mu} - \theta)\Big\}^2 + \tilde{V}_0^{-1}n^{-1}\Big\{\sum_{i=1}^{n}(\tilde{Y}_{0i} - \hat{\mu})\Big\}^2 + o_p(1) \\
&= \tilde{V}_1^{-1}n(1-\nu)^2(\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta)^2 + \tilde{V}_0^{-1}n\nu^2(\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta)^2 + o_p(1) \\
&= \frac{n}{\tilde{V}_1 + \tilde{V}_0}(\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta)^2 + o_p(1)\,.
\end{aligned}$$

The asymptotic variance of $\sqrt{n}(\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta)$ is given by:

$$\begin{aligned}
V &= AV(\sqrt{n}(\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0)) \\
&= \delta\sigma_{\epsilon1}^2 + \boldsymbol{\beta}_1^T\Sigma_Z\boldsymbol{\beta}_1 + \{(1-\delta)^2\delta^{-1}\sigma_{\epsilon1}^2 + 2(1-\delta)\sigma_{\epsilon1}^2\}E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z}) + \\
&\quad (1-\delta)\sigma_{\epsilon0}^2 + \boldsymbol{\beta}_0^T\Sigma_Z\boldsymbol{\beta}_0 + \{\delta^2(1-\delta)^{-1}\sigma_{\epsilon0}^2 + 2\delta\sigma_{\epsilon0}^2\}E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z}) - \\
&\quad 2E\{(\boldsymbol{Z}_i^T\boldsymbol{\beta}_1 - \mu_1)(\boldsymbol{Z}_i^T\boldsymbol{\beta}_0 - \mu_0)\}\,, \\
&= Var(\boldsymbol{Z}_i^T\boldsymbol{\beta}_1 - \boldsymbol{Z}_i^T\boldsymbol{\beta}_0) + \delta\sigma_{\epsilon1}^2 + (1-\delta)\sigma_{\epsilon0}^2 + \\
&\quad \{(1-\delta)^2\delta^{-1}\sigma_{\epsilon1}^2 + 2(1-\delta)\sigma_{\epsilon1}^2 + \delta^2(1-\delta)^{-1}\sigma_{\epsilon0}^2 + 2\delta\sigma_{\epsilon0}^2\}E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z})\,, \\
&= (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T\Sigma_Z(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + \{\delta + (1-\delta)^2\delta^{-1} + 2(1-\delta)\}\sigma_{\epsilon1}^2 + \\
&\quad \{(1-\delta) + \delta^2(1-\delta)^{-1} + 2\delta\}\sigma_{\epsilon0}^2\,, \\
&= (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T\Sigma_Z(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + \delta^{-1}\sigma_{\epsilon1}^2 + (1-\delta)^{-1}\sigma_{\epsilon0}^2\,.
\end{aligned}$$

To derive a closed form expression for V, we note that $E(\boldsymbol{Z}^T)\{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z}) = 1$ if we assume two linear regression models (2.4) and (2.5) both include an intercept, i.e., the vector $\boldsymbol{Z}$ has "1" as its first component.

Let

$$c_1 = \{(\tilde{V}_1 + \tilde{V}_0)/V\}^{-1},$$

It follows from Lemma 1 that $-2r(\theta)/c_1 = n(\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta)^2/V + o_p(1)$ converges in distribution to a χ_1^2 variable when $\theta = E(Y_1) - E(Y_0) = \mu_1 - \mu_0$, $n \to \infty$ and $n_1/n \to \delta \in (0, 1)$. □

2.9.3 Regularity Conditions for Theorem 2

Let $\boldsymbol{Z}$ be a d-dimensional vector; let $f(\boldsymbol{z})$ be the probability density of $\boldsymbol{Z}$; let $m(\boldsymbol{z}) = E(Y|\boldsymbol{Z} = \boldsymbol{z})$, let $g(\boldsymbol{z}) = \delta f(\boldsymbol{z})$. We assume that the following conditions hold for both $Y = Y_1$ and $Y = Y_0$:

- C1. $f(\boldsymbol{z})$ has bounded partial derivatives up to order $k(> d)$ almost surely.
- C2. $m(\boldsymbol{z})$ has bounded partial derivatives up to order $k(> d)$ almost surely.
- C3. $E(Y^2) < \infty$.
- C4. $\sqrt{n} E\{(1 - R)|m(\boldsymbol{Z})| I(g(\boldsymbol{Z}) < b_n)\} \to 0$ as $n \to \infty$.
- C5. The kernel function K is bounded, with bounded support and a finite variance. Moreover, the order k of the kernel function is greater than d, where the order of a kernel is defined as the order of the first non-zero moment of the kernel.
- C6. $nh_n^{2d}(b_n^2 \wedge (\log\log n)^{-1}) \to \infty$, $nh_n^{2k}/b_n^2 \to 0$, and $h_n^k/b_n^2 \to 0$, as $n \to \infty$.

2.9.4 Proof of Theorem 2

Proof. As in the proof of Theorem 1, we define μ to be a fixed number depending on n such that $\mu = \mu_0 + o(n^{-1/2})$. We obtain empirical likelihood estimates for the p's and q's given by:

$$\hat{p_i} = \frac{1}{n[1+\lambda_1(\tilde{Y}_{1i}^{kel} - \mu - \theta)]} \quad \text{and} \quad \hat{q_i} = \frac{1}{n[1+\lambda_0(\tilde{Y}_{0i}^{kel} - \mu)]},$$

where λ_1 and λ_0 are the solutions to the following equations:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{Y}_{1i}^{kel} - \mu - \theta}{1+\lambda_1(\tilde{Y}_{1i}^{kel} - \mu - \theta)} = 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{Y}_{0i}^{kel} - \mu}{1+\lambda_0(\tilde{Y}_{0i}^{kel} - \mu)} = 0.$$

Let $r^{kel}(\theta,\mu)$ be the empirical log-likelihood ratio statistic. Then

$$r^{kel}(\theta,\mu) = -\sum_{i=1}^{n} \log\left[1+\lambda_1(\tilde{Y}_{1i}^{kel} - \mu - \theta)\right] - \sum_{i=1}^{n} \log\left[1+\lambda_0(\tilde{Y}_{0i}^{kel} - \mu)\right].$$

Similar arguments as in the proof of Theorem 1 will lead to

$$-2r^{kel}(\theta,\hat{\mu}) = \frac{n}{\tilde{V}_1^{kel} + \tilde{V}_0^{kel}}(\bar{\tilde{Y}}_1^{kel} - \bar{\tilde{Y}}_0^{kel} - \theta)^2 + o_p(1),$$

where $\bar{\tilde{Y}}_1^{kel}$ and $\bar{\tilde{Y}}_0^{kel}$ are given by:

$$\begin{aligned}
\bar{\tilde{Y}}_1^{kel} &= \frac{1}{n}\sum_{i=1}^{n}\{R_i Y_{1i} + (1-R_i)\hat{m}_{1b_n}(\boldsymbol{Z}_i)\} \\
\bar{\tilde{Y}}_0^{kel} &= \frac{1}{n}\sum_{i=1}^{n}\{(1-R_i) Y_{0i} + R_i\hat{m}_{0b_n}(\boldsymbol{Z}_i)\}.
\end{aligned}$$

From LEMMA A.1. of Wang and Rao (2002), we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{Y}_{1i}^{kel} - \mu_1) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m_1(\boldsymbol{Z}_i) - \mu_1) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_i(Y_{1i} - m_1(\boldsymbol{Z}_i)) + \\
&\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_i(Y_{1i} - m_1(\boldsymbol{Z}_i))\frac{1-\delta}{\delta} + o_p(n^{-1/2}), \\
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{Y}_{0i}^{kel} - \mu_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m_0(\boldsymbol{Z}_i) - \mu_0) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-R_i)(Y_{0i} - m_0(\boldsymbol{Z}_i)) + \\
&\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-R_i)(Y_{0i} - m_0(\boldsymbol{Z}_i))\frac{\delta}{1-\delta} + o_p(n^{-1/2}).
\end{aligned}$$

Moreover, if we let $\sigma_j^2(\boldsymbol{Z}) = Var(Y_j|\boldsymbol{Z})$ for $j = 0, 1$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{Y}_{1i}^{kel} - \mu_1) \overset{d}{\to} N(0, V_1^{kel}(\theta)),$$
$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{Y}_{0i}^{kel} - \mu_0) \overset{d}{\to} N(0, V_0^{kel}(\theta)),$$

where $V_1^{kel}(\theta) = E(\sigma_1^2(\boldsymbol{Z}))/\delta + Var(m_1(\boldsymbol{Z}))$ and $V_0^{kel}(\theta) = E(\sigma_0^2(\boldsymbol{Z}))/(1-\delta) + Var(m_0(\boldsymbol{Z}))$. It follows from Wang and Rao (2002) that

$$\begin{aligned}\sqrt{n}(\bar{\tilde{Y}}_1^{kel} - \bar{\tilde{Y}}_0^{kel} - \theta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{(m_1(\boldsymbol{Z}_i) - \mu_1) + \frac{R_i}{\delta}(Y_{1i} - m_1(\boldsymbol{Z}_i))\} - \\ &\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{(m_0(\boldsymbol{Z}_i) - \mu_0) + \frac{1-R_i}{1-\delta}(Y_{0i} - m_0(\boldsymbol{Z}_i))\} + o_p(n^{-1/2}).\end{aligned}$$

The asymptotic variance of $\sqrt{n}(\bar{\tilde{Y}}_1^{kel} - \bar{\tilde{Y}}_0^{kel} - \theta)$ is therefore given by

$$\begin{aligned}V^{kel} &= AV(\sqrt{n}(\bar{\tilde{Y}}_1^{kel} - \bar{\tilde{Y}}_0^{kel} - \theta)) \\ &= \frac{1}{\delta}E(\sigma_1^2(\boldsymbol{Z})) + Var(m_1(\boldsymbol{Z})) + \frac{1}{1-\delta}E(\sigma_0^2(\boldsymbol{Z})) + Var(m_0(\boldsymbol{Z})) - \\ &\quad 2E(m_1(\boldsymbol{Z})m_0(\boldsymbol{Z})) + 2\mu_1\mu_0 \\ &= \frac{1}{\delta}E(\sigma_1^2(\boldsymbol{Z})) + \frac{1}{1-\delta}E(\sigma_0^2(\boldsymbol{Z})) + Var(m_1(\boldsymbol{Z}) - m_0(\boldsymbol{Z})).\end{aligned}$$

Defining the scaling constant

$$c_2 = \{(\tilde{V}_1^{kel} + \tilde{V}_0^{kel})/V^{kel}\}^{-1}, \tag{2.23}$$

it then follows immediately that $-2r^{kel}(\theta)/c_2 = n(\bar{\tilde{Y}}_1^{kel} - \bar{\tilde{Y}}_0^{kel} - \theta)^2/V^{kel}$ converges to a χ_1^2 random variable as $n_1, n_0 \to \infty$ when θ is the true parameter. □

Chapter 3

Mann-Whitney Test with Empirical Likelihood Methods for Pretest-Posttest Studies

3.1 Introduction

The methods in Chapter 2 focus on making inferences of the treatment effect $\eta = \mu_1 - \mu_0$ for pretest-posttest studies, where $\mu_1 = E(Y_1)$ and $\mu_0 = E(Y_0)$. Another important research problem for pretest-posttest studies is to test the difference of the distribution functions between the treatment group and the control group. Let $S_1(t) = P(Y_1 > t)$ and the $S_0(t) = P(Y_0 > t)$ be respectively the survival functions of the (non-negative) response variables Y_1 and Y_0. We say Y_1 is stochastically larger than Y_0 if $S_1(t) > S_0(t)$ for all $t > 0$. The formal statistical inference problem is to test $H_0 : F_1 = F_0$ against $H_1 : F_1 < F_0$, where F_1 and F_0 are the cumulative distribution functions of Y_1 and Y_0, respectively. For two independent samples, the nonparametric Mann-Whitney test (Mann and Whitney (1947)) has been a popular choice for testing the difference of two distribution functions. Owen (2001) presented a two-sample EL formulation of the problem but the solution to the constrained maximization problem turns out to be extremely difficult to find. The computational complexity is created by the use of a U-statistic in forming the constraints. Jing et al. (2009) proposed a jackknife empirical likelihood (JEL) method for handling U-statistics. It replaces the complicated nonlinear constraint by a simple linear constraint through the so-called jackknife pseudo values, and hence results in a computationally friendly formulation of the constrained maximization problem.

In this chapter, our goal is to develop empirical likelihood based methods for the Mann-Whitney test for pretest-posttest studies, with the major focus on incorporating the two unique features of the sample data: (i) the availability of baseline information for both groups; and (ii) the structure of the data with missing by design. Our proposed methods combine the standard nonparametric Mann-Whitney test with the empirical likelihood method of Huang et al. (2008) (the HQF method), the imputation-based empirical likelihood method we developed in Chapter 2, and the jackknife empirical likelihood method of Jing et al. (2009). The rest of the chapter is organized as the following. In Section 3.2, we review the Mann-Whitney test and the two-sample EL formulation of the test with two independent samples. A jackknife EL formulation of the test is also presented. We propose the adjusted Mann-Whitney test with HQF estimators and discuss its asymptotic properties in Section 3.3. In Section 3.4, we look at the two-sample EL and the JEL methods for the situation where the pretest responses are included through imputation. In Section 3.5, we introduce a two-sample jackknife EL method. Finite sample performances of the proposed methods are evaluated through simulation studies and the results are reported in Section 3.6. Some concluding remarks are given in Section 3.7.

3.2 Methods for Testing the Difference of Distributions of Two Independent Samples

In this section, we first review the standard Mann-Whitney test for testing the difference of distributions of two independent samples. Moreover, we will review the two-sample EL method for estimating the probability of one random variable being stochastically larger than another in two independent samples by Owen (2001). Last but not least, we summarize the JEL method for the two-sample U-statistic by Jing et al. (2009).

3.2.1 Standard Mann-Whitney Test

Suppose we have two independent samples $\{Y_1, \cdots, Y_n\}$ and $\{X_1, \cdots, X_m\}$ having continuous cumulative distribution functions F and G respectively. We say the random variable Y is "stochastically larger than" the random variable X if $P(Y > a) > P(X > a)$ for every a, i.e. $F(a) < G(a)$ for every a. The objective is to test the null hypothesis $H_0 : F = G$ against the alternative $H_1 : F < G$. We can see that this alternative hypothesis is meaningful especially when we consider testing the effect of treatment on some measurement. For instance, we want to assess the effect of a certain treatment on the survival time of the patients. Longer survival time (larger survival function) means greater effectiveness of the treatment. Then, rejecting the null hypothesis in favor of the one-sided alternative indicates there is some effect of the treatment on elongating the survival time of the patients. In order to test the null hypothesis that the two samples are identically distributed, i.e. $F = G$, against the alternative that the distribution of the first sample is stochastically larger than the distribution of the second sample, i.e. $F < G$, we may consider the rank statistics. If the Y_i's are a sample from a stochastically larger distribution, then the ranks of the Y_i's in the pooled sample $\{Y_1, \cdots, Y_n, X_1, \cdots, X_m\}$ should be relatively large. Therefore, the measure of the size of the ranks can be used as a test statistic. Wilcoxon

(1945) first proposed a test statistic

$$W = \sum_{i=1}^{n} R_{Ni},$$

where $R_{N1}, \cdots, R_{Nn}$ are the ranks of $Y_1, \cdots, Y_n$ in the pooled sample. Larger values of the Wilcoxon statistic means rejecting the null hypothesis. In Mann and Whitney (1947), the author constructed a U type statistic which counts the number of times that $Y_i \geq X_j$. Mathematically, it is

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} I(Y_i \geq X_j).$$

The Mann-Whitney test statistic is closely related to the Wilcoxon statistic through the formula

$$W = \frac{n(n+1)}{2} + U.$$

Wilcoxon (1945) considered the case $n = m$ and tabulated 3 points of the distribution of W. In Mann and Whitney (1947), the authors showed the formulation of the $2r$th moment of U and proved the limit distribution of the standardized test statistic is normal under the null hypothesis. It has also been shown that the Mann-Whitney test is consistent. The Mann-Whitney test statistic can also be written as

$$MW = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} I(Y_i \geq X_j) = \int \hat{G}(a) d\hat{F}(a).$$

In this case, we have

$$\sqrt{12mn/(m+n+1)}(MW - 1/2) \xrightarrow{D} N(0,1)$$

under the null hypothesis (van der Vaart (1998)). Following by this result, a one-sided Mann-Whitney test rejects the null hypothesis when

$$\frac{MW - 1/2}{\sqrt{(m+n+1)/12nm}} > Z_{1-\alpha},$$

where α is the level of significance of the test and $Z_{1-\alpha}$ is the $(1-\alpha)\%$ quantile of a standard normal distribution. Similarly, a two-sided Mann-Whitney test, which tests $H_0 : F = G$ against $H_1 : F < G$ or $F > G$, rejects the null hypothesis if

$$\frac{|MW - 1/2|}{\sqrt{(m+n+1)/12nm}} > Z_{1-\alpha/2}.$$

3.2.2 Two-Sample Empirical Likelihood and Mann-Whitney Test

Owen (2001) considered a two-sample EL formulation for the Mann-Whitney test. Let $\{Y_1, \cdots, Y_n\}$ and $\{X_1, \cdots, X_m\}$ be two independent samples with marginal cdf F and G respectively. Let $h(Y, X, \theta) = I(Y \geq X) - \theta$. It follows that $E\{h(Y, X, \theta)\} = 0$ defines $\theta = P(Y \geq X)$, and $\theta = \theta_0 = 1/2$ under $H_0 : F = G$. Here are some assumptions for facilitating the arguments. We assume that $\min\{n, m\} \to \infty$, and that $0 < E(h(Y, X, \theta_0)^2)$, which must be true if the expectation is taken with respect to F and G and Y and X have overlapping supports. Here θ_0 is the true value of θ. We also assume either $E(E(h(Y, X, \theta_0)|Y)^2) > 0$ or $E(E(h(Y, X, \theta_0)|X)^2) > 0$.

We maximize the loglikelihood function

$$\ell = \sum_{i=1}^{n} \log p_i + \sum_{j=1}^{m} \log q_j,$$

subject to the constraints:

$$\sum_{i=1}^{n} p_i = \sum_{j=1}^{m} q_j = 1,$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j [I(Y_i \geq X_j) - \theta] = 0,$$

where $\boldsymbol{p} = (p_1, \cdots, p_n)$ and $\boldsymbol{q} = (q_1, \cdots, q_m)$'s are the discrete probability measures that F and G respectively put on over $\{Y_1, \cdots, Y_n\}$ and $\{X_1, \cdots, X_m\}$. Define $H_{ij}(\theta) = H_{ij} =$

$I(Y_i \geq X_j) - \theta$. Using the Lagrange multiplier method, we obtain the estimates for p_i and q_j:

$$\begin{aligned}
\hat{p}_i &= \frac{1}{n + \lambda \sum_{r=1}^{m} \hat{q}_r H_{ir}} = \frac{1}{n} \frac{1}{1 + \lambda \tilde{H}_{i\cdot}/n}, \\
\hat{q}_j &= \frac{1}{m + \lambda \sum_{s=1}^{n} \hat{p}_s H_{sj}} = \frac{1}{m} \frac{1}{1 + \lambda \tilde{H}_{\cdot j}/m},
\end{aligned}$$

where $\tilde{H}_{i\cdot} = \sum_{r=1}^{m} \hat{q}_r H_{ir}$, and $\tilde{H}_{\cdot j} = \sum_{s=1}^{n} \hat{p}_s H_{sj}$. The Lagrange multiplier λ is determined through

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \hat{p}_i \hat{q}_j [I(Y_i \geq X_j) - \theta] = \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{p}_i \hat{q}_j H_{ij} = 0. \tag{3.1}$$

Define the profile empirical loglikelihood ratio as

$$\begin{aligned}
r(\theta_0) &= \sum_{i=1}^{n} \log n\hat{p}_i(\theta_0) + \sum_{j=1}^{m} \log m\hat{q}_j(\theta_0) \\
&= -\left[\sum_{i=1}^{n} \log\left(1 + \frac{\lambda}{n}\tilde{H}_{i\cdot}\right) + \sum_{j=1}^{m} \log\left(1 + \frac{\lambda}{m}\tilde{H}_{\cdot j}\right) \right].
\end{aligned}$$

The following arguments by Owen (2001) sketch the proof that

$$-2r(\theta_0) \to \chi_1^2$$

as $n \to \infty$. Since $\hat{p}$ and $\hat{q}$ can be written as:

$$\begin{aligned}
\hat{p}_i &= \frac{1}{n} \frac{1}{1 + \lambda \tilde{H}_{i\cdot}/n} = \frac{1}{n} \left[1 - \left(\frac{\lambda}{n}\tilde{H}_{i\cdot}\right) + \left(\frac{\lambda}{n}\tilde{H}_{i\cdot}\right)^2 - \left(\frac{\lambda}{n}\tilde{H}_{i\cdot}\right)^3 + \cdots \right], \\
\hat{q}_i &= \frac{1}{m} \frac{1}{1 + \lambda \tilde{H}_{\cdot j}/m} = \frac{1}{m} \left[1 - \left(\frac{\lambda}{m}\tilde{H}_{\cdot j}\right) + \left(\frac{\lambda}{m}\tilde{H}_{\cdot j}\right)^2 - \left(\frac{\lambda}{m}\tilde{H}_{\cdot j}\right)^3 + \cdots \right],
\end{aligned}$$

when plugging $\hat{p}$ and $\hat{q}$ into (3.1), we have:

$$\begin{aligned}
0 &= \bar{H}_{\cdot\cdot} - \lambda \left[\frac{1}{n^2 m} \sum_{i=1}^{n} \sum_{j=1}^{m} H_{ij} \tilde{H}_{i\cdot} + \frac{1}{nm^2} \sum_{i=1}^{n} \sum_{j=1}^{m} H_{ij} \tilde{H}_{\cdot j} \right] \\
&\quad + \lambda^2 \left[\frac{1}{n^3 m} \sum_{i=1}^{n} \sum_{j=1}^{m} H_{ij} \tilde{H}_{i\cdot}^2 + \frac{1}{nm^3} \sum_{i=1}^{n} \sum_{j=1}^{m} H_{ij} \tilde{H}_{\cdot j}^2 + \frac{1}{n^2 m^2} \sum_{i=1}^{n} \sum_{j=1}^{m} H_{ij} \tilde{H}_{i\cdot} \tilde{H}_{\cdot j} \right] + \cdots,
\end{aligned}$$

where $\bar{H}_{\cdot\cdot} = (nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}$. It can be argued that $||\lambda|| = O_p(n^{-1/2})$, and by ignoring its higher order terms, we have:

$$\lambda \doteq \left[\frac{1}{n^2 m}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\tilde{H}_{i\cdot} + \frac{1}{nm^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\tilde{H}_{\cdot j}\right]^{-1} \bar{H}_{\cdot\cdot}. \tag{3.2}$$

Define $\bar{H}_{i\cdot} = m^{-1}\sum_{j=1}^{m} H_{ij}$ and $\bar{H}_{\cdot j} = n^{-1}\sum_{i=1}^{n} H_{ij}$; then

$$\tilde{H}_{i\cdot} = \bar{H}_{i\cdot} - \frac{\lambda}{m^2}\sum_{r=1}^{m} H_{ir}\tilde{H}_{\cdot r} \quad \text{and} \quad \tilde{H}_{\cdot j} = \bar{H}_{\cdot j} - \frac{\lambda}{n^2}\sum_{s=1}^{n} H_{sj}\tilde{H}_{s\cdot}.$$

Replacing $\tilde{H}$ in (3.2) with $\bar{H}$, with the difference absorbed into the coefficient of higher order terms of λ, we have

$$\begin{aligned}
\lambda &\doteq \left[\frac{1}{n^2 m}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\bar{H}_{i\cdot} + \frac{1}{nm^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\bar{H}_{\cdot j}\right]^{-1} \bar{H}_{\cdot\cdot} \\
&= \left[\frac{1}{n^2 m^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\sum_{r=1}^{m} H_{ir} + \frac{1}{n^2 m^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\sum_{s=1}^{n} H_{sj}\right]^{-1} \bar{H}_{\cdot\cdot} \\
&= D^{-1}\bar{H}_{\cdot\cdot},
\end{aligned}$$

where

$$D = \frac{1}{n^2 m^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\sum_{r=1}^{m} H_{ir} + \frac{1}{n^2 m^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\sum_{s=1}^{n} H_{sj}.$$

Now, by keeping terms up to order λ^2 in the log likelihood ratio, we have

$$\begin{aligned}
-2\log r(\theta_0) &= 2\left[\sum_{i=1}^{n}\log\left(1 + \frac{\lambda}{n}\tilde{H}_{i\cdot}\right) + \sum_{j=1}^{m}\log\left(1 + \frac{\lambda}{m}\tilde{H}_{\cdot j}\right)\right] \\
&\doteq 2\sum_{i=1}^{n}\left(\frac{\lambda}{n}\tilde{H}_{i\cdot} - \frac{1}{2}\left(\frac{\lambda}{n}\tilde{H}_{i\cdot}\right)^2\right) + 2\sum_{j=1}^{m}\left(\frac{\lambda}{m}\tilde{H}_{\cdot j} - \frac{1}{2}\left(\frac{\lambda}{m}\tilde{H}_{\cdot j}\right)^2\right) \\
&\doteq 2\sum_{i=1}^{n}\left(\frac{\lambda}{n}\tilde{H}_{i\cdot} - \frac{1}{2}\left(\frac{\lambda}{n}\bar{H}_{i\cdot}\right)^2\right) + 2\sum_{j=1}^{m}\left(\frac{\lambda}{m}\tilde{H}_{\cdot j} - \frac{1}{2}\left(\frac{\lambda}{m}\bar{H}_{\cdot j}\right)^2\right)
\end{aligned}$$

Further replacing $\tilde{H}$'s by their expressions in terms of $\bar{H}$'s and keeping terms up to order λ^2, after some calculation, we have

$$-2\log r(\theta_0) \doteq 4\lambda \bar{H}_{\cdot\cdot} - 3\lambda^2 \left(\frac{1}{n^2} \sum_{i=1}^{n} \bar{H}_{i\cdot}^2 + \frac{1}{m^2} \sum_{j=1}^{m} \bar{H}_{\cdot j}^2 \right).$$

Plugging $\lambda \doteq D^{-1}\bar{H}_{\cdot\cdot}$ into the above expression, we have

$$-2\log r(\theta_0) \doteq \bar{H}_{\cdot\cdot}^2 (4D^{-1} - 3KD^{-2}), \tag{3.3}$$

where

$$K = \frac{1}{n^2} \sum_{i=1}^{n} \bar{H}_{i\cdot}^2 + \frac{1}{m^2} \sum_{j=1}^{m} \bar{H}_{\cdot j}^2.$$

It can be shown that as $\min\{n, m\} \to \infty$,

$$(4D^{-1} - 3KD^{-2}) Var(\bar{H}_{\cdot\cdot}) \to 1,$$

by using an ANOVA decomposition method on $h(X, Y, \theta_0) = I(Y \geq X) - \theta_0$. Therefore, it follows that the asymptotic distribution of $-2\log r(\theta_0)$ is χ_1^2. The α-level empirical likelihood ratio test rejects $H_0 : F = G$ if $-2r(\theta_0) > \chi_{1,\alpha}^2$ for $\theta_0 = 1/2$, where $\chi_{1,\alpha}^2$ is the upper 100αth quantile of the χ_1^2 distribution. We notice that $\bar{H}_{\cdot\cdot}$ is indeed the standard Mann-Whitney statistic from the previous section. Hence, using this two-sample EL technique to test $F = G$ is equivalent to using the standard Mann-Whitney test when the observations are all independent.

3.2.3 Jackknife Empirical Likelihood for Two-Sample U-Statistics

The computation difficulties in solving (3.1) are due to the formulation of the problem involving a U-statistic. Jing et al. (2009) proposed a so-called jackknife empirical likelihood method which can be used for such problems. The main idea of JEL is to construct the asymptotically independent jackknife pseudo values of the statistic of interest, and then apply the regular EL method to the mean of the pseudo values. We focus here on the JEL method for two-sample U-statistics.

Suppose we have two random samples $\{Y_1, \cdots, Y_n\}$ and $\{X_1, \cdots, X_m\}$ from two independent distributions. Define a two-sample U-statistic of degree (k_1, k_2) with a kernel h as follows:

$$\begin{aligned} U_{n,m} &= \binom{n}{k_1}^{-1} \binom{m}{k_2}^{-1} \sum_{1 \le i_1 < \cdots < i_{k_1} \le n} \sum_{1 \le j_1 < \cdots < j_{k_2} \le m} h(Y_{i1}, \cdots, Y_{ik_1}, X_{j1}, \cdots, X_{jk_2}) \\ &=: T(Y_1, \cdots, Y_n, X_1, \cdots, X_m). \end{aligned}$$

Notice the standard Mann-Whitney test statistic

$$MW = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} I(Y_i \ge X_j)$$

is a special case of $U_{n,m}$ for $k_1 = k_2 = 1$, and $h = I(Y_i \ge X_j)$. Let $\theta = Eh(Y_1, \cdots, Y_{k_1}, X_1, \cdots, X_{k_2})$ be the parameter of interest. The jackknife pseudo values can be constructed in the following way. Define

$$Z_i = \begin{cases} Y_i & i = 1, \cdots, n \\ X_{i-n} & i = n+1, \cdots, N, \end{cases}$$

where $N = n + m$. We can write

$$U_{n,m} = T(Y_1, \cdots, Y_n, X_1, \cdots, X_m) = T(Z_1, \cdots, Z_N) = T_N.$$

Denote

$$T_{N-1}^{(-i)} = T(Z_1, \cdots, Z_{i-1}, Z_{i+1}, \cdots, Z_N),$$

which is the statistic $U_{n,m}$ computed on the original data without the ith observation. Then the jackknife pseudo values can be defined as:

$$V_i = N \cdot T_N - (N-1) \cdot T_{N-1}^{(-i)}, \quad i = 1, \cdots, N.$$

Here we notice that $U_{n,m} = (1/N)\sum_{i=1}^{N} V_i$, and the V_i's have been shown to be asymptotically independent (Shi (1984)). Now we apply the EL method to the approximately independent r.v.s V_i, $i = 1, \cdots, N$. Let $\boldsymbol{p} = (p_1, \cdots, p_N)$ be the vector of probabilities assigned to each V_i. The empirical likelihood evaluated at θ can be given by

$$L(\theta) = \max\left\{ \prod_{i=1}^{N} p_i : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i (V_i - EV_i) = 0 \right\},$$

where it can be shown

$$EV_i = \begin{cases} \theta(\frac{N}{N-(k_1+k_2)})\left\{(m-1)\frac{k_1}{n} - (k_2 - 1)\right\} & i = 1, \cdots, n \\ \theta(\frac{N}{N-(k_1+k_2)})\left\{(n-1)\frac{k_2}{m} - (k_1 - 1)\right\} & i = n+1, \cdots, N. \end{cases}$$

Particularly, when $n = m$ and $k_1 = k_2$, $EV_i = \theta$. Solving this maximization problem by Lagrange multipliers method gives

$$\hat{p}_i(\theta) = \frac{1}{N} \frac{1}{1 + \lambda(V_i - EV_i)},$$

where λ is the solution to the equation

$$\frac{1}{N} \sum_{i=1}^{N} \frac{(V_i - EV_i)}{1 + \lambda(V_i - EV_i)} = 0.$$

Since $\prod_{i=1}^{N} p_i$ subject to $\sum_{i=1}^{N} p_i = 1$ is maximized when $p_i = 1/N$, then, the jackknife empirical log-likelihood ratio at θ is:

$$r_{\text{JEL}}(\theta) = \sum_{i=1}^{N} \log\{N\hat{p}_i(\theta)\} = -\sum_{i=1}^{N} \log\{1 + \lambda(V_i - EV_i)\}.$$

Define $g_{1,0}(y) = Eh(y, Y_2, \cdots, Y_{k_1}, X_1, \cdots, X_{k_2}) - \theta$ and $\sigma_{1,0}^2 = \text{var}(g_{1,0}(y))$; $g_{0,1}(x) = Eh(Y_1, Y_2, \cdots, Y_{k_1}, x, \cdots, X_{k_2}) - \theta$ and $\sigma_{0,1}^2 = \text{var}(g_{0,1}(x))$, Jing et al. (2009) have proved that assuming $Eh^2(Y_1, \cdots, Y_{k_1}, X_1, \cdots, X_{k_2}) < \infty$, $\sigma_{1,0}^2 > 0$, $\sigma_{0,1}^2 > 0$ and $0 < \liminf n/m \leq \limsup n/m < \infty$, then

$$-2r_{\text{JEL}}(\theta) \to \chi_1^2$$

in distribution. The JEL test on θ can therefore be constructed based on $-2r_{\text{JEL}}(\theta)$.

3.2.4 Notations under the Setting of Pretest-Posttest Studies

In this subsection, we introduce some notations we use throughout the rest of the chapter under the setting of pretest-posttest studies. Suppose a random sample of n subjects is selected from the target population, and values of the pretest responses $\boldsymbol{Z}$ (including relevant covariates) are measured to collect baseline information for all n subjects. Each subject is then randomly assigned to either a treatment group or a control group with probability δ and $1-\delta$ respectively. Let $R_i = 1$ if subject i is assigned to the treatment group and $R_i = 0$ if subject i is assigned to the control group. The marginal distributions of $\boldsymbol{Z}$ are assumed to be identical in the two groups because of the randomization. Following the concept of counterfactual outcome in causal inference (Rubin (1978)), we let Y_1 and Y_0 be the posttest response a subject potentially would have if he/she receives the treatment and control respectively. For convenience, we label the subjects in the treatment group from 1 to n_1 and the subjects in the control group to be from n_1+1 to n, and $n = n_1+n_0$. The observed data are denoted as $\{(R_i = 1, \boldsymbol{z}_i, y_{1i}), i = 1, \cdots, n_1\}$ for the treatment group, and $\{(R_i = 0, \boldsymbol{z}_i, y_{0i}), i = n_1+1, \cdots, n\}$ for the control group. Note that Y_1 and Y_0 can not be observed at the same time for the same subject. The variables Y_1 and Y_0 are missing by design. It follows from the randomization of subjects that $P(R = 1 \mid \boldsymbol{Z}, Y_1, Y_0) = P(R = 1)$. Let F_1 and F_0 be the marginal distribution function of Y_1 and Y_0 respectively. Our interest is to test hypothesis $H_0 : F_1 = F_0$ against $H_1 : F_1 < F_0$ or $F_1 > F_0$.

When we only consider the posttest responses from two groups in the analysis, we may directly apply one of the methods described previously to the data $\{Y_{11}, \cdots, Y_{1n_1}\}$ and $\{Y_{0(n_1+1)}, \cdots, Y_{0n}\}$. However, in the following sections, we want to extend these methods for incorporating the pretest responses and the baseline covariates into the analyses.

3.3 Adjusted Mann-Whitney Test Based on the HQF Estimators

In this section, we propose an adjusted Mann-Whitney test when both pretest and posttest responses are included in the analysis. From previous sections, we know that the standard Mann-Whitney test assumes that the observations from the two groups are independent from each other. However, after using the pretest information from one sample to correct for missingness in the other, the conventional Mann-Whitney test may no longer be valid. Instead, we propose a method which constructs the adjusted Mann-Whitney test statistic using the EL estimator proposed by Huang et al. (2008), denoted as HQF. We also derive the asymptotic distribution of the adjusted statistic.

As we introduced in Chapter 2, Huang et al. (2008) proposed an EL estimator for the treatment effect $\eta = \mu_1 - \mu_0$ which incorporates the baseline information. Their proposed strategy is to estimate μ_1 and μ_0 separately with appropriate constraints over the pretest variables for both the treatment and the control groups. The HQF estimator of μ_1 is computed as $\hat{\mu}_{1\text{HQF}} = \sum_{i=1}^{n_1} \hat{p}_i y_{1i}$, where $\hat{p}_i$ are obtained through the following EL method. Let $f(\boldsymbol{z}, y_1)$ be the joint density function of $(\boldsymbol{Z}, Y_1)$ related to the treatment group and $f(\boldsymbol{z})$ be the marginal density function of $\boldsymbol{Z}$. Let $p_i = f(\boldsymbol{z}_i, y_{1i})$ for $i = 1, \cdots, n_1$ and $r_i = f(\boldsymbol{z}_i)$ for $i = n_1 + 1, \cdots, n$. The empirical log-likelihood function is given by

$$\ell = \sum_{i=1}^{n_1} \log(p_i) + \sum_{i=n_1+1}^{n} \log(r_i)\,. \tag{3.4}$$

The $\hat{p}_i$ and $\hat{r}_i$ are obtained by maximizing (3.4) subject to $p_i > 0$, $r_i > 0$ and the following constraints:

$$\sum_{i=1}^{n_1} p_i = 1, \quad \sum_{i=n_1+1}^{n} r_i = 1, \tag{3.5}$$

$$\sum_{i=1}^{n_1} p_i a_1(\boldsymbol{z}_i) = \sum_{i=n_1+1}^{n} r_i a_1(\boldsymbol{z}_i) = a_1\,, \tag{3.6}$$

where $a_1(\boldsymbol{z}) = E(Y_1 \mid \boldsymbol{Z} = \boldsymbol{z})$, and $a_1 = E\{a(\boldsymbol{Z})\}$. The actual form of $a_1(\boldsymbol{z})$ is typically unknown, but one could use a guessed form in practice. Using the Lagrange multiplier method, the resulting estimates for p_i and r_i are:

$$\hat{p}_i = \frac{1}{n_1}\frac{1}{1+\lambda\{a_1(\boldsymbol{z}_i) - a_1\}}, \quad i = 1, \cdots, n_1, \tag{3.7}$$

$$\hat{r}_i = \frac{1}{n_0}\frac{1}{1+\tau\{a_1(\boldsymbol{z}_i) - a_1\}}, \quad i = n_1+1, \cdots, n, \tag{3.8}$$

and the Lagrange multipliers λ and τ are determined by solving

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{a_1(\boldsymbol{z}_i) - a_1}{1+\lambda\{a_1(\boldsymbol{z}_i) - a_1\}} = 0\,, \tag{3.9}$$

$$\frac{1}{n_0}\sum_{i=n_1+1}^{n}\frac{a_1(\boldsymbol{z}_i) - a_1}{1+\tau\{a_1(\boldsymbol{z}_i) - a_1\}} = 0\,. \tag{3.10}$$

It has been shown by Huang et al. (2008) that $\hat{\mu}_{1\text{HQF}}$ is more efficient than the naive estimator $\hat{\mu}_1 = n_1^{-1}\sum_{i=1}^{n_1} y_{1i}$. The HQF estimator $\hat{\mu}_{0\text{HQF}} = \sum_{j=n_1+1}^{n}\hat{q}_j y_{0j}$ can similarly be computed using $a_0(\boldsymbol{z}_i)$, $i = 1, \cdots, n$ and y_{0j}, $j = n_1+1, \cdots, n$, where $a_0(\boldsymbol{z}) = E(Y_0 \mid \boldsymbol{Z} = \boldsymbol{z})$.

Our proposed adjusted Mann-Whitney test statistic using the HQF estimators is given by

$$MW_{\text{HQF}} = \sum_{j=n_1+1}^{n}\sum_{i=1}^{n_1}\hat{p}_i\hat{q}_j I(Y_{1i} \geq Y_{0j}). \tag{3.11}$$

where the $\hat{p}_i$'s and $\hat{q}_j$'s are those used for the HQF estimators $\hat{\mu}_{1\text{HQF}}$ and $\hat{\mu}_{0\text{HQF}}$. Note that $\delta = P(R = 1)$ and assume that the baseline information $(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n)$ is an iid sample from $\boldsymbol{Z}$. We have the following result concerning the asymptotic distribution of MW_{HQF}.

Theorem 3. *Suppose that* $0 < \delta < 1$ *and* $E|a_k(\boldsymbol{Z})|^3 < \infty$ *for* $k = 1, 0$. *Then under the null hypothesis* H_0: $F_1 = F_0$,

$$\sqrt{n}\big(MW_{\text{HQF}} - 1/2\big) \xrightarrow{d} N\big(0, E(A_i^2)\big)\,,$$

where

$$
\begin{aligned}
A_i &= -\frac{R_i}{\delta}(1 - F_0(Y_{1i})) + \frac{1-R_i}{1-\delta}(1 - F_1(Y_{0i})) \\
&+ \begin{pmatrix} -E\{I(Y_1 \geq Y_0)\boldsymbol{\psi}_1(\boldsymbol{Z})\}/\delta \\ -E\{I(Y_1 \geq Y_0)\boldsymbol{\psi}_0(\boldsymbol{Z})\}/(1-\delta) \end{pmatrix}^T \\
&\times \begin{pmatrix} \{E(\boldsymbol{\psi}_1(\boldsymbol{Z})\boldsymbol{\psi}_1^T(\boldsymbol{Z}))\}^{-1} & 0 \\ 0 & \{E(\boldsymbol{\psi}_0(\boldsymbol{Z})\boldsymbol{\psi}_0^T(\boldsymbol{Z}))\}^{-1} \end{pmatrix} (R_i - \delta) \begin{pmatrix} \boldsymbol{\psi}_1(\boldsymbol{Z}_i) \\ -\boldsymbol{\psi}_0(\boldsymbol{Z}_i) \end{pmatrix}
\end{aligned}
$$

with $\boldsymbol{\psi}_k(\boldsymbol{Z}_i) = (1, a_k(\boldsymbol{Z}_i))^T, k = 1, 0$ *and* $\delta = P(R_i = 1)$.

Proof. Following the arguments in the Appendix of Huang et al. (2008), $\hat{p}_i$ of equation (3.7) can be reparameterized as

$$
\hat{p}_i = \frac{1}{n_1}\frac{1}{1+\lambda_1\{a_1(\boldsymbol{z}_i) - a_1\}} = \frac{1}{n}\frac{1}{\boldsymbol{\xi}_1^T\boldsymbol{\psi}_1(\boldsymbol{z}_i)}, \quad i = 1, \cdots, n_1, \tag{3.12}
$$

with $\boldsymbol{\xi}_1 = (n_1(1-\lambda_1 a_1)/n, n_1\lambda_1/n)^T$. Similarly, $\hat{q}_j$ can be reparameterized as

$$
\hat{q}_j = \frac{1}{n_0}\frac{1}{1+\lambda_0\{a_0(\boldsymbol{z}_i) - a_0\}} = \frac{1}{n}\frac{1}{\boldsymbol{\xi}_0^T\boldsymbol{\psi}_0(\boldsymbol{z}_i)}, \quad j = n_1+1, \cdots, n, \tag{3.13}
$$

with $\boldsymbol{\xi}_0 = (n_0(1-\lambda_0 a_0)/n, n_0\lambda_0/n)^T$. Let $\hat{\boldsymbol{\xi}}_1$ and $\hat{\boldsymbol{\xi}}_0$ be the estimates of $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_0$ respectively; then $\hat{\boldsymbol{\xi}}_1$ and $\hat{\boldsymbol{\xi}}_0$ are the solutions to estimating equations

$$
\begin{aligned}
U_1(\boldsymbol{\xi}_1) &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{R_i}{\boldsymbol{\xi}_1^T\boldsymbol{\psi}_1(\boldsymbol{z}_i)} - \frac{1-R_i}{1-\boldsymbol{\xi}_1^T\boldsymbol{\psi}_1(\boldsymbol{z}_i)}\right]\boldsymbol{\psi}_1(\boldsymbol{z}_i) = \mathbf{0}, \\
U_0(\boldsymbol{\xi}_0) &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1-R_i}{\boldsymbol{\xi}_0^T\boldsymbol{\psi}_0(\boldsymbol{z}_i)} - \frac{R_i}{1-\boldsymbol{\xi}_0^T\boldsymbol{\psi}_1(\boldsymbol{z}_i)}\right]\boldsymbol{\psi}_0(\boldsymbol{z}_i) = \mathbf{0}.
\end{aligned}
$$

Denote $\boldsymbol{\xi}_{10} = (\delta, 0)^T$ and $\boldsymbol{\xi}_{00} = (1-\delta, 0)^T$. We notice

$$
\begin{aligned}
\boldsymbol{\xi}_{10}^T\boldsymbol{\psi}_1(\boldsymbol{z}_i) &= \delta \quad \Longrightarrow \quad E[U_1(\boldsymbol{\xi}_{10}^T)] = \mathbf{0} \\
\boldsymbol{\xi}_{00}^T\boldsymbol{\psi}_0(\boldsymbol{z}_i) &= 1-\delta \quad \Longrightarrow \quad E[U_0(\boldsymbol{\xi}_{00}^T)] = \mathbf{0}.
\end{aligned}
$$

Assuming that $E|a_1(\boldsymbol{Z})|^3 < \infty$ and $E|a_0(\boldsymbol{Z})|^3 < \infty$, then by a Taylor expansion, we have

$$\sqrt{n}\left[\begin{pmatrix}\hat{\boldsymbol{\xi}}_1\\ \hat{\boldsymbol{\xi}}_0\end{pmatrix} - \begin{pmatrix}\boldsymbol{\xi}_{10}\\ \boldsymbol{\xi}_{00}\end{pmatrix}\right] = D^{-1} \times \sqrt{n}\begin{pmatrix}U_1(\boldsymbol{\xi}_{10})\\ U_0(\boldsymbol{\xi}_{00})\end{pmatrix} + o_p(1),$$

where

$$\begin{aligned} D &= \lim_{n\to\infty}\begin{pmatrix}-\partial U_1/\partial \boldsymbol{\xi}_1^T & 0\\ 0 & -\partial U_0/\partial \boldsymbol{\xi}_0^T\end{pmatrix}\Bigg|_{(\boldsymbol{\xi}_1=\boldsymbol{\xi}_{10},\boldsymbol{\xi}_0=\boldsymbol{\xi}_{00})} \\ &= \begin{pmatrix}(1/\delta(1-\delta))E[\boldsymbol{\psi}_1(\boldsymbol{Z})\boldsymbol{\psi}_1^T(\boldsymbol{Z})] & 0\\ 0 & (1/(1-\delta)\delta)E[\boldsymbol{\psi}_0(\boldsymbol{Z})\boldsymbol{\psi}_0^T(\boldsymbol{Z})]\end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} \begin{pmatrix}U_1(\boldsymbol{\xi}_{10})\\ U_1(\boldsymbol{\xi}_{00})\end{pmatrix} &= \begin{pmatrix}1/n\sum_{i=1}^n((R_i-\delta)/(\delta(1-\delta)))\boldsymbol{\psi}_1(\boldsymbol{z}_i)\\ 1/n\sum_{i=1}^n(((1-R_i)-(1-\delta))/(\delta(1-\delta)))\boldsymbol{\psi}_0(\boldsymbol{z}_i)\end{pmatrix} \\ &= \frac{1}{n}\sum_{i=1}^n\left(\frac{R_i-\delta}{\delta(1-\delta)}\right)\begin{pmatrix}\boldsymbol{\psi}_1(\boldsymbol{z}_i)\\ -\boldsymbol{\psi}_0(\boldsymbol{z}_i)\end{pmatrix} \end{aligned}$$

Now, we plug $\hat{p}_i(\hat{\boldsymbol{\xi}}_1)$ and $\hat{q}_j(\hat{\boldsymbol{\xi}}_0)$ into the adjusted Mann-Whitney statistic in equation (3.11).

Under the null hypothesis $H_0 : F_1 = F_0$ and by the Taylor expansion we have

$$\begin{aligned}
\sqrt{n}(MW_{\text{HQF}} - 1/2) &= \sqrt{n}\Big\{ \sum_{j=1}^{n}\sum_{i=1}^{n} R_i(1-R_j)I(Y_{1i} \geq Y_{0j})\hat{p}_i(\hat{\boldsymbol{\xi}}_1)\hat{q}_j(\hat{\boldsymbol{\xi}}_0) - 1/2 \Big\} \\
&= \sqrt{n}\Big[\sum_{j=1}^{n}\sum_{i=1}^{n} R_i(1-R_j)I(Y_{1i} \geq Y_{0j})\hat{p}_i(\boldsymbol{\xi}_{10})\hat{q}_j(\boldsymbol{\xi}_{00}) - 1/2 \Big] + \\
&\quad \begin{pmatrix} E(\partial MW_{\text{HQF}}/\partial \boldsymbol{\xi}_1^T) \\ E(\partial MW_{\text{HQF}}/\partial \boldsymbol{\xi}_0^T) \end{pmatrix}^T_{\boldsymbol{\xi}_1=\boldsymbol{\xi}_{10},\boldsymbol{\xi}_1=\boldsymbol{\xi}_{00}} \cdot \sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_{10} \\ \hat{\boldsymbol{\xi}}_0 - \boldsymbol{\xi}_{00} \end{pmatrix} + o_p(1) \\
&= \underbrace{\sqrt{n}\Big[\sum_{j=1}^{n}\sum_{i=1}^{n} R_i(1-R_j)I(Y_{1i} \geq Y_{0j})\hat{p}_i(\boldsymbol{\xi}_{10})\hat{q}_j(\boldsymbol{\xi}_{00}) - 1/2 \Big] +}_{(*)} \\
&\quad \begin{pmatrix} -E[I(Y_1 \geq Y_0)\boldsymbol{\psi}_1(\boldsymbol{Z})]/\delta \\ -E[I(Y_1 \geq Y_0)\boldsymbol{\psi}_0(\boldsymbol{Z})]/(1-\delta) \end{pmatrix}^T \times \\
&\quad \begin{pmatrix} E[\boldsymbol{\psi}_1(\boldsymbol{Z})\boldsymbol{\psi}_1^T(\boldsymbol{Z})] & 0 \\ 0 & E[\boldsymbol{\psi}_0(\boldsymbol{Z})\boldsymbol{\psi}_0^T(\boldsymbol{Z})] \end{pmatrix}^{-1} \times \\
&\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(R_i - \delta)\begin{pmatrix} \boldsymbol{\psi}_1(\boldsymbol{z}_i) \\ -\boldsymbol{\psi}_0(\boldsymbol{z}_i) \end{pmatrix} + o_p(1).
\end{aligned}$$

We can notice that $\hat{p}_i(\boldsymbol{\xi}_{10}) = 1/(\delta n)$ and $\hat{q}_i(\boldsymbol{\xi}_{00}) = 1/((1-\delta)n)$. Therefore, the term with the double-summation in the above equation is in fact the standard Mann-Whitney test statistic. Let

$$\begin{aligned}
\tilde{F}_1(y) &= \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\delta}I(Y_{1i} \leq y) \\
\tilde{F}_0(y) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1-R_i}{1-\delta}I(Y_{0i} \leq y)
\end{aligned}$$

Then we have

$$
\begin{aligned}
(*) &= \sqrt{n}\Big[\sum_{j=1}^{n}\sum_{i=1}^{n} R_i(1-R_j)I(Y_{1i}\geq Y_{0j})\hat{p}_i(\boldsymbol{\xi}_{10})\hat{q}_j(\boldsymbol{\xi}_{00}) - 1/2\Big] \\
&= \sqrt{n}\Big[\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{R_i(1-R_j)}{\delta(1-\delta)} I(Y_{1i}\geq Y_{0j}) - 1/2\Big] \\
&= \sqrt{n}\Big(\int \tilde{F}_0 d\tilde{F}_1 - \int F_0 dF_1\Big),
\end{aligned}
$$

where F_1 and F_0 are the marginal cdf of Y_1 and Y_0. Define $\phi(h_0,h_1) = \int h_0 dh_1 : (h_0,h_1) \mapsto \mathbb{R}$. Then by Lemma 20.10 in van der Vaart (1998), ϕ is Hadamard differentiable and

$$
\phi'(h_0,h_1)\big|_{(F_0,F_1)} = h_1 F_0\big|_{-\infty}^{\infty} - \int h_1 dF_0 + \int h_0 dF_1.
$$

Now by the Functional Delta method (Theorem 20.8 in van der Vaart (1998)), we have

$$
\begin{aligned}
(*) &= \sqrt{n}\Big(\int \tilde{F}_0 d\tilde{F}_1 - \int F_0 dF_1\Big) \\
&= \sqrt{n}(\phi(\tilde{F}_0,\tilde{F}_1) - \phi(F_0,F_1)) \\
&= \phi'(\sqrt{n}(\tilde{F}_0 - F_0), \sqrt{n}(\tilde{F}_1 - F_1)) + o_p(1) \\
&= \sqrt{n}(\tilde{F}_1 - F_1)\cdot F_0\big|_{-\infty}^{\infty} - \int \sqrt{n}(\tilde{F}_1 - F_1) dF_0 + \int \sqrt{n}(\tilde{F}_0 - F_0) dF_1 + o_p(1) \\
&= 0 - \sqrt{n}\int \tilde{F}_1 dF_0 + \sqrt{n}\int \tilde{F}_0 dF_1 + o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{-\infty}^{\infty} \frac{R_i}{\delta} I(Y_{1i}\leq y) dF_0(y) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{-\infty}^{\infty} \frac{1-R_i}{1-\delta} I(Y_{0i}\leq y) dF_1(y) + o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{R_i}{\delta}(1 - F_0(Y_{1i})) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{1-R_i}{1-\delta}(1 - F_1(Y_{0i})) + o_p(1)\,.
\end{aligned}
$$

It follows that

$$
\sqrt{n}(MW_{\text{HQF}} - 1/2) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i + o_p(1)\,, \tag{3.14}
$$

where

$$
\begin{aligned}
A_i &= -\frac{R_i}{\delta}(1 - F_0(Y_{1i})) + \frac{1 - R_i}{1 - \delta}(1 - F_1(Y_{0i})) \\
&\quad + \begin{pmatrix} -1/\delta E\{I(Y_1 \geq Y_0)\boldsymbol{\psi}_1(\boldsymbol{Z})\} \\ -1/(1-\delta)E\{I(Y_1 \geq Y_0)\boldsymbol{\psi}_0(\boldsymbol{Z})\} \end{pmatrix}^T \\
&\quad \times \begin{pmatrix} \{E(\boldsymbol{\psi}_1(\boldsymbol{Z})\boldsymbol{\psi}_1^T(\boldsymbol{Z}))\}^{-1} & 0 \\ 0 & \{E(\boldsymbol{\psi}_0(\boldsymbol{Z})\boldsymbol{\psi}_0^T(\boldsymbol{Z}))\}^{-1} \end{pmatrix} \cdot (R_i - \delta) \cdot \begin{pmatrix} \boldsymbol{\psi}_1(\boldsymbol{Z}_i) \\ -\boldsymbol{\psi}_0(\boldsymbol{Z}_i) \end{pmatrix}
\end{aligned}
$$

Since R is independent of Y_1 and Y_0, A_i's are iid random variables with $E(A_i) = 0$. Hence, by the Central Limit Theorem,

$$\sqrt{n}(MW_{\text{HQF}} - 1/2) \xrightarrow{d} N(0, E(A_i^2)),$$

under the null hypothesis. □

A two-sided α-level adjusted Mann-Whitney test rejects $H_0 : F_1 = F_0$ when

$$\left|\{\hat{E}(A_i^2)\}^{-1/2}\sqrt{n}(MW_{\text{HQF}} - 1/2)\right| \geq Z_{1-\alpha/2}, \tag{3.15}$$

where $Z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$-th quantile of $N(0, 1)$, $\hat{E}(A_i^2) = n^{-1}\sum_{i=1}^n \hat{A}_i^2$, and $\hat{A}_i$ is the simple plug-in estimator for A_i.

We have shown that $\sqrt{n}\big(MW_{\text{HQF}} - 1/2\big) = n^{-1/2}\sum_{i=1}^n A_i + o_p(1)$ where the A_i's are iid random variables with $E(A_i) = 0$. An important observation is that this result holds even if $a_k(\boldsymbol{Z}) = E(Y_k|\boldsymbol{Z})$, $k = 1, 0$ is not correctly specified. When $a_k(\boldsymbol{Z})$, $k = 1, 0$ is misspecified, the test based on (3.15) is still valid with size approximately equal to α. However, in this case, we expect the test to be less efficient in terms of a decrease in power.

3.4 Empirical Likelihood Based Mann-Whitney Test with Imputation

In this section, we present the EL based Mann-Whitney test for pretest-posttest studies under the imputation approach discussed in Chapter 2, which incorporates the baseline information and the missing-by-design data structure through an imputation model. We extend the methods we described in Section 3.2.2 and Section 3.2.3 by imputing Y_1^* and Y_0^* for subjects in the control group and in the treatment group respectively based on their pretest responses and other baseline covariates, and applying the two-sample EL method and the JEL method to the "enlarged" samples.

The imputation technique we consider in this section is the so-called stochastic regression imputation (Little and Rubin (1987)), which replaces each missing value in the data by a predicted value from fitting a regression model to the complete cases (as in the regression imputation) plus a random residual term. The reason we use the stochastic regression imputation here is that we would like to preserve the distributions of the posttest responses approximately after imputation. The following linear models are assumed to be true.

$$
\begin{aligned}
Y_{1i} &= \boldsymbol{Z}_i^T \boldsymbol{\beta}_1 + \epsilon_{1i}, i = 1, \cdots, n, & (3.16)\\
Y_{0i} &= \boldsymbol{Z}_i^T \boldsymbol{\beta}_0 + \epsilon_{0i}, i = 1, \cdots, n, & (3.17)
\end{aligned}
$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ are the regression parameters for treatment and control, and the ϵ_{1i}'s and ϵ_{0i}'s are independent errors with zero mean and variance $\sigma_{\epsilon 1}^2$ and $\sigma_{\epsilon 0}^2$ respectively. The assumed models imply that, for the initial sample of n selected subjects, the posttest response would follow model (3.16) if the subject is assigned to the treatment group and would follow model (3.17) if the subject is assigned to the control group. The observed sample data are $\{(R_i = 1, \boldsymbol{Z}_i, Y_{1i}), i = 1, \cdots, n_1\}$ for the treatment group and $\{(R_i = 0, \boldsymbol{Z}_i, Y_{0i}), i = n_1+1, \cdots, n\}$ for the control group. The information of the pretest responses $\{\boldsymbol{Z}_i, i = n_1+1, \cdots, n\}$ can be used to impute the potential Y_{1i}^* for subjects $i = n_1+1, \cdots, n$

through model (3.16), and the information of the pretest responses $\{\boldsymbol{Z}_i, i = 1, \cdots, n_1\}$ can be used to impute the potential Y_{0i}^* for subjects $i = 1, \cdots, n_1$ through model (3.17). Let

$$\begin{aligned}\hat{\boldsymbol{\beta}}_1 &= (\sum_{i=1}^{n} R_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T)^{-1} \sum_{i=1}^{n} R_i \boldsymbol{Z}_i Y_{1i}, \\ \hat{\boldsymbol{\beta}}_0 &= (\sum_{i=1}^{n} (1-R_i) \boldsymbol{Z}_i \boldsymbol{Z}_i^T)^{-1} \sum_{i=1}^{n} (1-R_i) \boldsymbol{Z}_i Y_{0i},\end{aligned}$$

be the ordinary least squares estimators for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ using observed data. Define

$$\begin{aligned}Y_{1i}^* &= \boldsymbol{Z}_i^T \hat{\boldsymbol{\beta}}_1 + \epsilon_{1i}^*, \quad \text{for} \quad i = n_1+1, \cdots, n \\ Y_{0i}^* &= \boldsymbol{Z}_i^T \hat{\boldsymbol{\beta}}_0 + \epsilon_{0i}^*, \quad \text{for} \quad i = 1, \cdots, n_1\end{aligned}$$

to be the imputed values obtained through the stochastic regression imputation method, where ϵ_{1i}^* is a randomly selected element from the residual vector, $\{\boldsymbol{r}_{1i} = Y_{1i} - \boldsymbol{Z}_i^T \hat{\boldsymbol{\beta}}_1, i = 1, \cdots, n_1\}$, determined from fitting regression model (3.16) to the observed data from the treatment group; and ϵ_{0i}^* is a randomly selected element from the residual vector, $\{\boldsymbol{r}_{0i} = Y_{0i} - \boldsymbol{Z}_i^T \hat{\boldsymbol{\beta}}_0, i = n_1+1, \cdots, n\}$, determined from fitting regression model (3.17) to the observed data from the control group. Let $\{\tilde{Y}_{1i} = R_i Y_{1i} + (1-R_i) Y_1^*, i = 1, \cdots, n\}$ and $\{\tilde{Y}_{0i} = (1-R_i) Y_{0i} + R_i Y_0^*, i = 1, \cdots, n\}$. With the imputation strategy, we now have two "enlarged" samples: $\{\tilde{Y}_{1i}, i = 1, \cdots, n\}$ for the treatment, $\{\tilde{Y}_{0i}, i = 1, \cdots, n\}$ for the control, and both samples are of size n. In Chapter 2, we showed that the EL-based test for the treatment effect using the imputed samples is more powerful than the tests without using the baseline information. The goal of this section is to develop tests for H_0: $F_1 = F_0$ against H_1: $F_1 < F_0$ using the enlarged sample data with imputed values.

We first consider the two-sample EL method discussed in Section 3.2.2. Let $\theta = E(I(\tilde{Y}_1 > \tilde{Y}_0))$. Let $\boldsymbol{p} = (p_1, \cdots, p_n)$ and $\boldsymbol{q} = (q_1, \cdots, q_n)$ be the discrete probability measures over the two samples $\{\tilde{Y}_{11}, \cdots, \tilde{Y}_{1n}\}$ and $\{\tilde{Y}_{01}, \cdots, \tilde{Y}_{0n}\}$, respectively. The two-sample empirical likelihood function is given by

$$\ell(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{n} \log p_i + \sum_{j=1}^{n} \log q_j.$$

Let $\tilde{p}_i = \tilde{p}_i(\theta)$ and $\tilde{q}_j = \tilde{q}_j(\theta)$ be obtained by maximizing $\ell(\boldsymbol{p}, \boldsymbol{q})$ subject to the constraints

$$\sum_{i=1}^{n} p_i = \sum_{j=1}^{n} q_j = 1,$$
$$\sum_{i=1}^{n}\sum_{j=1}^{n} p_i q_j [I(\tilde{Y}_{1i} \geq \tilde{Y}_{0j}) - \theta] = 0,$$

for a fixed θ. Define $H^*_{ij}(\theta) = H^*_{ij} = I(\tilde{Y}_{1i} \geq \tilde{Y}_{0j}) - \theta$. Then $\tilde{p}_i$ and $\tilde{q}_i$ are given by:

$$\begin{aligned}
\tilde{p}_i &= \frac{1}{n + \lambda \sum_{r=1}^{n} \tilde{q}_r H^*_{ir}} = \frac{1}{n}\frac{1}{1 + \lambda \tilde{H}^*_{i\cdot}/n}, \\
\tilde{q}_j &= \frac{1}{n + \lambda \sum_{s=1}^{n} \tilde{p}_s H^*_{sj}} = \frac{1}{n}\frac{1}{1 + \lambda \tilde{H}^*_{\cdot j}/n},
\end{aligned}$$

where $\tilde{H}^*_{i\cdot} = \sum_{r=1}^{n} \tilde{q}_r H^*_{ir}$, and $\tilde{H}^*_{\cdot j} = \sum_{s=1}^{n} \tilde{p}_s H^*_{sj}$. The Lagrange multiplier λ is determined through

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{p}_i \tilde{q}_j [I(\tilde{Y}_{1i} \geq \tilde{Y}_{0j}) - \theta] = \sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{p}_i \tilde{q}_j H^*_{ij} = 0. \tag{3.18}$$

The empirical log-likelihood ratio statistic on θ with the imputed samples is computed as

$$\begin{aligned}
\tilde{r}(\theta) &= \sum_{i=1}^{n} \log\{n\tilde{p}_i(\theta)\} + \sum_{j=1}^{n} \log\{n\tilde{q}_j(\theta)\} \\
&= -\left[\sum_{i=1}^{n} \log\left(1 + \frac{\lambda}{n}\tilde{H}^*_{i\cdot}\right) + \sum_{j=1}^{n} \log\left(1 + \frac{\lambda}{n}\tilde{H}^*_{\cdot j}\right)\right].
\end{aligned}$$

Unfortunately, due to the complicated dependence structures among the $\tilde{Y}_{1i}$'s and the $\tilde{Y}_{0j}$'s, the asymptotic distribution of $\tilde{r}(\theta)$ does not seem to have a tractable form. We propose to use a bootstrap calibrated α-level test as follows: Reject H_0 if $-2\tilde{r}(\theta_0) > b_\alpha$, where $\theta_0 = 1/2$ and b_α is the approximate upper α-quantile of the sampling distribution of $-2\tilde{r}(\theta_0)$ obtained through the following bootstrap procedures. Let $\tilde{\theta} = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} I(\tilde{Y}_{1i} \geq \tilde{Y}_{0j})$.

(1) Select bootstrap samples $s_1^{\#}$ of size n_1 and $s_0^{\#}$ of size $n - n_1$ from the original treatment sample and control sample, respectively, using simple random sampling with

replacement; denote the two bootstrap sample data sets as $\{(\boldsymbol{Z}_i^{\#}, Y_{1i}^{\#}), i \in s_1^{\#}\}$ and $\{(\boldsymbol{Z}_j^{\#}, Y_{0j}^{\#}), j \in s_0^{\#}\}$;

(2) Use the stochastic regression imputation method to obtain imputed values for $Y_{1i}^{\#}$, $i \in s_0^{\#}$ and for $Y_{0j}^{\#}$, $j \in s_1^{\#}$; compute $\tilde{r}(\theta)$ at $\theta = \tilde{\theta}$, denoted as $\tilde{r}^{\#}(\tilde{\theta})$, using the two imputed bootstrap samples;

(3) Repeat steps (1) and (2) B times, independently, to obtain the sequence $\{-2\tilde{r}_1^{\#}(\tilde{\theta}), \cdots, -2\tilde{r}_B^{\#}(\tilde{\theta})\}$; let b_α be the $100(1-\alpha)$th sample quantile of the sequence.

The jackknife EL method described in Section 3.2.3 can also be applied here to reduce the computational burden of the test procedures. After imputation, our data become $\{\tilde{Y}_{1i} = R_i Y_{1i} + (1-R_i)Y_{1i}^*, i = 1, \cdots, n\}$ for the treatment group, and $\{\tilde{Y}_{0i} = (1-R_i)Y_{0i} + R_i Y_{0i}^*, i = 1, \cdots, n\}$ for the control. Let

$$MW_{\text{IMP}} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I(\tilde{Y}_{1i} \geq \tilde{Y}_{0j}),$$

and $\theta = E(I(\tilde{Y}_{1i} \geq \tilde{Y}_{0j}))$. We construct the jackknife pseudo values as the following. Let

$$\tilde{Z}_i = \begin{cases} \tilde{Y}_{1i} & i = 1, \cdots, n \\ \tilde{Y}_{0(i-n)} & i = n+1, \cdots, 2n. \end{cases}$$

We can write

$$MW_{\text{IMP}} = T(\tilde{Y}_{11}, \cdots, \tilde{Y}_{1n}, \tilde{Y}_{01}, \cdots, \tilde{Y}_{0n}) = T(\tilde{Z}_1, \cdots, \tilde{Z}_{2n}) = T_{2n}.$$

Define $T_{2n-1}^{(-i)} = T(\tilde{Z}_1, \cdots, \tilde{Z}_{i-1}, \tilde{Z}_{i+1}, \cdots, \tilde{Z}_{2n})$, as the statistic computed on the data with ith observation deleted. Specifically, we have:

$$\begin{aligned} T_{2n-1}^{(-i)} &= \frac{1}{n(n-1)} \sum_{\substack{1\leq k\leq n \\ k\neq i}} \sum_{1\leq l\leq n} I(\tilde{Y}_{1k} \geq \tilde{Y}_{0l}), \quad \text{when } 1 \leq i \leq n; \\ T_{2n-1}^{(-i)} &= \frac{1}{n(n-1)} \sum_{1\leq k\leq n} \sum_{\substack{1\leq l\leq n \\ l\neq i}} I(\tilde{Y}_{1k} \geq \tilde{Y}_{0l}), \quad \text{when } n+1 \leq i \leq 2n. \end{aligned}$$

Then the jackknife pseudo values can be defined as:

$$\tilde{V}_i = 2nT_{2n} - (2n-1)T_{2n-1}^{(-i)}, \quad i = 1, \cdots, 2n.$$

And it can be shown $MW_{\text{IMP}} = 1/(2n)\sum_{i=1}^{2n} \tilde{V}_i$ and $E(\tilde{V}_i) = \theta$. Since $\tilde{Y}_{1i}$'s and $\tilde{Y}_{0i}$'s are not independent r.v.'s, the jackknife pseudo values $\tilde{V}_i$'s are no longer asymptotically independent. Let $\boldsymbol{p} = (p_1, \cdots, p_{2n})$ be the vector of probabilities assigned to each $\tilde{V}_i$. The empirical likelihood evaluated at θ is given by

$$L(\theta) = \max\left\{\prod_{i=1}^{2n} p_i : \sum_{i=1}^{2n} p_i = 1, \sum_{i=1}^{2n} p_i\tilde{V}_i = \theta\right\}.$$

Using the Lagrange multipliers method, we will obtain

$$\tilde{p}_i(\theta) = \frac{1}{2n}\frac{1}{1+\lambda(\tilde{V}_i - \theta)},$$

where λ is the solution to the equation

$$\frac{1}{2n}\sum_{i=1}^{2n}\frac{(\tilde{V}_i - \theta)}{1+\lambda(\tilde{V}_i - \theta)} = 0.$$

Since $\prod_{i=1}^{2n} p_i$, subject to $\sum_{i=1}^{2n} p_i = 1$ is maximized when $p_i = 1/2n$, then the jackknife empirical log-likelihood ratio at θ is given by:

$$\tilde{r}_{\text{JEL}}(\theta) = \sum_{i=1}^{2n}\log(2n\tilde{p}_i(\theta)) = -\sum_{i=1}^{2n}\log\{1+\lambda(\tilde{V}_i - \theta)\}.$$

Since the $\tilde{V}_i$'s are not asymptotically independent, the asymptotic distribution of $\tilde{r}_{\text{JEL}}(\theta)$ does not have a tractable form. A bootstrap calibrated α-level test can be conducted as follows. Reject H_0 when $\tilde{r}_{\text{JEL}}(\theta_0) > b'_\alpha$, where $\theta_0 = 1/2$ and b'_α can be obtained through the following procedure: Let $\tilde{\theta}_{\text{JEL}} = (2n)^{-1}\sum_{i=1}^{2n}\tilde{V}_i$,

(1) Select bootstrap samples $s_1^{\#}$ of size n_1 and $s_0^{\#}$ of size $n - n_1$ from the original treatment sample and control sample, respectively, using simple random sampling with replacement; denote the two bootstrap sample data sets as $\{(\boldsymbol{Z}_i^{\#}, Y_{1i}^{\#}), i \in s_1^{\#}\}$ and $\{(\boldsymbol{Z}_j^{\#}, Y_{0j}^{\#}), j \in s_0^{\#}\}$;

(2) Use the stochastic regression imputation method to obtain imputed values for $Y_{1i}^{\#}$, $i \in s_0^{\#}$ and for $Y_{0j}^{\#}$, $j \in s_1^{\#}$; compute $\tilde{r}_{\text{JEL}}(\theta)$ at $\theta = \tilde{\theta}_{\text{JEL}}$, denoted as $\tilde{r}_{\text{JEL}}^{\#}(\tilde{\theta}_{\text{JEL}})$, using the two imputed bootstrap samples;

(3) Repeat steps (1) and (2) B times, independently, to obtain the sequence $\{-2\tilde{r}_{\text{JEL}1}^{\#}(\tilde{\theta}_{\text{JEL}}), \cdots, -2\tilde{r}_{\text{JEL}B}^{\#}(\tilde{\theta}_{\text{JEL}})\}$; let b'_α be the $100(1-\alpha)$th sample quantile of the sequence.

One limitation of the stochastic regression imputation technique we considered in this section is that after the imputation, the distribution of the imputed data is not exactly the same as the distribution of the observed data. In other words, $E(I(\tilde{Y}_{1i} > \tilde{Y}_{0i})) \neq E(I(Y_{1i} > Y_{0i}))$. This may not be an ideal situation when we want to test the difference of the distributions or make inference on $\theta = E(I(Y_{1i} > Y_{0i}))$. Another disadvantage of the imputation based methods is that the computation becomes much slower since the sample sizes for both treatment and control groups are extended. Although the JEL method is computationally more efficient than the EL method, it still suffers from slow computation after imputation. When we pool the expanded sample of the treatment group and the control group together to construct the jackknife pseudo values, the computation becomes slower as the sample size n increases. In the next section, we propose a two-sample JEL method which constructs the jackknife pseudo values separately for the treatment group and the control group, and incorporates the pretest response through a constraint of the EL problem.

3.5 Two-sample Jackknife EL Method for Mann-Whitney Test

The empirical likelihood based Mann-Whitney test presented in Section 3.3 incorporates baseline information through the constraint (3.6) but it relies on the asymptotic normality of MW_{HQF}. The EL ratio statistic on θ cannot be computed under the setting where $\hat{p}_i$ and $\hat{q}_j$ are computed separately using the HQF approach. The imputation approach described in Section 3.4 is not only computationally heavy but also technically difficult. The asymptotic distribution of the EL ratio statistic $\tilde{r}(\theta)$ or $\tilde{r}_{\text{JEL}}(\theta)$ is not readily available and the proposed bootstrap calibration methods are ad hoc procedures. In this section, we present a two-sample jackknife empirical likelihood method for the Mann-Whitney test which allows simple computations, direct incorporations of pretest measures through additional constraints, and rigorous justification of bootstrap calibrations for the EL ratio test.

To simplify notation, let $n_0 = n - n_1$ and denote the sample data for the treatment and the control groups as $\{(\boldsymbol{Z}_{1i}, Y_{1i}), i = 1, \cdots, n_1\}$ and $\{(\boldsymbol{Z}_{0j}, Y_{0j}), j = 1, \cdots, n_0\}$, respectively. Note that $\boldsymbol{Z}_{1i}$ and $\boldsymbol{Z}_{0j}$ share a common distribution but Y_{1i} and Y_{0j} might not. Let $\boldsymbol{Y}_1 = (Y_{11}, \cdots, Y_{1n_1})$ and $\boldsymbol{Y}_0 = (Y_{01}, \cdots, Y_{0n_0})$. Let

$$T_{n_1,n_0}(\boldsymbol{Y}_1, \boldsymbol{Y}_0) = \frac{1}{n_1}\frac{1}{n_0}\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} I(Y_{1i} \geq Y_{0j})\,.$$

Let $\theta = E\big(I(Y_{1i} \geq Y_{0j})\big) = P(Y_1 \geq Y_0)$, where Y_1 and Y_0 are the original response variables under the treatment and the control, respectively. It follows that $E\big(T_{n_1,n_0}(\boldsymbol{Y}_1, \boldsymbol{Y}_0)\big) = \theta$ for any sample sizes n_1 and n_0. Define the two-sample jackknife pseudo values as

$$\begin{aligned} U_i &= n_1 T_{n_1,n_0}(\boldsymbol{Y}_1, \boldsymbol{Y}_0) - (n_1 - 1)T_{n_1-1,n_0}(\boldsymbol{Y}_1[-i], \boldsymbol{Y}_0)\,, \quad i = 1, \cdots, n_1\,, \\ V_j &= n_0 T_{n_1,n_0}(\boldsymbol{Y}_1, \boldsymbol{Y}_0) - (n_0 - 1)T_{n_1,n_0-1}(\boldsymbol{Y}_1, \boldsymbol{Y}_0[-j])\,, \quad j = 1, \cdots, n_0\,, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{Y}_1[-i] &= (Y_{11}, \cdots, Y_{1(i-1)}, Y_{1(i+1)}, \cdots, Y_{1n_1})\,, \\ \boldsymbol{Y}_0[-j] &= (Y_{01}, \cdots, Y_{0(j-1)}, Y_{0(j+1)}, \cdots, Y_{0n_0})\,. \end{aligned}$$

and

$$T_{n_1-1,n_0}(\boldsymbol{Y}_1[-i],\boldsymbol{Y}_0) = \frac{1}{(n_1-1)n_0}\sum_{\substack{1\le k\le n_1\\ k\ne i}}\sum_{1\le j\le n_0} I(Y_{1k}\ge Y_{0j}),$$
$$T_{n_1,n_0-1}(\boldsymbol{Y}_1,\boldsymbol{Y}_0[-j]) = \frac{1}{n_1(n_0-1)}\sum_{1\le i\le n_1}\sum_{\substack{1\le k\le n_0\\ k\ne j}} I(Y_{1i}\ge Y_{0k}).$$

It is apparent that $E(U_i) = E(V_j) = \theta$ for all i and j. It also follows from Shi (1984) that the U_i's and the V_j's are asymptotically independent. Let $\boldsymbol{p} = (p_1,\cdots,p_{n_1})$ and $\boldsymbol{q} = (q_1,\cdots,q_{n_0})$ be such that $\sum_{i=1}^{n_1} p_i = 1, p_i \ge 0$ and $\sum_{j=1}^{n_0} q_j = 1, q_j \ge 0$, assigning probability p_i to U_i and assigning probability q_j to V_j. The two-sample empirical log-likelihood is given by

$$\ell(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n_1}\log p_i + \sum_{j=1}^{n_0}\log q_j,$$

We first consider the following constraints

$$\sum_{i=1}^{n_1} p_i = 1 \quad\text{and}\quad \sum_{j=1}^{n_0} q_j = 1, \tag{3.19}$$
$$\sum_{i=1}^{n_1} p_i U_i = \theta \quad\text{and}\quad \sum_{j=1}^{n_0} q_j V_j = \theta. \tag{3.20}$$

Constraints (3.19) are the normalization constraints and constraints (3.20) are induced by the parameter of interest, θ, using the jackknife pseudo values U_i and V_j. Suppose $\hat{\boldsymbol{p}} = (\hat{p}_1,\cdots,\hat{p}_{n_1})$ and $\hat{\boldsymbol{q}} = (\hat{q}_1,\cdots,\hat{q}_{n_0})$ maximize $\ell(\boldsymbol{p},\boldsymbol{q})$ subject to (3.19) only; $\tilde{\boldsymbol{p}}(\theta) = (\tilde{p}_1(\theta),\cdots,\tilde{p}_{n_1}(\theta))$ and $\tilde{\boldsymbol{q}}(\theta) = (\tilde{q}_1(\theta),\cdots,\tilde{q}_{n_0}(\theta))$ maximize $\ell(\boldsymbol{p},\boldsymbol{q})$ subject to the constraints (3.19) and (3.20), for fixed θ. Then the two-sample jackknife empirical loglikelihood ratio statistic on θ, is defined as

$$r_{\text{JEL2}}(\theta) = \ell(\tilde{\boldsymbol{p}}(\theta),\tilde{\boldsymbol{q}}(\theta)) - \ell(\hat{\boldsymbol{p}},\hat{\boldsymbol{q}}).$$

The pretest responses and other baseline covariates can be incorporated by adding the following additional group of constraints to the maximization process:

$$\sum_{i=1}^{n_1} p_i a_1(\boldsymbol{Z}_i) = \sum_{j=1}^{n_0} q_i a_1(\boldsymbol{Z}_j), \tag{3.21}$$

$$\sum_{i=1}^{n_1} p_i a_0(\boldsymbol{Z}_i) = \sum_{j=1}^{n_0} q_i a_0(\boldsymbol{Z}_j), \tag{3.22}$$

where $a_k(\boldsymbol{z}) = E(Y_k \mid \boldsymbol{Z} = \boldsymbol{z})$ for $k = 1, 0$. If $\boldsymbol{Z}$ is univariate (denoted as Z) and both $a_1(Z)$ and $a_0(Z)$ are linear functions of Z, then (3.21) and (3.22) reduce to a single constraint $\sum_{i=1}^{n_1} p_i Z_i = \sum_{j=1}^{n_0} q_j Z_j$ under the normalization constraints (3.19). Suppose $\hat{\boldsymbol{p}}' = (\hat{p}'_1, \cdots, \hat{p}'_{n_1})$ and $\hat{\boldsymbol{q}}' = (\hat{q}'_1, \cdots, \hat{q}'_{n_0})$ maximize $\ell(\boldsymbol{p}, \boldsymbol{q})$ subject to the normalization constraints (3.19), and the baseline information constraints (3.21) and (3.22); $\tilde{\boldsymbol{p}}'(\theta) = (\tilde{p}'_1(\theta), \cdots, \tilde{p}'_{n_1}(\theta))$ and $\tilde{\boldsymbol{q}}'(\theta) = (\tilde{q}'_1(\theta), \cdots, \tilde{q}'_{n_0}(\theta))$ maximize $\ell(\boldsymbol{p}, \boldsymbol{q})$ subject to the constraints (3.19), (3.21), (3.22), and constraints (3.20), for fixed θ. The jackknife empirical log-likelihood ratio statistic on θ is

$$\tilde{r}_{\text{JEL2}}(\theta) = \ell(\tilde{\boldsymbol{p}}'(\theta), \tilde{\boldsymbol{q}}'(\theta)) - \ell(\hat{\boldsymbol{p}}', \hat{\boldsymbol{q}}').$$

The above formulation of the two-sample EL ratio statistic on θ becomes a special case of the EL inferences on a common mean with multiple samples in the presence of heteroscedasticity. That is, the U_i's and the V_j's have the common mean θ but different variances. Tsao and Wu (2006) and Fu et al. (2009) contain extensive discussions on the EL inferences for a common mean, including a weighted EL approach for multiple samples. Depending on the actual formulation of the EL function (unweighted or weighted) and the type of constraints involved, the asymptotic distribution of the EL ratio statistic is typically a scaled χ^2, with the scaling constant involving unknown population parameters. However, a bootstrap procedure which mimics the original constrained maximization process can be used to bypass the scaling constant. Such a procedure can also be rigorously justified. Our proposed bootstrap calibrated α-level two-sample jackknife EL method which rejects H_0: $F_1 = F_0$ if $-2\tilde{r}_{\text{JEL2}}(\theta_0) > \tilde{b}_\alpha$ for $\theta_0 = 1/2$, where $\tilde{b}_\alpha$ can be obtained by the following procedure: Let $\tilde{\theta}_{\text{JEL2}} = \sum_{i=1}^{n_1} \hat{p}'_i U_i = \sum_{j=1}^{n_0} \hat{q}'_j V_i$.

(1) Select bootstrap samples $s_1^{\#}$ of size n_1 and $s_0^{\#}$ of size n_0 from the original treatment sample and control sample, respectively, using simple random sampling with replacement; denote the two bootstrap sample data sets as $\{(\boldsymbol{Z}_i^{\#}, Y_{1i}^{\#}), i \in s_1^{\#}\}$ and $\{(\boldsymbol{Z}_j^{\#}, Y_{0j}^{\#}), j \in s_0^{\#}\}$;

(2) Construct the jackknife pseudo values $U_i^{\#}$'s and $V_j^{\#}$'s based on the bootstrap samples $s_1^{\#}$ and $s_0^{\#}$; then apply the two-sample JEL method to $U_i^{\#}$'s and $V_j^{\#}$'s and calculate the corresponding empirical log-likelihood ratio statistic $\tilde{r}_{\text{JEL2}}(\theta)$, with $\theta = \tilde{\theta}_{\text{JEL2}}$, denoted as $\tilde{r}_{\text{JEL2}}^{\#}(\tilde{\theta}_{\text{JEL2}})$;

(3) Repeat steps (1) and (2) B times, independently, to obtain the sequence $\{-2\tilde{r}_{\text{JEL2}1}^{\#}(\tilde{\theta}_{\text{JEL2}}), \cdots, -2\tilde{r}_{\text{JEL2}B}^{\#}(\tilde{\theta}_{\text{JEL2}})\}$; let $\tilde{b}_\alpha$ be the $100(1-\alpha)$th sample quantile of the sequence.

3.6 Simulation Studies

In this section, we present the results from simulation studies to evaluate finite sample performance of the methods we discussed in this chapter. We focus on comparing the empirical sizes and the empirical powers of different methods when testing the null hypothesis $H_0 : F_1 = F_0$.

We consider three scenarios (A), (B) and (C). For scenario (A), we only include the posttest responses into the analyses. The methods we considered in scenario (A) are: (i) the standard Mann-Whitney test statistic with asymptotic normality from Section 3.2.1 (MW); (ii) the two-sample EL method with χ_1^2 approximation from Section 3.2.2 (EL); (iii) the jackknife EL method for two-sample U-statistic with χ_1^2 approximation from Section 3.2.3 (JEL); and (iv) the two-sample jackknife EL method with only the normalization constraint (3.19) and constraint (3.20) from Section 3.5 (JEL2). In scenario (B), we incorporate the pretest responses and other baseline covariates into the analyses, and consider the following methods: (i) the adjusted Mann-Whitney test based on the HQF estimators with asymptotic normal distribution from Section 3.3 (HQFMW); (ii) the adjusted Mann-Whitney test based on the HQF estimators with bootstrap calibration from Section 3.3 (HQFMWb); (iii) the two-sample EL method with imputation and bootstrap calibration from Section 3.4 (ELimp); (iv) the JEL method with imputation and bootstrap calibration from Section 3.4 (JELimp); and (v) the two-sample JEL method with constraints (3.19), (3.20), and additional constraints (3.21) and (3.22), which involve the baseline information, from Section 3.5 (JEL2p).

In each simulation study, we generated 1000 simulated data sets and 500 bootstrap samples when the bootstrap method is used. Two simulation models are used to generate the posttest responses for the treatment group (Y_{1i}) and the control group (Y_{0j}) for scenarios (A) and (B). Model (I) is specified as

$$\begin{aligned} Y_{1i} &= \beta_{10} + \beta_{11} X_{1i} + \beta_{12} Z_{1i} + e_{1i}\,, \quad i = 1, \cdots, n_1\,, \\ Y_{0j} &= \beta_{00} + \beta_{01} X_{0j} + \beta_{02} Z_{0j} + e_{0j}\,, \quad j = 1, \cdots, n_0\,, \end{aligned}$$

where X denotes baseline covariate "gender" and Z represents the pretest response. Model (II) includes a nonlinear term and an interaction term:

$$\begin{aligned} Y_{1i} &= \gamma_{10} + \gamma_{11} X_{1i} + \gamma_{12} Z_{1i} + \gamma_{13} Z_{1i}^{1/2} + \gamma_{14} X_{1i} Z_{1i} + e'_{1i}\,, \quad i = 1, \cdots, n_1\,, \\ Y_{0j} &= \gamma_{00} + \gamma_{01} X_{0j} + \gamma_{02} Z_{0j} + \gamma_{03} Z_{0i}^{1/2} + \gamma_{04} X_{0i} Z_{0i} + e'_{0j}\,. \quad j = 1, \cdots, n_0\,. \end{aligned}$$

The X_{i1}'s and X_{0j}'s are generated from a Bernoulli distribution with $p = 0.5$; the Z_{1i}'s and Z_{0j}'s are generated from a standard exponential distribution; the error terms are generated independently as $e_{1i} \sim N(0, \sigma_{e0}^2)$ and $e_{0j} \sim N(0, \sigma_{e1}^2)$ for Model (I), and $e'_{1i} \sim N(0, \sigma^2)$, $e'_{0j} \sim N(0, \sigma^2)$ for Model (II).

For model (I) parameters, we first set $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12})^T = (1, 1, 1.2)^T$ and $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01}, \beta_{02})^T = (1, 1, 1.2)^T$. The error term variances σ_{e1}^2 and σ_{e0}^2 are chosen such that the correlation coefficients between the posttest responses Y_{1i} and Y_{0j} and their linear predictors $\beta_{10}+\beta_{11}X_{1i}+\beta_{12}Z_{1i}$ and $\beta_{00}+\beta_{01}X_{0j}+\beta_{02}Z_{0j}$ are 0.80. This setting corresponds to H_0: $F_1 = F_0$, denoted as Case 1. We further consider Case 2 to Case 5, where we reset the values of β_{12} as 1.7, 2.2, 2.7, 3.2, respectively. Those cases represent different degrees of departure from the H_0. For model (II) parameters, we also consider five cases of different combinations of $\boldsymbol{\gamma}_1 = (\gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{14})^T$ and $\boldsymbol{\gamma}_0 = (\gamma_{00}, \gamma_{01}, \gamma_{02}, \gamma_{03}, \gamma_{04})^T$. We set $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_0 = (1, 0.5, 0.5, -1.5, 2)^T$ for Case 1, and set $\gamma_{10} = 1.5, 2, 2.5, 3$ from Case 2 to Case 5, respectively. The error term variances are simply chosen as $\sigma^2 = 4$. We consider sample sizes for the treatment group and the control group to be $n_1 = n_0 = 45$ for each simulation.

Scenario (C) considers model misspecifications for the HQF method and the imputation-based methods. In scenario (C), the true model is Model (II) which includes a nonlinear term and an interaction term. For the HQFMW method and the two-sample JEL method, we assume:

$$a_k(x_{ki}, z_{ki}) = a_{k0} + a_{k1}x_{ki} + a_{k2}z_{ki}, \quad \text{where } k = 1, 0.$$

For the imputation based methods, we assume the working regression models to be:

$$E(y_{ki}|x_{ki}, z_{ki}) = a_{k0} + a_{k1}x_{ki} + a_{k2}z_{ki}, \quad \text{where } k = 1, 0.$$

Simulation results of scenario (A), (B) and (C) are summarized in Table 3.1, Table 3.2, and Table 3.3 respectively. In each table, the empirical size and the power of the tests with 5% nominal significance are reported. The empirical size of the test for each method is listed in the column of Case 1 in each table. In the column of Case 2 to Case 5, we record the empirical power of the tests based on each method. From Table 3.1, we can see that, under both Model (I) and Model (II), the empirical sizes of the tests are similar and around 5% for all the methods. We also notice that the empirical power of the tests based on the standard Mann-Whitney statistic are slightly higher than those of the other three methods for every case. The empirical power of the tests based on the EL, JEL, and JEL2 method are close under the case of independent data.

From Table 3.2, under both Model (I) and Model (II), we can see that the empirical size of the tests of both HQFMW and HQFMWb are very close to 5%. However, the sizes of the tests based on the EL and JEL method with imputation are slightly lower than 5%. The empirical size of the test of JEL2p is equal to 5% under Model (I), but a little lower than 5% under model (II). The HQFMW method outperforms the ELimp, JELimp, and the JEL2p method in terms of larger empirical powers through Case 2 to Case 5 under both models. The empirical power of the tests based on JEL2p are lower than the ones of the HQFMW methods, yet are higher than the ones of the imputation based methods. Such results suggest we should reconsider the imputation technique and investigate more on the imputation methods which preserve the distribution of the observed response data in the future work. An important observation from Table 3.2 and Table 3.1 is that incorporating the baseline information increases the power of the test for all the methods considered for all cases.

Table 3.3 reports the empirical size and power of the tests under model misspecification. We want to look at the performance of each test by comparing the numbers inside Table 3.3, and additionally, we compare Table 3.3 to the second portion of Table 3.2 where we have correctly specified working models for each method. In Table 3.3, for the HQF adjusted Mann-Whitney methods, the empirical sizes of the tests are close to 5%. The empirical sizes of the tests based on both imputation based EL and JEL methods are further away from 5% than their counterparts in Table 3.2. This is what we are expecting since the stochastic regression imputation is not robust against misspecification. The empirical size of the test based on the two-sample JEL method is fairly close to 5%. The empirical power of the tests of HQFMW and HQFMWb are the largest compared to those of the ELimp, JELimp and the JEL2p method for every case in Table 3.3. Furthermore, if we compare the numbers in Table 3.3 to those in Table 3.2, we notice that the power of the tests for both the HQF adjusted Mann-Whitney methods does not drop a great deal, and a similar conclusion is found for the two-sample JEL method.

Table 3.1: Scenario (A): Empirical Power of Testing $H_0 : F_1 = F_0$

Model	(n_1, n_0)	Method	Case1	Case2	Case3	Case4	Case5
Model (I)	(45,45)	MW	0.047	0.181	0.439	0.637	0.767
		EL	0.050	0.183	0.425	0.602	0.722
		JEL	0.059	0.176	0.414	0.592	0.729
		JEL2	0.049	0.185	0.409	0.568	0.689
Model (II)	(45,45)	MW	0.055	0.156	0.434	0.773	0.951
		EL	0.057	0.157	0.432	0.758	0.942
		JEL	0.052	0.138	0.412	0.765	0.947
		JEL2	0.046	0.135	0.437	0.761	0.946

3.7 Concluding Remarks

Chen et al. (2013) studied the Mann-Whitney test with covariate adjustments for missing data and observational study. They considered a kernel estimator of the conditional distribution function after accommodating the missingness by inverse response probabilities and constructed the adjusted Mann-Whitney test statistic using the kernel estimators of the distribution functions. However, their proposed method does not apply directly to the settings considered in this chapter.

In Chapter 2, we have shown that the imputation based approach is very efficient for the estimation of the treatment effect for pretest-posttest studies. Our simulation results show that the approach is not efficient for constructing Mann-Whitney test for the difference of two distribution functions. This is probably due to the fact that the imputed values retain the mean responses, i.e., $E(\tilde{Y}_{ki}) = E(Y_{ki})$, but do not necessarily restore the distribution functions. In other words, we have $E\{I(\tilde{Y}_{1i} > \tilde{Y}_{0i})\} \neq E\{I(Y_{1i} > Y_{0i})\}$ even under the stochastic regression imputation with a true model.

The two-sample jackknife empirical likelihood method for the Mann-Whitney test is promising, due to its less demanding computational procedures and the flexibility in incorporating baseline information. It is related to the common mean problem previously

Table 3.2: Scenario (B): Empirical Power of Testing $H_0 : F_1 = F_0$

Model	(n_1, n_0)	Method	Case1	Case2	Case3	Case4	Case5
Model (I)	(45,45)	HQFMW	0.049	0.271	0.604	0.771	0.862
		HQFMWb	0.052	0.259	0.599	0.762	0.868
		ELimp	0.036	0.239	0.544	0.744	0.815
		JELimp	0.037	0.247	0.542	0.719	0.805
		JEL2p	0.050	0.266	0.568	0.755	0.826
Model (II)	(45,45)	HQFMW	0.057	0.180	0.564	0.902	0.995
		HQFMWb	0.050	0.155	0.528	0.884	0.994
		ELimp	0.037	0.137	0.485	0.803	0.967
		JELimp	0.035	0.137	0.487	0.814	0.969
		JEL2p	0.038	0.167	0.528	0.854	0.977

Table 3.3: Scenario (C): Empirical Power of Testing $H_0 : F_1 = F_0$

Model	(n_1, n_0)	Method	Case1	Case2	Case3	Case4	Case5
Model (II)	(45,45)	HQFMW	0.049	0.169	0.535	0.863	0.986
		HQFMWb	0.044	0.157	0.516	0.845	0.987
		ELimp	0.026	0.131	0.428	0.741	0.931
		JELimp	0.026	0.131	0.412	0.744	0.935
		JEL2p	0.048	0.156	0.512	0.837	0.974

discussed by Tsao and Wu (2006) and Fu et al. (2009). The weighted empirical likelihood method is shown to be efficient for estimating the common mean with multiple samples. Using the approach for the Mann-Whitney test with samples from pretest-posttest studies is currently under investigation.

Chapter 4

Empirical Likelihood Method for Pretest-Posttest Studies under Complex Survey Design

4.1 Introduction

In this chapter, we extend our discussion of the empirical likelihood method for pretest-posttest study to the context of complex survey data. In the literature of empirical likelihood, Chen and Qin (1993) first applied the empirical likelihood method to the field of sample surveys with available auxiliary information, where they assumed that the sampling design is simple random sampling without replacement. For a more general sampling design, Chen and Sitter (1999) proposed a pseudo-empirical likelihood approach to account for the effect of sampling from a finite population. Their idea was to weight the standard empirical likelihood as in the Horvitz-Thompson estimator (Horvitz and Thompson (1952)). However, they only focused on the point estimation. Wu and Rao (2006) proposed a slightly different pseudo-empirical likelihood approach, and further developed the asymptotic distribution of the pseudo-empirical log likelihood ratio statistics for constructing confidence intervals and conducting hypothesis tests based on a single survey sample, under several sampling designs. We notice that in the absence of auxiliary information, the pseudo MLE of Chen and Sitter (1999) is identical to the Horvitz-Thompson estimator, while that of Wu and Rao (2006) is reduced to the Hájek estimator (Hájek (1971)). Wu and Yan (2012) extended Wu and Rao (2006) to the two-sample problem setting where the two samples are assumed to be sampled from two different finite populations. Moreover, we refer to Rao and Wu (2009) for an excellent overview of empirical likelihood method in survey sampling.

The International Tobacco Control (ITC) Policy Evaluation Project Four Country (4C) Survey is a prospective cohort study designed to evaluate the psychosocial and behavioural impact of key national-level tobacco control policies enacted over a period of eleven years (2002-2014), in at least one of four countries: the United States, Canada, the United Kingdom, and Australia. Over $2,000$ adult smokers were recruited by probability sampling methods in each of the four countries at the initial cohort. At each subsequent wave (approximately annual intervals for a period of eleven years) the sample was formed by recontact of earlier respondents and replenishment with new respondents to ensure that there were approximately $2,000$ in each country who completed the survey. The sampling design of the ITC 4C survey is random sampling within strata which are defined by geographic region and community size in each country. The ITC 4C Survey was developed by an interdisciplinary team of tobacco control experts across the four countries, with backgrounds in psychology, public health, epidemiology, economics, community medicine, marketing, sociology, and statistics/biostatistics. The questions of the ITC 4C survey are from the following domains: demographic variables, smoking behaviour, warning labels, advertising and promotion, light/mild brand descriptors, taxation and purchase behaviour, stop-smoking medications and alternative nicotine products, cessation and quitting behaviour as well as key psychosocial measures. All aspects of the study protocol and survey measures are standardized across the four countries. For more details about the sampling methods, survey protocol and administration, and other related information, we refer to the ITC Four Country Wave 1 Technical Report (2004) and the ITC Four Country Wave 2-8 Technical Report (2011).

At Wave 7 of the ITC 4C survey, an embedded pilot study was conducted to evaluate whether an online version of the survey would be a viable option for further waves in the ITC 4C survey. More specifically, the study was to determine the amount of cost savings that could be achieved if some of the cohort participants completed the survey on-line, and to determine whether some of the people could be reached that might otherwise be lost. After the pilot study, it was decided that the web survey option would be offered to all respondents starting from Wave 8. The ITC 4C Wave 8 Recontact Survey employed a mixed mode approach, combining web and phone data collection. Each recontact respondent of Wave 8 received either an email invitation (given he/she provided an email address at Wave 7) or a mailed letter invitation to respond online. Among 5135 recontact respondents of Wave 8, 2006 (39%) answered by web survey. Our objective is to assess whether the distributions of the responses to certain questions from people using the web survey are different from those of people using the telephone survey. Analyzing such differences is of interest because the answering mode might affect the answer to certain types of question. For example, phone respondents might tend to give the last response option because they heard it most recently; however, web respondents can see all response options at once. Also, for instance, some people may be uncomfortable answering certain types of questions in front of the interviewer over the telephone, but wouldn't have such an issue if answering over the internet. By adopting the setting of the pretest-posttest study, we consider the data of those 5135 participants at Wave 7 as the baseline information ("pretest responses"), the data at Wave 8 as the "posttest responses", and the web survey mode as the "treatment". We want to develop the methods for estimating the treatment effect while accommodating the survey design.

Unlike the randomized design of the pretest-posttest study we introduced in the previous two chapters, the recontact respondents in Wave 8 were self-selected to one of the treatment groups, i.e. answering by web or phone. Without the benefit of randomization, treatment groups may differ systematically with respect to relevant characteristics, and thus, may not be directly comparable (Rosenbaum and Rubin (1984)). In a standard pretest-posttest design, with randomization, distributions of covariates are balanced across treatment groups. However, in an observational study, treatment exposure may be associated with covariates which are also associated with the potential responses. Therefore, in order to make inference on the treatment effect for observational data, methods are required to adjust for confounding of exposure to treatment with subject characteristics (Lunceford and Davidian (2004)). Methods based on propensity score modelling are becoming increasingly popular for making causal inference with observational data. The propensity score is defined as the probability of treatment exposure conditional on observed baseline covariates (Rosenbaum and Rubin (1983)). Rosenbaum and Rubin (1984) considered stratification or subclassification based on the estimated propensity score and estimating the treatment effect as the average of within-stratum effects. Methods based on propensity score matching are also popular in the medical literature (Austin (2008)). There is another alternative class of estimators discussed in (Rosenbaum (1998), Lunceford and Davidian (2004)), which are constructed by inverse weighting the estimated propensity score in the fashion of the Horvitz-Thompson estimator. The authors of Lunceford and Davidian (2004) also identified the most efficient semiparametric estimator based on a propensity score weighting method using the theory of Robins et al. (1994).

In this chapter, our objective is to propose an estimator of treatment effect based on propensity score stratification under survey data and derive its variance estimation. Alternatively, we want to develop an estimator based on empirical likelihood (EL) method by constructing weights for each subject using their estimated propensity score, and applying the two-sample pseudo empirical likelihood method by Wu and Yan (2012). The rest of the chapter is organized as follows. In Section 4.2, we derive the estimator based on the propensity score stratification and its variance estimation. In Section 4.3, we propose the estimator based on the two-sample pseudo EL method. An application of the proposed methods to the ITC 4C survey data will be given in Section 4.4. In Section 4.5, we consider the two-sample pseudo EL method under a simpler setting where we have a randomized study with survey data. Some concluding remarks are given in Section 4.6.

4.2 Estimator Based on Propensity Score Stratification and Its Variance Estimation

4.2.1 The Propensity Score and Stratification

Adopting the counterfactual framework, we let Y_1 and Y_0 be the responses an individual potentially would exhibit if he/she receives treatment and control, respectively. Y_1 and Y_0 will never be observed simultaneously. Let R be the indicator of treatment exposure ($R = 1$ if in treatment, and $R = 0$ if in control). The treatment effect can be expressed by

$$\theta = \mu_1 - \mu_0 = E(Y_1) - E(Y_0),$$

where the expectation of $Y_1(Y_0)$ is taken with respect to the hypothetical distribution of the potential response $Y_1(Y_0)$. In a randomized study, the potential outcomes (Y_1, Y_0) are statistically independent of the treatment assignment R, or $(Y_1, Y_0) \perp\!\!\!\perp R$. Therefore, the treatment effect θ can be identified from the observed data since $E(Y_1|R = 1) = E(Y_1)$ and $E(Y_0|R = 0) = E(Y_0)$. However, in an observational study, the treatment exposure R is no longer necessarily independent of (Y_1, Y_0), and some subject characteristics may be associated with both treatment exposure and the potential responses. Assume X contains all possible confounders which are associated with both treatment exposure and the potential responses. Then, we have

$$(Y_1, Y_0) \perp\!\!\!\perp R | X,$$

the assumption of strongly ignorable treatment assignment (Rosenbaum and Rubin (1983)). Under this assumption, θ can be identified from the observed data since

$$E\{E(Y_m|R = m, X)\} = E\{E(Y_m|X)\} = E(Y_m), \quad m = 1, 0.$$

The propensity score is the conditional probability that an individual will be exposed to treatment given observed covariates X, and mathematically:

$$e(X) = P(R = 1|X).$$

Rosenbaum and Rubin (1983) proved that subclassification on the propensity score will balance X, i.e. the distribution of X is the same for treated and untreated subjects within subclasses that are homogeneous in propensity score $e(X)$. In practice, the true propensity score is often unknown; therefore, it is common to estimate it based on the observed data by assuming a logistic regression model:

$$e(X, \boldsymbol{\beta}) = \{1 + exp(-X^T \boldsymbol{\beta})\}^{-1}.$$

The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ can be obtained by solving:

$$\sum_{i=1}^{n} \frac{Z_i - e(X_i, \boldsymbol{\beta})}{e(X_i, \boldsymbol{\beta})(1 - e(X_i, \boldsymbol{\beta}))} \partial/\partial\boldsymbol{\beta}(e(X_i, \boldsymbol{\beta})) = 0,$$

(Lunceford and Davidian (2004)).

Plugging the MLE $\hat{\boldsymbol{\beta}}$ into $e(X, \boldsymbol{\beta})$ leads to the estimated propensity score $\hat{e}_i(X_i, \hat{\boldsymbol{\beta}})$. Then the sample can be divided into K subsamples (termed "strata") according to the sample quantiles of the $\hat{e}_i$. Specifically, let $\hat{q}_k, \quad k = 1, \cdots, K$ be the k-th sample quantile of the estimated propensity score such that the proportion of $\hat{e}_i \leq \hat{q}_k$ is roughly k/K, and $\hat{q}_0 = 0$, $\hat{q}_K = 1$. Then we define the subsample Q_k to be:

$$Q_k = \{i : \hat{e}_i \in (\hat{q}_{k-1}, \hat{q}_k]\}, \quad k = 1, \cdots, K.$$

Rosenbaum and Rubin (1983, 1984) suggested the use of quintiles, i.e. $K = 5$, which is a popular choice of K in much of the literature of propensity score stratification. If the $\hat{e}_i$'s are good approximations of the true propensity scores, then within each Q_k, the treatment exposure is approximately random, and the distribution of X is approximately the same for treated and untreated units. Because the approximation may be imperfect, it is important to assess the balance of the covariates achieved by stratification based on the estimated propensity score. For balance diagnosis under propensity score stratification, we treat each of the covariates contained in X as the response which is subject to a two-way (2 treatments $\times$ K strata) analysis of variance. Large values of F ratios for the main (treatment) effects or for the two-way interaction suggest inadequate fit of the propensity score model (Rosenbaum and Rubin (1984)). Graphic tools such as a box-plot of each covariate against treatment within Q_k is also a useful way of diagnosing departures from balance. For more details and illustration on balance diagnosis under propensity score stratification, we refer to Rosenbaum and Rubin (1983) and Rosenbaum and Rubin (1984). For balance diagnosis under other propensity score methods, e.g. propensity score matching, we refer to Austin (2009).

4.2.2 Estimator of θ

Now we introduce some notations in the setting of the ITC 4C survey data. Let h index the country, and N_h be the finite population size of country h; then we have overall a finite population $\{1, \cdots, N\}$, where $N = \sum_h N_h$. Let Y_{1i} be the potential outcome that an individual i from the finite population will exhibit if he/she is exposed to the treatment, and Y_{0i} be the potential outcome that an individual i from the finite population will exhibit if he/she is exposed to the control; then the population-level parameter of interest is $\theta = E(Y_1) - E(Y_0)$. We denote the set of sampled units as $s = \{i : i \in \text{ sample}\}$, which is a combined set of sampled units from each country h, that is the union over h of sets $s_h = \{i : i \in \text{ sample } \ \& \ \ i \in h\}$. Let n be the sample size, and n_h be the number of subjects sampled from country h, where $n = \sum_h n_h$. Suppose $d_i = 1/\pi_i$ is the basic design (or inflation) weight, with $\pi_i = P(i \in s)$ being the inclusion probability of subject i. Then we define the rescaled weights to be:

$$w_i = \frac{d_i n_h}{\hat{N}_h} = \frac{d_i n_h}{\sum_{i \in s_h} d_i}.$$

After we fit an adequate propensity model to the sample data, and we obtain the estimated propensity score for each subject $i \in s$, then we subclassify the sample into K subsamples according to the sample quantile of estimated propensity score, where each subsample $Q_k = \{i \in s : \hat{e}_i \in (\hat{q}_{k-1}, \hat{q}_k]\}$. Let $H_{ik} = 1$ if unit i is in Q_k, and $H_{ik} = 0$ otherwise. Then the sample-level estimator of θ based on propensity score stratification can be written as a weighted sum of the difference of the sample mean of observed Y_1 and Y_0 within subsample Q_k, formally:

$$\hat{\theta}_{str} = \sum_{k=1}^{K} \left(\frac{\sum_{i \in s} H_{ik} w_i}{\sum_{i \in s} w_i} \right) \left\{ \frac{\sum_{i \in s} H_{ik} R_i w_i Y_{1i}}{\sum_{i \in s} H_{ik} R_i w_i} - \frac{\sum_{i \in s} H_{ik}(1 - R_i) w_i Y_{0i}}{\sum_{i \in s} H_{ik}(1 - R_i) w_i} \right\} \tag{4.1}$$

Let

$$\hat{\mu}_{1k} = \frac{\sum_{i \in s} H_{ik} R_i w_i Y_{1i}}{\sum_{i \in s} H_{ik} R_i w_i} \tag{4.2}$$

$$\hat{\mu}_{0k} = \frac{\sum_{i \in s} H_{ik}(1 - R_i) w_i Y_{0i}}{\sum_{i \in s} H_{ik}(1 - R_i) w_i}, \tag{4.3}$$

then $\hat{\mu}_{1k} - \hat{\mu}_{0k}$ estimates the difference of the population-level average of observed Y_1 and Y_0 within the population domain covered by the k-th sample quantile of the estimated propensity score. The factor $\sum_{i \in s}(H_{ik} w_i)/\sum_{i \in s} w_i$ estimates the proportion of the population covered by the k-th sample quantile of the estimated propensity score.

In the next subsection, we look at the variance estimation of $\hat{\theta}_{str}$ from both model-based and design-based perspectives.

4.2.3 Variance Estimations

Model-based variance estimator

Define

$$\hat{\theta}_k = \hat{\mu}_{1k} - \hat{\mu}_{0k} = \frac{\sum_{i\in s} H_{ik} R_i w_i Y_{1i}}{\sum_{i\in s} H_{ik} R_i w_i} - \frac{\sum_{i\in s} H_{ik}(1-R_i) w_i Y_{0i}}{\sum_{i\in s} H_{ik}(1-R_i) w_i};$$

then

$$\hat{\theta}_{str} = \sum_{k=1}^{K} \left(\frac{\sum_{i\in s} H_{ik} w_i}{\sum_{i\in s} w_i} \right) \hat{\theta}_k.$$

We assume that within each Q_k, $E_{ps}(R_i \mid w_{.}, Y_{1.}, Y_{0.})$ is approximately a constant δ_k (the quantile average value), where E_{ps} denotes expectation taken with respect to the mechanism of "choosing" the treatment, i.e. expectation with respect to the treatment determination. If E_p denotes the expectation with respect to the sampling design, assuming $Y_{1.}$ and $Y_{0.}$ fixed, and E_{psc} is $E_{ps}(. \mid w_{.}, Y_{1.}, Y_{0.})$, then we have

$$E_p E_{psc} \sum_{i\in s} H_{ik} R_i w_i Y_{1i} \simeq \delta_k \sum_h \alpha_h T_{1hk},$$

where T_{1hk} is the population total of Y_1 in the k-th propensity score quantile (a domain) in country h, and $\alpha_h = n_h/N_h$. Similarly,

$$E_p E_{psc} \sum_{i\in s} H_{ik} R_i w_i \simeq \delta_k \sum_h \alpha_h N_{hk},$$

$$E_p E_{psc} \sum_{i\in s} H_{ik}(1-R_i) w_i Y_{1i} \simeq (1-\delta_k) \sum_h \alpha_h T_{0hk},$$

$$E_p E_{psc} \sum_{i\in s} H_{ik}(1-R_i) w_i \simeq (1-\delta_k) \sum_h \alpha_h N_{hk},$$

where T_{0hk} is the population total of y_0 in the k-th propensity score quantile in country h, and N_{hk} is the size of the domain with propensity score in the k-th quantile in country h. Moreover,

$$E_p E_{psc} \sum_{i\in s} H_{ik} w_i \simeq \sum_h \alpha_h N_{hk},$$

$$E_p E_{psc} \sum_{i \in s} w_i \simeq \sum_h \alpha_h N_h.$$

If E_ξ denotes the expectation taken with respect to the superpopulation model for Y_1 and Y_0, we write $E_\xi T_{1hk} \simeq N_{hk}\mu_{1hk}$ and $E_\xi T_{0hk} \simeq N_{hk}\mu_{0hk}$, where μ_{1hk} and μ_{0hk} are the population mean of Y_1 and Y_0 respectively in the domain represented by the k-th propensity score sample quantile in country h. Consider the ultimate estimand to be

$$\theta_N = \sum_{k=1}^{K} \left(\frac{\sum_h \alpha_h N_{hk}}{\sum_h \alpha_h N_h} \right) \{\mu_{1k} - \mu_{0k}\},$$

where

$$\mu_{1k} = \frac{\sum_h \alpha_h N_{hk} \mu_{1hk}}{\sum_h \alpha_h N_{hk}},$$

and

$$\mu_{0k} = \frac{\sum_h \alpha_h N_{hk} \mu_{0hk}}{\sum_h \alpha_h N_{hk}},$$

thus,

$$\theta_N = \sum_{k=1}^{K} \frac{\sum_h \alpha_h N_{hk}\mu_{1hk} - \sum_h \alpha_h N_{hk}\mu_{0hk}}{\sum_h \alpha_h N_h}.$$

The mean μ_{1k} is the expectation of Y_1, given all covariates, averaged over the distribution of covariates that pertains in Q_k, which happens to be about the same in the two treatment groups by the balancing property of the propensity score. Then we have

$$\mu_{1k} \simeq \frac{\sum_{i \in s} H_{ik} R_i w_i \mu_{1i}}{\sum_{i \in s} H_{ik} R_i w_i}$$

where μ_{1i} is the E_ξ expectation of Y_{1i}, given all covariates. Similarly for μ_{0k}. Then, the error of $\hat{\theta}_k$ is

$$\frac{\sum_{i \in s} H_{ik} R_i w_i Y_{1i}}{\sum_{i \in s} H_{ik} R_i w_i} - \frac{\sum_{i \in s} H_{ik} (1 - R_i) w_i Y_{0i}}{\sum_{i \in s} H_{ik} (1 - R_i) w_i} - \mu_{1k} + \mu_{0k},$$

which is approximately

$$\frac{\sum_{i \in s} H_{ik} R_i w_i (Y_{1i} - \mu_{1i})}{\sum_{i \in s} H_{ik} R_i w_i} - \frac{\sum_{i \in s} H_{ik} (1 - R_i) w_i (Y_{0i} - \mu_{0i})}{\sum_{i \in s} H_{ik} (1 - R_i) w_i}.$$

This can be shown to have expectation 0 with respect to E_ξ conditional on the sample and the R_i, and so the MSE can be estimated through an estimate of the approximate MSE with respect to ξ conditional on the sample and the R_i. This conditional MSE is

$$\frac{\sum_{i\in s} H_{ik} R_i w_i^2 \sigma_{1i}^2}{(\sum_{i\in s} H_{ik} R_i w_i)^2} + \frac{\sum_{i\in s} H_{ik}(1-R_i) w_i^2 \sigma_{0i}^2}{(\sum_{i\in s} H_{ik}(1-R_i) w_i)^2},$$

where σ_{1i}^2 is the variance of Y_{1i}, given the covariates. Thus the estimated MSE of $\hat{\theta}_k$ is

$$\hat{V}_k = \frac{\sum_{i\in s} H_{ik} R_i w_i^2 \hat{\sigma}_{1i}^2}{(\sum_{i\in s} H_{ik} R_i w_i)^2} + \frac{\sum_{i\in s} H_{ik}(1-R_i) w_i^2 \hat{\sigma}_{0i}^2}{(\sum_{i\in s} H_{ik}(1-R_i) w_i)^2},$$

where $\hat{\sigma}_{1i}^2$ is the square of the residual for i of the regression of Y_{1i} on the covariates in quantile k (weighted or unweighted), and $\hat{\sigma}_{0i}^2$ is defined similarly. The estimated MSE of $\hat{\theta}_{str}$ could be

$$\hat{V} = \sum_{k=1}^{K} \left(\frac{\sum_{i\in s} H_{ik} w_i}{\sum_{i\in s} w_i} \right)^2 \hat{V}_k.$$

Design-based variance estimator

To facilitate the argument, again let H_{ik} be the indicator whether subject i belongs in the subsample Q_k, which is determined based on the sample quantiles of the estimated propensity score. Also for simplicity, assume that $\hat{N}_h = N_h$. In the sense of the sampling design, if

$$\hat{\mu}_{1k} = \frac{\sum_{i\in s} R_i H_{ik} w_i Y_{1i}}{\sum_{i\in s} R_i H_{ik} w_i},$$

then $\hat{\mu}_{1k}$ is approximately unbiased for

$$\mu_{1kN} = \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik} Y_{1i}}{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik}}.$$

Notice that we can write the numerator of $\hat{\mu}_{1k}$ in a Hansen-Hurwitz form as

$$\sum_{i\in Q_k} R_i w_i Y_{1i} = \sum_{i\in s} R_i H_{ik} w_i Y_{1i} = \sum_h \alpha_h \sum_{i\in s_h} d_i R_i H_{ik} y_{1i} = \sum_h \alpha_h \frac{1}{n_h} \sum_{i\in s_h} \frac{R_i H_{ik} y_{1i}}{z_i}$$

where the size measure $z_i = 1/(d_i n_h)$, d_i is the basic design (or inflation) weight, which is the reciprocal of the inclusion probability π_i, s_h is the sample of units from country h, and n_h is the size of s_h. Similarly, for the control group,

$$\hat{\mu}_{0k} = \frac{\sum_{i\in s}(1-R_i)H_{ik}w_iY_{0i}}{\sum_{i\in s}(1-R_i)H_{ik}w_i};$$

then $\hat{\mu}_{0k}$ is approximately unbiased for

$$\mu_{0kN} = \frac{\sum_h \alpha_h \sum_{i=1}^{N_h}(1-R_i)H_{ik}Y_{0i}}{\sum_h \alpha_h \sum_{i=1}^{N_h}(1-R_i)H_{ik}}.$$

Now $\hat{\theta}_k = \hat{\mu}_{1k} - \hat{\mu}_{0k}$ is approximately design-unbiased for

$$\theta_{kN} = \mu_{1kN} - \mu_{0kN} = \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} R_iH_{ik}Y_{1i}}{\sum_h \alpha_h \sum_{i=1}^{N_h} R_iH_{ik}} - \frac{\sum_h \alpha_h \sum_{i=1}^{N_h}(1-R_i)H_{ik}Y_{0i}}{\sum_h \alpha_h \sum_{i=1}^{N_h}(1-R_i)H_{ik}}.$$

As we did when discussing the model-based variance estimation, if we assume that within the domain k, the expectation of R_i given other variables is δ_k (this assumption is not used later when we derive the design-based variance estimator), then $\hat{\theta}_k = \hat{\mu}_{1k} - \hat{\mu}_{0k}$ estimates

$$\frac{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}Y_{1i}}{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}} - \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}Y_{0i}}{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}}.$$

Now

$$\frac{\sum_{i\in s} H_{ik}w_i}{\sum_{i\in s} w_i}$$

estimates

$$\frac{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}}{\sum_h \alpha_h N_h},$$

the proportion of the population covered by the k-th sample quantile. So it can be argued that $\hat{\theta}_{str}$ estimates a finite population quantity that approximates

$$\frac{\sum_h \alpha_h \sum_{i=1}^{N_h} Y_{1i} - \sum_h \alpha_h \sum_{i=1}^{N_h} Y_{0i}}{\sum_h \alpha_h N_h},$$

which essentially is the difference of the population mean of Y_1 and Y_0.

Next we develop the design-based variance estimation. The estimator for the overall difference is given by:

$$\begin{aligned}\hat{\theta}_{str} &= \sum_{k=1}^{K}\left(\frac{\sum_{i\in s} H_{ik}w_i}{\sum_{i\in s} w_i}\right)\hat{\theta}_k \\ &= \sum_{k=1}^{K}\left\{\frac{\sum_{i\in s} H_{ik}w_i}{\sum_{i\in s} w_i}\cdot\frac{\sum_{i\in s} R_iH_{ik}w_iY_{1i}}{\sum_{i\in s} R_iH_{ik}w_i} - \frac{\sum_{i\in s} H_{ik}w_i}{\sum_{i\in s} w_i}\cdot\frac{\sum_{i\in s}(1-R_i)H_{ik}w_iY_{0i}}{\sum_{i\in s}(1-R_i)H_{ik}w_i}\right\}\end{aligned}$$

We first look at the variance estimation of the first component

$$\frac{\sum_{i\in s} H_{ik}w_i}{\sum_{i\in s} w_i}\cdot\frac{\sum_{i\in s} R_iH_{ik}w_iY_{1i}}{\sum_{i\in s} R_iH_{ik}w_i},$$

and we notice this is a product of two ratio estimators. Let $\hat{a}_1/\hat{a}_0$ and $\hat{b}_1/\hat{b}_2$ be two ratio estimators which estimate a_1/a_2 and b_1/b_2 respectively. By linearization we have,

$$\begin{aligned}\frac{\hat{a}_1}{\hat{a}_2} &\doteq \frac{a_1}{a_2} + \frac{1}{a_2}(\hat{a}_1 - \frac{a_1}{a_2}\hat{a}_2), \\ \frac{\hat{b}_1}{\hat{b}_2} &\doteq \frac{b_1}{b_2} + \frac{1}{b_2}(\hat{b}_1 - \frac{b_1}{b_2}\hat{b}_2).\end{aligned}$$

In our setting,

$$\begin{aligned}\hat{a}_1 &= \sum_{i\in s} H_{ik}w_i, \quad \hat{a}_2 = \sum_{i\in s} w_i \\ a_1 &= \sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}, \quad a_2 = \sum_h \alpha_h N_h, \\ \hat{b}_1 &= \sum_{i\in s} R_iH_{ik}w_iY_{1i}, \quad \hat{b}_2 = \sum_{i\in s} R_iH_{ik}w_i, \\ b_1 &= \sum_h \alpha_h \sum_{i=1}^{N_h} R_iH_{ik}Y_{1i}, \quad b_2 = \sum_h \alpha_h \sum_{i=1}^{N_h} R_iH_{ik}.\end{aligned}$$

Now, the product of the two ratio estimators is:

$$\frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{b}_1}{\hat{b}_2} \doteq \frac{a_1}{a_2}\cdot\frac{b_1}{b_2} + \frac{b_1}{b_2}\cdot\frac{1}{a_2}(\hat{a}_1 - \frac{a_1}{a_2}\hat{a}_2) + \frac{a_1}{a_2}\cdot\frac{1}{b_2}(\hat{b}_1 - \frac{b_1}{b_2}\hat{b}_2) + \frac{1}{a_2}\cdot\frac{1}{b_2}(\hat{a}_1 - \frac{a_1}{a_2}\hat{a}_2)(\hat{b}_1 - \frac{b_1}{b_2}\hat{b}_2).$$

In the above formula, the initial (constant) term is $O(1)$, and the next two terms should each be $O_p(n^{-1/2})$. It can then be shown that the last term is $O_p(n^{-1})$, and thus can be neglected in the above formula. Now the above formula becomes:

$$\frac{\hat{a}_1}{\hat{a}_2} \cdot \frac{\hat{b}_1}{\hat{b}_2} \doteq \frac{a_1}{a_2} \cdot \frac{b_1}{b_2} + \frac{b_1}{b_2} \cdot \frac{1}{a_2}(\hat{a}_1 - \frac{a_1}{a_2}\hat{a}_2) + \frac{a_1}{a_2} \cdot \frac{1}{b_2}(\hat{b}_1 - \frac{b_1}{b_2}\hat{b}_2),$$

which can be written as:

$$\begin{aligned}
\frac{\hat{a}_1}{\hat{a}_2} \cdot \frac{\hat{b}_1}{\hat{b}_2} - \frac{a_1}{a_2} \cdot \frac{b_1}{b_2} &= \frac{\sum_{i\in s} H_{ik} w_i}{\sum_{i\in s} w_i} \cdot \frac{\sum_{i\in s} R_i H_{ik} w_i Y_{1i}}{\sum_{i\in s} R_i H_{ik} w_i} - \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}}{\sum_h \alpha_h N_h} \cdot \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik} Y_{1i}}{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik}} \\
&\doteq \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik} Y_{1i}}{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik}} \cdot \frac{1}{\sum_h \alpha_h N_h} \left(\sum_{i\in s} w_i H_{ik} - \frac{a_1}{a_2} \sum_{i\in s} w_i \right) + \\
&\quad \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}}{\sum_h \alpha_h N_h} \cdot \frac{1}{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik}} \left(\sum_{i\in s} R_i H_{ik} w_i Y_{1i} - \frac{b_1}{b_2} \sum_{i\in s} R_i H_{ik} w_i \right) \\
&= \frac{b_1}{b_2} \cdot \frac{1}{a_2} \left(\sum_{i\in s} w_i H_{ik} - \frac{a_1}{a_2} \sum_{i\in s} w_i \right) + \frac{a_1}{a_2} \cdot \frac{1}{b_2} \left(\sum_{i\in s} R_i H_{ik} w_i Y_{1i} - \frac{b_1}{b_2} \sum_{i\in s} R_i H_{ik} w_i \right) \\
&= \frac{b_1}{b_2} \cdot \frac{1}{a_2} \sum_{i\in s} w_i \gamma_i + \frac{a_1}{a_2} \cdot \frac{1}{b_2} \sum_{i\in s} w_i R_i H_{ik} u_{1i} \\
&= \frac{1}{a_2 b_2} \sum_{i\in s} w_i \left\{ b_1 \gamma_i + a_1 R_i H_{ik} u_{1i} \right\} = \frac{1}{a_2 b_2} \sum_h \alpha_h \frac{1}{n_h} \sum_{i\in s_h} \frac{b_1 \gamma_i + a_1 R_i H_{ik} u_{1i}}{z_i}.
\end{aligned}$$

where

$$\gamma_i = H_{ik} - a_1/a_2 = H_{ik} - \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}}{\sum_h \alpha_h N_h},$$

and

$$u_{1i} = y_{1i} - \frac{b_1}{b_2} = Y_{1i} - \mu_{1kN} = Y_{1i} - \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik} Y_{1i}}{\sum_h \alpha_h \sum_{i=1}^{N_h} R_i H_{ik}}.$$

Now for the control group, we define

$$\begin{aligned}
\hat{c}_1 &= \sum_{i\in s} (1 - R_i) H_{ik} w_i Y_{0i}, \quad \hat{c}_2 = \sum_{i\in s} (1 - R_i) H_{ik} w_i, \\
c_1 &= \sum_h \alpha_h \sum_{i=1}^{N_h} (1 - R_i) H_{ik} Y_{0i}, \quad c_2 = \sum_h \alpha_h \sum_{i=1}^{N_h} (1 - R_i) H_{ik}
\end{aligned}$$

Then similarly, we have

$$\frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{c}_1}{\hat{c}_2}-\frac{a_1}{a_2}\cdot\frac{c_1}{c_2} = \frac{\sum_{i\in s} H_{ik}w_i}{\sum_{i\in s} w_i}\cdot\frac{\sum_{i\in s}(1-R_i)H_{ik}w_iY_{0i}}{\sum_{i\in s}(1-R_i)H_{ik}w_i} - \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} H_{ik}}{\sum_h \alpha_h N_h}\cdot\frac{\sum_h \alpha_h \sum_{i=1}^{N_h}(1-R_i)H_{ik}Y_{0i}}{\sum_h \alpha_h \sum_{i=1}^{N_h}(1-R_i)H_{ik}}$$

$$\doteq \frac{1}{a_2c_2}\sum_{i\in s} w_i\Big\{c_1\gamma_i + a_1(1-R_i)H_{ik}u_{0i}\Big\} = \frac{1}{a_2c_2}\sum_h \alpha_h \frac{1}{n_h}\sum_{i\in s_h}\frac{c_1\gamma_i + a_1(1-R_i)H_{ik}u_{0i}}{z_i},$$

where

$$u_{0i} = Y_{0i} - \frac{c_1}{c_2} = Y_{0i} - \mu_{0kN} = Y_{0i} - \frac{\sum_h \alpha_h \sum_{i=1}^{N_h}(1-R_i)H_{ik}Y_{0i}}{\sum_h \alpha_h \sum_{i=1}^{N_h}(1-R_i)H_{ik}}.$$

The error in $\hat{\theta}_k$ is:

$$\hat{\theta}_k - \theta_{kN} = \frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{b}_1}{\hat{b}_2} - \frac{a_1}{a_2}\cdot\frac{b_1}{b_2} - \left(\frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{c}_1}{\hat{c}_2} - \frac{a_1}{a_2}\cdot\frac{c_1}{c_2}\right).$$

Its “finite population” variance is

$$Var_p\left(\frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{b}_1}{\hat{b}_2}\right) + Var_p\left(\frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{c}_1}{\hat{c}_2}\right) - 2Cov_p\left(\frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{b}_1}{\hat{b}_2},\quad \frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{c}_1}{\hat{c}_2}\right).$$

Now, an estimator of the variance of $(\hat{a}_1/\hat{a}_2)\cdot(\hat{b}_1/\hat{b}_2)$ is:

$$\widehat{Var}(\frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{b}_1}{\hat{b}_2}) \quad = \quad \frac{1}{\hat{a}_2^2}\cdot\frac{1}{\hat{b}_2^2}\sum_h \alpha_h^2 \frac{1}{n_h(n_h-1)}\left(\sum_{i\in s_h}\left(\frac{\hat{b}_1\hat{\gamma}_i + \hat{a}_1R_iH_{ik}\hat{u}_{1i}}{z_i} - \hat{r}_1\right)^2\right),$$

where $\hat{u}_{1i} = Y_{1i} - \hat{\mu}_{1k} = Y_{1i} - \hat{b}_1/\hat{b}_2$ and

$$\hat{\gamma}_i = H_{ik} - \frac{\hat{a}_1}{\hat{a}_2} = H_{ik} - \frac{\sum_{i\in s} H_{ik}w_i}{\sum_{i\in s} w_i},$$

and

$$\hat{r}_1 = \frac{1}{n_h}\sum_{i\in s_h}\frac{\hat{b}_1\hat{\gamma}_i + \hat{a}_1R_iH_{ik}\hat{u}_{1i}}{z_i};$$

and an estimator of the variance of $(\hat{a}_1/\hat{a}_2)\cdot(\hat{c}_1/\hat{c}_2)$ is:

$$\widehat{Var}(\frac{\hat{a}_1}{\hat{a}_2}\cdot\frac{\hat{c}_1}{\hat{c}_2}) \quad = \quad \frac{1}{\hat{a}_2^2}\cdot\frac{1}{\hat{c}_2^2}\sum_h \alpha_h^2 \frac{1}{n_h(n_h-1)}\left(\sum_{i\in s_h}\left(\frac{\hat{c}_1\hat{\gamma}_i + \hat{a}_1(1-R_i)H_{ik}\hat{u}_{0i}}{z_i} - \hat{r}_0\right)^2\right),$$

where $\hat{u}_{0i} = Y_{0i} - \hat{\mu}_{0k} = Y_{0i} - \hat{c}_1/\hat{c}_2$ and

$$\hat{r}_0 = \frac{1}{n_h} \sum_{i \in s_h} \frac{\hat{c}_1 \hat{\gamma}_i + \hat{a}_1 (1 - R_i) H_{ik} \hat{u}_{0i}}{z_i};$$

and the covariance term can be estimated by:

$$\frac{1}{\hat{a}_2^2} \cdot \frac{1}{\hat{b}_2} \cdot \frac{1}{\hat{c}_2} \sum_h \alpha_h^2 \frac{1}{n_h(n_h - 1)} \left(\sum_{i \in s_h} \left(\frac{\hat{b}_1 \hat{\gamma}_i + \hat{a}_1 R_i H_{ik} \hat{u}_{1i}}{z_i} - \hat{r}_1 \right) \left(\frac{\hat{c}_1 \hat{\gamma}_i + \hat{a}_1 (1 - R_i) H_{ik} \hat{u}_{0i}}{z_i} - \hat{r}_0 \right) \right).$$

Finally, the variance estimator for $\hat{\theta}_{str}$ is:

$$\sum_{k=1}^{K} \widehat{Var} \left\{ \left(\frac{\sum_{i \in s} H_{ik} w_i}{\sum_{i \in s} w_i} \right) \hat{\theta}_k \right\}.$$

4.3 Propensity Score Weighting and Two-Sample Pseudo EL Method

There is an alternative type of estimator of $\theta = E(Y_1) - E(Y_0)$, which constructs weights for each individual using his/her propensity score (Rosenbaum (1998), Lunceford and Davidian (2004)). The rationale of the propensity score weighting estimator is that the expectation of the potential responses Y_1 and Y_0 can be identified from the observed data weighted by the inverse of the propensity score, or specifically:

$$E\left\{ \frac{RY_1}{e(X)} \right\} = E\left[E\left\{ \frac{RY_1}{e(X)} \middle| Y_1, X \right\} \right] = E\left[\frac{Y_1}{e(X)} E\{R | Y_1, X\} \right] = E(Y_1),$$

and

$$E\left\{ \frac{(1-R)Y_0}{1-e(X)} \right\} = E\left[E\left\{ \frac{(1-R)Y_0}{1-e(X)} \middle| Y_0, X \right\} \right] = E\left[\frac{Y_0}{1-e(X)} E\{(1-R) | Y_1, X\} \right] = E(Y_0).$$

The estimator based on propensity score weighting (Rosenbaum (1998)) is then:

$$\hat{\theta}_{IPW} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{R_i Y_{1i}}{\hat{e}_i} - \frac{(1-R_i) Y_{0i}}{1 - \hat{e}_i} \right\},$$

where $\hat{e}_i$ is the estimated propensity score for subject i.

Now we extend the idea under the context of survey data. If we write the above estimator as a finite population level quantity:

$$\theta_N = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{R_iY_{1i}}{\hat{e}_i} - \frac{(1-R_i)Y_{0i}}{1-\hat{e}_i}\right\},$$

then the design-unbiased HT estimator is given by

$$\frac{1}{N}\sum_{i\in s} d_i\left\{\frac{R_iY_{1i}}{\hat{e}_i} - \frac{(1-R_i)Y_{0i}}{1-\hat{e}_i}\right\}, \tag{4.4}$$

where $d_i = 1/\pi_i$, and π_i is the probability of subject i being selected into the sample. Define

$$d_{1i}^* = \frac{R_id_i}{\hat{e}_i} \quad \text{and} \quad d_{0i}^* = \frac{(1-R_i)d_i}{1-\hat{e}_i};$$

then (4.4) can be written as:

$$\frac{1}{N}\sum_{i\in s}\{d_{1i}^*R_iY_{1i} - d_{0i}^*(1-R_i)Y_{0i}\}, \tag{4.5}$$

since $R_i^2 = R_i$ and $(1-R_i)^2 = (1-R_i)$. Let

$$\begin{aligned}
\tilde{d}_{1i}^* &= \frac{d_{1i}^*}{\sum_{i\in s} d_{1i}^*} = \frac{d_i/\hat{e}_i}{\sum_{i\in s_1}(d_i/\hat{e}_i)}, \quad \text{if } i \in s_1, \\
\tilde{d}_{0i}^* &= \frac{d_{0i}^*}{\sum_{i\in s} d_{0i}^*} = \frac{d_i/(1-\hat{e}_i)}{\sum_{i\in s_0}(d_i/(1-\hat{e}_i))}, \quad \text{if } i \in s_0.
\end{aligned}$$

where s_1 and s_0 are the collections of subjects in sample s who are exposed to the treatment and the control respectively. Moreover, $\tilde{d}_{1i}^* = 0$, if $i \in s_0$ and $\tilde{d}_{0i}^* = 0$, if $i \in s_1$. Since N is usually unknown, an alternative to (4.5) is the Hájek estimator:

$$\sum_{i\in s}\left\{\left(\frac{d_{1i}^*}{\sum_{i\in s} d_{1i}^*}\right)R_iY_{1i} - \left(\frac{d_{0i}^*}{\sum_{i\in s} d_{0i}^*}\right)(1-R_i)Y_{0i}\right\}$$

$$= \sum_{i \in s_1} \frac{d_i Y_{1i}/\hat{e}_i}{\sum_{i \in s_1}(d_i/\hat{e}_i)} - \sum_{i \in s_0} \frac{d_i Y_{0i}/(1-\hat{e}_i)}{\sum_{i \in s_0}(d_i/(1-\hat{e}_i))} = \sum_{i \in s_1} \tilde{d}_{1i}^* Y_{1i} - \sum_{i \in s_0} \tilde{d}_{0i}^* Y_{0i}. \tag{4.6}$$

Intuitively, if we look at $(d_i/\hat{e}_i)^{-1} = \pi_i \hat{e}_i$ as the probability that subject i is selected into the treatment sample s_1, then $\sum_{i \in s_1}(d_i/\hat{e}_i)$ should be approximately equal to the population size N, and similarly for the control group. We write the above estimator in terms of two separate quantities of s_1 and s_0 to facilitate our discussion of two-sample pseudo empirical likelihood method later on.

Moreover, we assume that the sample is stratified by countries. Let $\alpha_h = n_h/\hat{N}_h$. Then the population level quantity can be written as:

$$\theta_N = \frac{1}{\sum_h \alpha_h N_h} \sum_h \alpha_h \sum_{i=1}^{N_h} \left\{ \frac{R_i Y_{1i}}{\hat{e}_i} - \frac{(1-R_i)Y_{0i}}{1-\hat{e}_i} \right\}$$

Our proposed estimator is given by:

$$\hat{\theta} = \frac{\sum_h \alpha_h \sum_{i \in (s_1 \bigcap s_h)} d_i Y_{1i}/\hat{e}_i}{\sum_h \alpha_h \sum_{i \in (s_1 \bigcap s_h)}(d_i/\hat{e}_i)} - \frac{\sum_h \alpha_h \sum_{i \in (s_0 \bigcap s_h)} d_i Y_{0i}/(1-\hat{e}_i)}{\sum_h \alpha_h \sum_{i \in (s_0 \bigcap s_h)}(d_i/(1-\hat{e}_i))}$$

$$= \sum_{i \in s_1} \frac{w_{1i}^* Y_{1i}}{\sum_{i \in s_1} w_{1i}^*} - \sum_{i \in s_0} \frac{w_{0i}^* Y_{0i}}{\sum_{i \in s_0} w_{0i}^*} = \sum_{i \in s_1} \tilde{w}_{1i}^* Y_{1i} - \sum_{i \in s_0} \tilde{w}_{0i}^* Y_{0i}. \tag{4.7}$$

where $w_{1i}^* = \alpha_h d_i/\hat{e}_i$, $i \in s_1$, $w_{0i}^* = \alpha_h d_i/(1-\hat{e}_i)$, $i \in s_0$, and $\tilde{w}_{mi}^* = w_{mi}^*/\sum_{i \in s_m} w_{mi}^*$, $m = 1, 0$. $\hat{\theta}$ is an approximately design-unbiased estimator of θ_N. Next we want to show that $\hat{\theta}$ is a maximum pseudo empirical likelihood estimator, and discuss its asymptotic properties.

The idea of pseudo empirical log-likelihood function was first proposed by Chen and Sitter (1999) for point estimation, where the authors constructed the log-likelihood function based on the sample data as

$$\hat{\ell}_{HT}(p) = \sum_{i \in s} d_i \log p_i,$$

the Horvitz-Thompson (HT) estimator of the finite population level log-likelihood $\ell_N = \sum_{i=1}^{N} \log p_i$. Wu and Rao (2006) extended the idea of pseudo EL for constructing hypothesis tests and confidence intervals based on a single complex survey sample. In Wu and Yan (2012), the authors considered a two-sample pseudo empirical likelihood for two independent samples selected from two separate finite populations. In this section, we want to adapt the idea of Wu and Yan (2012) to the setting of our problem, where individuals in one finite population are exposed to two treatment groups, and we want to estimate the treatment effect using sample data. Suppose the number of individuals in s_1 and s_0 are n_1 and n_0 respectively, then we consider the following pseudo empirical log-likelihood function:

$$\ell_{pel}(\boldsymbol{p}_1, \boldsymbol{p}_0) = \frac{1}{2}\sum_{i=1}^{n_1} \tilde{w}_{1i}^{*} \log(p_{1i}) + \frac{1}{2}\sum_{j=1}^{n_0} \tilde{w}_{0j}^{*} \log(p_{0j}),$$

where $\tilde{w}_{1i}^{*}$ and $\tilde{w}_{0j}^{*}$ are defined previously. As in Wu and Yan (2012), putting 1/2 in front each summation is to facilitate the reformulation of the constraints which are to be specified below. We maximize the $\ell_{pel}(\boldsymbol{p}_1, \boldsymbol{p}_0)$ subject to the following constraints:

$$\sum_{i=1}^{n_1} p_{1i} = 1, \quad \sum_{j=1}^{n_0} p_{0j} = 1; \tag{4.8}$$

$$\sum_{i=1}^{n_1} p_{1i} Y_{1i} - \sum_{j=1}^{n_0} p_{0j} Y_{0j} = \theta_N; \tag{4.9}$$

The maximum pseudo EL estimator of θ is computed as:

$$\hat{\theta}_{pel} = \sum_{i=1}^{n_1} \hat{p}_{1i} Y_{1i} - \sum_{j=1}^{n_0} \hat{p}_{0j} Y_{0j},$$

where $\hat{p}_{1i} = \tilde{w}_{1i}^{*}$ and $\hat{p}_{0j} = \tilde{w}_{0j}^{*}$ maximize $\ell_{pel}(\boldsymbol{p}_1, \boldsymbol{p}_0)$ subject to the normalization constraint (4.8). Then the resulting estimator $\hat{\theta}_{pel}$ is given by:

$$\hat{\theta}_{pel} = \sum_{i=1}^{n_1} \tilde{w}_{1i}^{*} Y_{1i} - \sum_{j=1}^{n_0} \tilde{w}_{0j}^{*} Y_{0j}, \tag{4.10}$$

which is the same as (4.7), thus approximately design-unbiased for $\hat{\theta}_N$. Since constraint (4.8) is equivalent to $\sum_{i=1}^{n_1} p_{1i} = 1$ and $\sum_{i=1}^{n_1}(1/2)p_{1i} + \sum_{j=1}^{n_0}(1/2)p_{0j} = 1$, and constraint (4.9) is equivalent to $\sum_{i=1}^{n_1}(1/2)p_{1i}(2Y_{1i}) + \sum_{j=1}^{n_0}(1/2)p_{0j}(-2Y_{0j}) = \theta_N$, we can rewrite the constraints (4.8) and (4.9) as the following:

$$\sum_{i=1}^{n_1}\left(\frac{1}{2}\right)p_{1i} + \sum_{j=1}^{n_0}\left(\frac{1}{2}\right)p_{0j} = 1, \tag{4.11}$$

$$\sum_{i=1}^{n_1}\left(\frac{1}{2}\right)p_{1i}\boldsymbol{u}_{1j} + \sum_{j=1}^{n_0}\left(\frac{1}{2}\right)p_{0j}\boldsymbol{u}_{0j} = \mathbf{0}, \tag{4.12}$$

where $\boldsymbol{u}_{1i} = (1, 2Y_{1i})^T - \boldsymbol{\eta}$, $\boldsymbol{u}_{0j} = (0, -2Y_{0j})^T - \boldsymbol{\eta}$, and $\boldsymbol{\eta} = (1/2, \theta_N)^T$. Using the Lagrange multiplier method to maximize $\ell_{pel}(\boldsymbol{p}_1, \boldsymbol{p}_0)$ subject to constraints (4.11) and (4.12) gives us the $\hat{p}_{1i}(\theta)$ and $\hat{p}_{0j}(\theta)$ as:

$$\begin{aligned}
\hat{p}_{1i}(\theta_N) &= \tilde{w}_{1i}^*/(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{1i}), \quad i = 1, \cdots, n_1 \\
\hat{p}_{0j}(\theta_N) &= \tilde{w}_{0j}^*/(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{0j}), \quad j = 1, \cdots, n_0.
\end{aligned}$$

The Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to:

$$\sum_{i=1}^{n_1}\frac{1}{2}\frac{\tilde{w}_{1i}^*\boldsymbol{u}_{1i}}{(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{1i})} + \sum_{j=1}^{n_0}\frac{1}{2}\frac{\tilde{w}_{0j}^*\boldsymbol{u}_{0j}}{(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{0j})} = \mathbf{0}, \tag{4.13}$$

which can be solved by the algorithm developed in Wu (2004). Define the pseudo empirical log-likelihood ratio statistic for θ_N as:

$$\begin{aligned}
r_{pel}(\theta_N) &= \ell_{pel}(\hat{\boldsymbol{p}}_1(\theta_N), \hat{\boldsymbol{p}}_0(\theta_N)) - \ell_{pel}(\hat{\boldsymbol{p}}_1(\hat{\theta}_{pel}), \hat{\boldsymbol{p}}_0(\hat{\theta}_{pel})) \\
&= -\left\{\frac{1}{2}\sum_{i=1}^{n_1}\tilde{w}_{1i}^*\log(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{1i}) + \frac{1}{2}\sum_{j=1}^{n_0}\tilde{w}_{0j}^*\log(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{0j})\right\}.
\end{aligned}$$

By using arguments similar to those in the proof of Theorem 2.2 in Wu and Yan (2012), we derive the asymptotic property of $r_{pel}(\theta_N)$. Substituting $1/(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{1i}) = 1 - \boldsymbol{\lambda}^T\boldsymbol{u}_{1i}/(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{1i})$ and $1/(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{0j}) = 1 - \boldsymbol{\lambda}^T\boldsymbol{u}_{0j}/(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{0j})$ into (4.13), then we have

$$\left\{\sum_{i=1}^{n_1}\frac{1}{2}\frac{\tilde{w}_{1i}^*\boldsymbol{u}_{1i}\boldsymbol{u}_{1i}^T}{(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{1i})} + \sum_{j=1}^{n_0}\frac{1}{2}\frac{\tilde{w}_{0j}^*\boldsymbol{u}_{0j}\boldsymbol{u}_{0j}^T}{(1 + \boldsymbol{\lambda}^T\boldsymbol{u}_{0j})}\right\}\boldsymbol{\lambda} = \frac{1}{2}\sum_{i=1}^{n_1}\tilde{w}_{1i}^*\boldsymbol{u}_{1i} + \frac{1}{2}\sum_{j=1}^{n_0}\tilde{w}_{0j}^*\boldsymbol{u}_{0j}.$$

We note that the right-hand side of the above equation is:

$$\boldsymbol{U} = \frac{1}{2}\sum_{i=1}^{n_1}\tilde{w}_{1i}^*\boldsymbol{u}_{1i} + \frac{1}{2}\sum_{j=1}^{n_0}\tilde{w}_{0j}^*\boldsymbol{u}_{0j} = (0, \sum_{i=1}^{n_1}\tilde{w}_{1i}^*Y_{1i} - \sum_{j=1}^{n_0}\tilde{w}_{0j}^*Y_{0j} - \theta_N)^T,$$

and it can be shown that $\boldsymbol{\lambda} \doteq \boldsymbol{D}^{-1}\boldsymbol{U}$ where $\boldsymbol{D}$ is given by

$$\boldsymbol{D} = (1/2)\sum_{i=1}^{n_1}\tilde{w}_{1i}^*\boldsymbol{u}_{1i}\boldsymbol{u}_{1i}^T + (1/2)\sum_{j=1}^{n_0}\tilde{w}_{0j}^*\boldsymbol{u}_{0j}\boldsymbol{u}_{0j}^T. \tag{4.14}$$

Then

$$\begin{aligned} -2r_{pel}(\theta_N) &= 2\bigg\{\frac{1}{2}\sum_{i=1}^{n_1}\tilde{w}_{1i}^*\log(1+\boldsymbol{\lambda}^T\boldsymbol{u}_{1i}) + \frac{1}{2}\sum_{j=1}^{n_0}\tilde{w}_{0j}^*\log(1+\boldsymbol{\lambda}^T\boldsymbol{u}_{0j})\bigg\} \\ &\doteq 2\bigg\{\frac{1}{2}\sum_{i=1}^{n_1}\tilde{w}_{1i}^*\Big(\boldsymbol{\lambda}^T\boldsymbol{u}_{1i} - \frac{1}{2}\boldsymbol{\lambda}^T\boldsymbol{u}_{1i}\boldsymbol{u}_{1i}^T\boldsymbol{\lambda}\Big) + \frac{1}{2}\sum_{j=1}^{n_0}\tilde{w}_{0j}^*\Big(\boldsymbol{\lambda}^T\boldsymbol{u}_{0j} - \frac{1}{2}\boldsymbol{\lambda}^T\boldsymbol{u}_{0j}\boldsymbol{u}_{0j}^T\boldsymbol{\lambda}\Big)\bigg\} \\ &= \boldsymbol{U}^T\boldsymbol{D}^{-1}\boldsymbol{U} = d^{(22)}\Big(\sum_{i=1}^{n_1}\tilde{w}_{1i}^*Y_{1i} - \sum_{j=1}^{n_0}\tilde{w}_{0j}^*Y_{0j} - \theta_N\Big). \end{aligned}$$

Therefore, $-2r_{pel}(\theta_N)/c$ converges in distribution to a χ_1^2 random variable as $n \to \infty$, where c is given by:

$$c = d^{(22)}\bigg\{V_p\Big(\sum_{i=1}^{n_1}\tilde{w}_{1i}^*Y_{1i} - \sum_{j=1}^{n_0}\tilde{w}_{0j}^*Y_{0j}\Big)\bigg\}, \tag{4.15}$$

$d^{(22)}$ is the second diagonal element of the matrix $\boldsymbol{D}^{-1}$, and V_p stands for the design-based variance. Standard regularity conditions of deriving the asymptotic distribution of the pseudo EL ratio statistic are presented in Wu and Rao (2006). Additional conditions might also be needed with the estimation of the propensity score and the propensity score weighting. Since $\boldsymbol{D}$ involves the unknown parameter θ_N, in practice, we can plug in $\theta_N = \hat{\theta}_{pel} = \sum_{i=1}^{n_1}\tilde{w}_{1i}^*Y_{1i} - \sum_{j=1}^{n_0}\tilde{w}_{0j}^*Y_{0j}$ into $\boldsymbol{D}$ to obtain an estimated $\hat{c} = \hat{d}^{(22)}\hat{V}_p(\hat{\theta}_{pel})$. A designed based variance estimator $\hat{V}_p(\hat{\theta}_{pel})$ will be given below. We can then construct a $(1-\alpha)100\%$ pseudo EL confidence interval by

$$\{\theta | -2r_{pel}(\theta)/\hat{c} \le \chi_1^2(\alpha)\}.$$

Now we derive the designed based variance estimator $\hat{V}_p(\hat{\theta}_{pel})$. We write the $\hat{\theta}_{pel}$ in the following way:

$$
\begin{aligned}
\hat{\theta}_{pel} &= \sum_{i=1}^{n_1} \tilde{w}_{1i}^* Y_{1i} - \sum_{j=1}^{n_0} \tilde{w}_{0j}^* Y_{0j} = \sum_{i \in s_1} \frac{w_{1i}^* Y_{1i}}{\sum_{i \in s_1} w_{1i}^*} - \sum_{i \in s_0} \frac{w_{0i}^* Y_{0i}}{\sum_{i \in s_0} w_{0i}^*} \\
&= \frac{\sum_h \alpha_h \sum_{i \in (s_1 \bigcap s_h)} d_i Y_{1i}/\hat{e}_i}{\sum_h \alpha_h \sum_{i \in (s_1 \bigcap s_h)} (d_i/\hat{e}_i)} - \frac{\sum_h \alpha_h \sum_{i \in (s_0 \bigcap s_h)} d_i Y_{0i}/(1-\hat{e}_i)}{\sum_h \alpha_h \sum_{i \in (s_0 \bigcap s_h)} (d_i/(1-\hat{e}_i))}
\end{aligned}
$$

If we write both the numerators and the denominators in a Hansen-Hurwitz form, then

$$
\begin{aligned}
\hat{\theta}_{pel} &= \frac{\sum_h \alpha_h (1/n_h) \sum_{i \in (s_1 \bigcap s_h)} (Y_{1i}/z_{1i}^*)}{\sum_h \alpha_h (1/n_h) \sum_{i \in (s_1 \bigcap s_h)} (1/z_{1i}^*)} - \frac{\sum_h \alpha_h (1/n_h) \sum_{i \in (s_0 \bigcap s_h)} (Y_{0i}/z_{0i}^*)}{\sum_h \alpha_h (1/n_h) \sum_{i \in (s_0 \bigcap s_h)} (1/z_{0i}^*)} \\
&=: \hat{\mu}_{1pel} - \hat{\mu}_{0pel},
\end{aligned}
$$

where $z_{1i}^* = \hat{e}_i/(d_i n_h)$, $z_{0i}^* = (1-\hat{e}_i)/(d_i n_h)$, and n_h is the number of individuals in the sample s who come from country h. The θ_N can be written as:

$$
\begin{aligned}
\theta_N &= \frac{1}{\sum_h \alpha_h N_h} \sum_h \alpha_h \sum_{i=1}^{N_h} \left\{ \frac{R_i Y_{1i}}{\hat{e}_i} - \frac{(1-R_i) Y_{0i}}{1-\hat{e}_i} \right\} \\
&= \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} (R_i Y_{1i}/\hat{e}_i)}{\sum_h \alpha_h N_h} - \frac{\sum_h \alpha_h \sum_{i=1}^{N_h} ((1-R_i) Y_{1i}/(1-\hat{e}_i))}{\sum_h \alpha_h N_h} \\
&=: \mu_{1N} - \mu_{0N}.
\end{aligned}
$$

Then

$$
\begin{aligned}
&\hat{\mu}_{1pel} - \mu_{1N} \\
&= \frac{1}{\sum_h \alpha_h \frac{1}{n_h} \sum_{i \in (s_1 \bigcap s_h)} \frac{1}{z_{1i}^*}} \left\{ \sum_h \alpha_h \frac{1}{n_h} \sum_{i \in (s_1 \bigcap s_h)} \left(\frac{Y_{1i}}{z_{1i}^*} \right) - \mu_{1N} \sum_h \alpha_h \frac{1}{n_h} \sum_{i \in (s_1 \bigcap s_h)} \left(\frac{1}{z_{1i}^*} \right) \right\} \\
&= \frac{1}{\sum_h \alpha_h \frac{1}{n_h} \sum_{i \in (s_1 \bigcap s_h)} \frac{1}{z_{1i}^*}} \left\{ \sum_h \alpha_h \frac{1}{n_h} \sum_{i \in (s_1 \bigcap s_h)} \left(\frac{r_{1i}}{z_{1i}^*} \right) \right\},
\end{aligned}
$$

where $r_{1i} = Y_{1i} - \mu_{1N}$. Similarly,

$$
\begin{aligned}
&\hat{\mu}_{0pel} - \mu_{0N} \\
= &\ \frac{1}{\sum_h \alpha_h \frac{1}{n_h} \sum_{i\in(s_0 \bigcap s_h)} \frac{1}{z^*_{0i}}} \left\{ \sum_h \alpha_h \frac{1}{n_h} \sum_{i\in(s_0 \bigcap s_h)} \left(\frac{Y_{0i}}{z^*_{0i}} \right) - \mu_{0N} \sum_h \alpha_h \frac{1}{n_h} \sum_{i\in(s_0 \bigcap s_h)} \left(\frac{1}{z^*_{0i}} \right) \right\} \\
= &\ \frac{1}{\sum_h \alpha_h \frac{1}{n_h} \sum_{i\in(s_0 \bigcap s_h)} \frac{1}{z^*_{0i}}} \left\{ \sum_h \alpha_h \frac{1}{n_h} \sum_{i\in(s_0 \bigcap s_h)} \left(\frac{r_{0i}}{z^*_{0i}} \right) \right\},
\end{aligned}
$$

where $r_{0i} = Y_{0i} - \mu_{0N}$. The error in $\hat{\theta}_{pel}$ is

$$\hat{\theta}_{pel} - \theta_N = (\hat{\mu}_{1pel} - \mu_{1N}) - (\hat{\mu}_{0pel} - \mu_{0N}),$$

and its "finite population" variance is

$$V_p(\hat{\theta}_{pel}) = V_p(\hat{\mu}_{1pel}) + V_p(\hat{\mu}_{0pel}) - 2Cov_p(\hat{\mu}_{1pel}, \hat{\mu}_{0pel}).$$

An estimator of $V_p(\hat{\mu}_{1pel})$ is

$$\hat{V}_p(\hat{\mu}_{1pel}) = \frac{1}{(\sum_{i\in s_1} w^*_{1i})^2} \sum_h \alpha_h^2 \frac{1}{n_h(n_h-1)} \left\{ \sum_{i\in s_h} \left(\frac{R_i \hat{r}_{1i}}{z^*_{1i}} - \hat{t}_{r1} \right)^2 \right\},$$

where $\hat{r}_{1i} = Y_{1i} - \hat{\mu}_{1pel}$ and $\hat{t}_{r1} = (1/n_h) \sum_{i\in(s_1 \bigcap s_h)} (\hat{r}_{1i}/z^*_{1i})$. An estimator of $V_p(\hat{\mu}_{0pel})$ is

$$\hat{V}_p(\hat{\mu}_{0pel}) = \frac{1}{(\sum_{i\in s_0} w^*_{0i})^2} \sum_h \alpha_h^2 \frac{1}{n_h(n_h-1)} \left\{ \sum_{i\in s_h} \left(\frac{(1-R_i) \hat{r}_{0i}}{z^*_{0i}} - \hat{t}_{r0} \right)^2 \right\},$$

where $\hat{r}_{0i} = Y_{0i} - \hat{\mu}_{0pel}$ and $\hat{t}_{r0} = (1/n_h) \sum_{i\in(s_0 \bigcap s_h)} (\hat{r}_{0i}/z^*_{0i})$. And an estimator for $Cov_p(\hat{\mu}_{1pel}, \hat{\mu}_{0pel})$ is

$$
\begin{aligned}
&\widehat{Cov}_p(\hat{\mu}_{1pel}, \hat{\mu}_{0pel}) \\
= &\ \frac{1}{(\sum_{i\in s_1} w^*_{1i})(\sum_{i\in s_0} w^*_{0i})} \sum_h \alpha_h^2 \frac{1}{n_h(n_h-1)} \left\{ \sum_{i\in s_h} \left(\frac{R_i \hat{r}_{1i}}{z^*_{1i}} - \hat{t}_{r1} \right) \left(\frac{(1-R_i)\hat{r}_{0i}}{z^*_{0i}} - \hat{t}_{r0} \right) \right\}.
\end{aligned}
$$

Therefore,

$$\hat{V}_p(\hat{\theta}_{pel}) = \hat{V}_p(\hat{\mu}_{1pel}) + \hat{V}_p(\hat{\mu}_{0pel}) - 2\widehat{Cov}_p(\hat{\mu}_{1pel}, \hat{\mu}_{0pel}).$$

4.4 Application to the ITC 4 Country Survey Data

The data we use in this application is the Wave 7-8 M7/P7-M8 continuers (subjects who participate in both Wave 7 and Wave 8) of the ITC 4 Country survey. The Wave 7 survey was conducted from October 2008 to July 2009, and the Wave 8 survey was conducted from July 2010 to June 2011, for all 4 countries: Australia, Canada, UK, and USA. The sample sizes of Wave 7-8 Recontact is 5135, which consists of 1292 individuals from Australia, 1374 individuals from Canada, 1325 individuals from the UK, and 1144 individuals from the US. We treat the Wave 7 data as the baseline information, the Wave 8 data as the posttest responses, and "answer by web" as the treatment exposure. We want to estimate the treatment effect using the methods we developed previously in this chapter.

4.4.1 Variables and Data Management

Most questions in the questionnaire are standardized throughout Wave 1-Wave 9, and the variable names of different waves differ by the first letter. The variable names of Wave 7 begin with the letter "g", and the variable names of Wave 8 begin with the letter "h". The response variable we use in the analysis is hFR245v, "cigarettes per day (CPD)" at Wave 8, which is a derived continuous variable. The same variable at Wave 7 is gFR245v, and we regard it as the "pretest response" of CPD. The following is a list of the baseline covariates we included in the data analysis:

- age: a continuous variable
- COUNTRY: a categorical variable
- ethnic: "ethnicity", a categorical variable
- gCH801: "visited doctor since last survey day", a binary variable
- gDE111: "marital status", a categorical variable

- gDE212v: "income categories", a categorical variable
- gDE312v: "education categories", a categorical variable
- gDI503: "depression: little interest or pleasure", a binary variable
- gDI504: "depression: feeling down or hopeless", a binary variable
- gDI505: "depression: diagnosed with depression", a binary variable (answered if gDI503 and gDI504 are yes)
- gDI701: "frequency of alcohol drinks consumed in last 12 months", a categorical variable
- gFR309v: "smoking status", a derived categorical variable
- gPR101: "describe your health", a categorical variable
- sex: a binary variable

The treatment variable is hMode, a categorical variable which is coded as: 1 -"telephone English", 2 -"telephone French", 3 -"internet English", and 4 -"internet French". We derived a binary treatment indicator variable "trt" as: 1 -"telephone" and 2 -"internet" based on hMode. The question corresponding to gDI505 is only answered by the participants who answered "yes" to both gDI503 and gDI504; thus we derive a binary variable "gDI505r" as: 1 -"yes", and 2 -"no" or not answered. Also, we dichotomized gDI701 to "gDI701r" as: 1 - "at least once per week" and 2 - "less than once per week". And we dichotomized gFR309v to "gsmkstat" as: 1 - "smoker" and 2 - "non-smoker". Moreover, we derived a binary variable "email", to indicate whether the respondent had been email invited or not, based on the responses to the question gAI512: "what would be best email address to contact you on". We coded "email" 1 - "yes", if the answer to gAI512 is "respondent willing and offers email address", and 2 - "no", if answered otherwise to gAI512. The weights variable we use is hDE963v, the longitudinal weight for M7/P7-M8 continuers rescaled to sum to country sample size. The rate of missing data is on average about 5% for each variable we included in the analysis. We omitted the entire observation of an individual if his/her answer to one of our included variables is missing.

4.4.2 Propensity Score Model and Balance Diagnosis

The propensity score model we assumed in the analysis is:

$$\log\left\{\frac{P(\mathrm{trt}_i = 2)}{1 - P(\mathrm{trt}_i = 2)}\right\} = \boldsymbol{X}_i^T\boldsymbol{\beta}, \quad i \in s,$$

where the covariate matrix $\boldsymbol{X}$ = (1, age, COUNTRY, ethnic, gCH801, gDE111, gDE212v, gDE312v, gDI503, gDI504, gDI505r, gDI701r, gsmkstat, gPR101, sex, gFR245v), and $\boldsymbol{\beta}$ is the coefficient. The estimated propensity score for each individual i is:

$$\hat{e}_i = \{1 + \exp(-\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}})\}^{-1}, \quad i \in s,$$

where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$. We fit the propensity score model separately for the group of individuals with email invitation and for the group of individuals without email invitation. This was also suggested in a previous analysis using this data by Hajducek et al. (2012), because these two groups were inherently different. Also, the performance of the balance diagnosis was not good when we fit a propensity score model for the whole dataset; yet it was much more improved when the propensity score model was fitted separately for the email invitation group and the no email invitation group. From the following table, we can see that most of the people who received the email invitation chose the web survey, whereas the majority of the no email invitation group chose to answer by phone.

	Telephone	Internet
Email invited	800	1321
No Email invited	1909	477

For the propensity score stratification method, we stratify the sample into $K = 5$ strata based on the sample quintile of the estimated propensity score. We also perform this stratification separately for the email invited and no email invited groups. We perform balance diagnosis for each of the 15 covariates included in the propensity score model. We treat each covariate as the response and fit a regression model to examine the "treatment effect". For example Table 4.1 shows the summary of fitting the logistic regression of gDI503 (a binary variable) against trt only. It indicates that, prior to stratification, trt has a significant "effect" on gDI503, i.e. gDI503 is not balanced between treatment groups. Table 4.2 is the summary of fitting a logistic regression of gDI503 against trt, quint, and the interaction of trt and quint (trt×quint). It indicates that after stratification, trt is no longer statistically significant; thus, the distribution of gDI503 is balanced between treatment groups. Table 4.3 and Table 4.4 summarize the results of fitting the linear regression model of gFR245v (a continuous variable) with the main treatment effect only and with the treatment, quint and trt×quint interaction. After stratification the mean of gFR245v is balanced between treatment groups. We further look at the box-plots of gFR245v against treatment groups for each stratum (quint) in Figure 4.1. From the box-plots we can see the quantiles of gFR245v between two treatment groups are close for most strata. In the stratum 4, we notice a slight difference of the upper quartiles and the maximums of gFR245v between two treatment groups. We could consider refining the propensity score model following the way discussed in Rosenbaum and Rubin (1984). However, in this analysis, we assume our initial propensity score model is adequate.

4.4.3 Data Analysis

Our objective is to estimate the effect of "answering by internet" (treatment) on the response variable "hFR245v" (CPD at Wave 8). We let $\hat{\theta}_{str}$ represent the the propensity score stratification estimator and $\hat{\theta}_{pel}$ represent the maximum pseudo EL estimator based on propensity score weighting. We only consider the design-based variance estimators in the data analysis. For $\hat{\theta}_{str}$, the confidence intervals are constructed using the normal approximation. For $\hat{\theta}_{pel}$, we report two types of confidence intervals. One is the normal based confidence interval, and the other is constructed based on the pseudo empirical log-likelihood ratio (PELr) statistic:

$$\{\theta | -2r_{pel}(\theta)/\hat{c} \leq \chi_1^2(0.05)\}.$$

We first estimate the treatment effect separately for the email invitation group and the no email invitation group. Table 4.5 summarizes the results. For the email invitation group, $\hat{\theta}_{str}$ is equal to -1.514283 with variance estimate 0.288476 and the 95% confidence interval $(-2.566999, -0.461568)$. This result suggests a significant difference of hFR245v between the two different answering modes (treatment groups). In the meanwhile, $\hat{\theta}_{pel}$ is equal to -1.474795 with variance estimate 0.3285852. We see that these results are fairly close to those of $\hat{\theta}_{str}$ for the email invitation group. The 95% CI of $\hat{\theta}_{pel}$ based on the normal approximation and the PELr are very similar and equal to $(-2.598313, -0.3512768)$ and $(-2.592939, -0.3457211)$ respectively. We obtain the same conclusion that the difference of hFR245v is significant between the two survey modes based on the results of $\hat{\theta}_{pel}$.

For the no email invitation group, we observe a discrepancy between $\hat{\theta}_{str}$ (-0.8966309) and $\hat{\theta}_{pel}$ (-0.3144521). One possible reason is that for the propensity score stratification method, some residual confounding within strata may remain. Lunceford and Davidian (2004) discussed a modified propensity score stratification method which can reduce residual within-stratum confounding. The variance estimates of $\hat{\theta}_{str}$ and $\hat{\theta}_{pel}$ are close and respectively equal to 0.4099932 and 0.4631958. Based on the 95% CI's of both $\hat{\theta}_{str}$ and $\hat{\theta}_{pel}$, we conclude that there is no significant difference of hFR245 between treatment groups in the no email invitation group.

In addition, we combine the email invitation group and the no email invitation group together, and apply the two-sample pseudo EL method based on propensity score weighting for estimating the treatment effect. The results are summarized in Table 4.6. The point estimate of the treatment effect is equal to -0.8812598 with design-based variance estimate 0.2069272. The 95% confidence interval based on the pseudo empirical loglikelihood ratio

Table 4.1: Email invited group, gDI503, before propensity score stratification

Coefficients	Estimate	Std. Error	z value	Pr(> $\|z\|$)	
(Intercept)	-1.24395	0.08483	-14.663	< 2e-16	***
factor(trt)2	-0.25597	0.11078	-2.311	0.0209	*

Table 4.2: Email invited group, gDI503, after propensity score stratification

Coefficients	Estimate	Std. Error	z value	Pr(> $\|z\|$)	
(Intercept)	-0.35555	0.13824	-2.572	0.0101	*
factor(trtf)2	-0.03391	0.19768	-0.172	0.8638	
factor(quint)2	-0.96801	0.23094	-4.192	2.77e-05	***
factor(quint)3	-1.07600	0.25494	-4.221	2.44e-05	***
factor(quint)4	-1.79621	0.30581	-5.874	4.26e-09	***
factor(quint)5	-1.97394	0.34470	-5.727	1.02e-08	***
factor(trtf)2:factor(quint)2	0.05476	0.31192	0.176	0.8606	
factor(trtf)2:factor(quint)3	-0.34282	0.33773	-1.015	0.3101	
factor(trtf)2:factor(quint)4	0.30353	0.38031	0.798	0.4248	
factor(trtf)2:factor(quint)5	0.04612	0.42362	0.109	0.9133	

is $(-1.7709,\ 0.013081)$, and again it is very close to the normal based 95% CI which is $(-1.7728, 0.010329)$. This result tells that the difference of hFR245v between treatment groups in the combined sample is almost but not quite statistically significant at the 5% level.

Table 4.3: Email invited group, gFR245v, before propensity score stratification

Coefficients	Estimate	Std. Error	t value	Pr(> $\|z\|$)	
(Intercept)	13.4705	0.3913	34.426	< 2e-16	***
factor(trtf)2	-1.2996	0.4958	-2.621	0.00883	**

Table 4.4: Email invited group, gFR245v, after propensity score stratification

Coefficients	Estimate	Std. Error	t value	$Pr(>\|z\|)$	
(Intercept)	19.09583	0.71685	26.638	< 2e-16	***
factor(trtf)2	-0.92324	1.02348	-0.902	0.367	
factor(quint)2	-5.11402	1.06983	-4.780	1.87e-06	***
factor(quint)3	-7.24155	1.14311	-6.335	2.89e-10	***
factor(quint)4	-9.46250	1.13344	-8.348	< 2e-16	***
factor(quint)5	-9.86922	1.18702	-8.314	< 2e-16	***
factor(trtf)2:factor(quint)2	0.96077	1.45799	0.659	0.510	
factor(trtf)2:factor(quint)3	0.99571	1.49372	0.667	0.505	
factor(trtf)2:factor(quint)4	1.74169	1.48820	1.170	0.242	
factor(trtf)2:factor(quint)5	0.02686	1.52033	0.018	0.986	

Table 4.5: Data analyses separately for the email invitation group and the no email invitation group

	Method	$\hat{\theta}$	$\widehat{Var}(\hat{\theta})$	95% CI
Email invited	$\hat{\theta}_{str}$ (Normal based CI)	-1.514283	0.288476	(-2.566999, -0.461568)
	$\hat{\theta}_{pel}$ (Normal based CI)	-1.474795	0.3285852	(-2.598313, -0.3512768)
	$\hat{\theta}_{pel}$ (PELr CI)	-1.474795	0.3285852	(-2.592939, -0.3457211)
No email invited	$\hat{\theta}_{str}$ (Normal based CI)	-0.8966309	0.4099932	(-2.149933, 0.356671)
	$\hat{\theta}_{pel}$ (Normal based CI)	-0.3144521	0.4631958	(-1.648398, 1.019494)
	$\hat{\theta}_{pel}$ (PELr CI)	-0.3144521	0.4631958	(-1.647891, 1.024455)

Table 4.6: Pseudo two-sample EL method (the email invitation group and the no email invitation group combined)

	$\hat{\theta}_{pel}$	$\widehat{Var}(\hat{\theta}_{pel})$	95% CI
Normal based CI	-0.8812598	0.2069272	(-1.772849, 0.01032946)
PELr CI	-0.8812598	0.2069272	(-1.770908, 0.01308058)

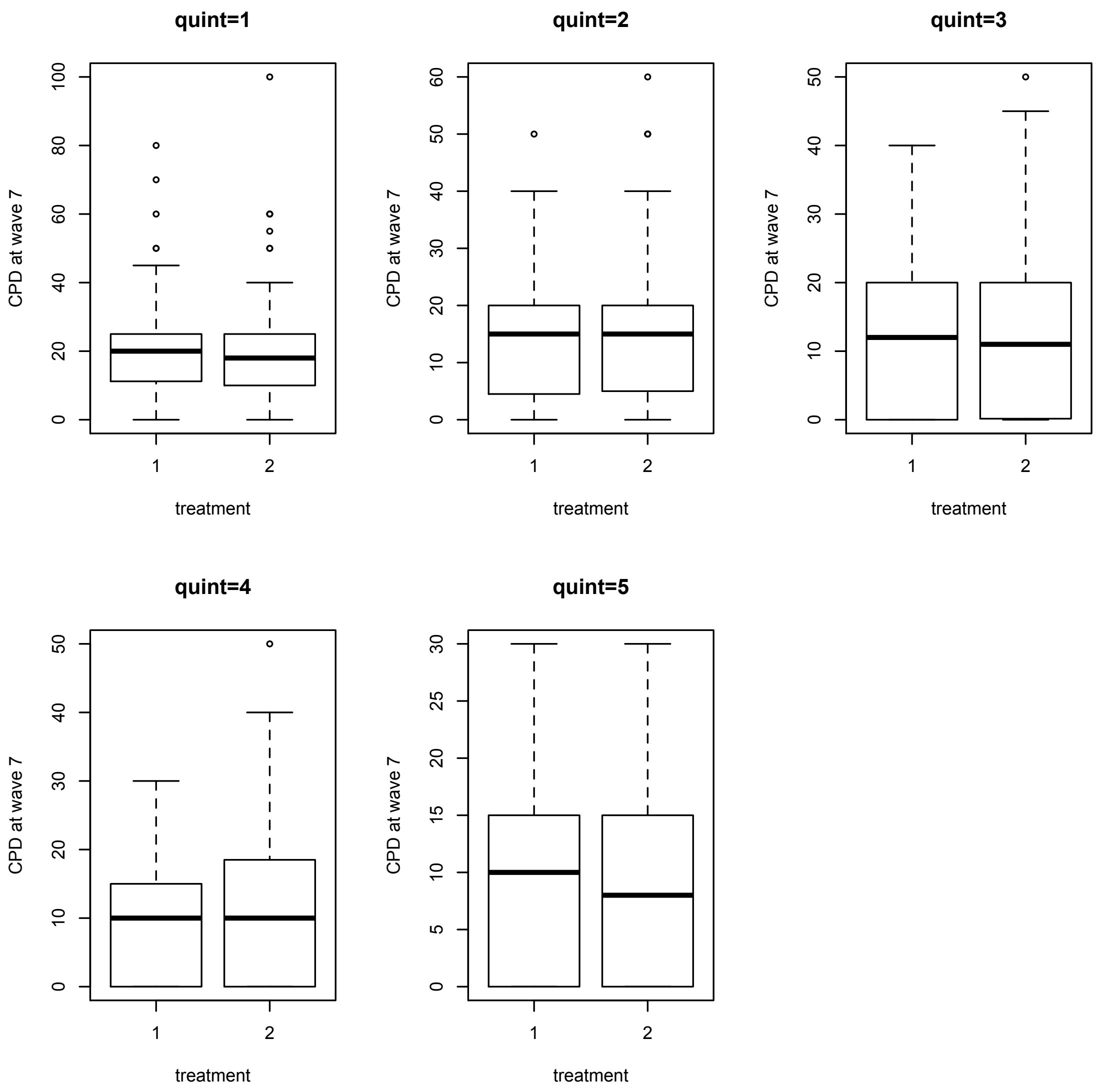

Figure 4.1: Email invited group: gFR245v

4.5 Two-Sample Pseudo EL Method with Imputation for Survey Data

In this section, we consider a randomized study with survey data. Suppose that there is one finite population, and a study sample is chosen from the finite population with a probability sampling design. The baseline covariates of each individual in the study sample are recorded as the pretest responses. Then the study sample is randomized into two groups: treatment and control. After a period of time, the posttest responses are measured. Assume the counterfactual responses are missing by design. Our goal is to develop an imputation based pseudo two-sample EL approach and discuss the asymptotic properties for the pseudo EL ratio statistic of our proposed estimator.

4.5.1 Notations

Suppose there is a finite population $\{1, \cdots, N\}$. A study sample s of size n is drawn from the finite population using a certain sampling design p. Let $d_i = 1/\pi_i, i \in s$ be the basic design weights and $\tilde{d}_i = d_i / \sum_{j \in s} d_j$ be the normalized weights, where $\pi_i = \Pr(i \in s)$ is the inclusion probability. Let $\boldsymbol{Z}_i, i = 1, \cdots, N$ be the vector of baseline covariates, and $(Y_{1i}, Y_{0i}), i = 1, \cdots, N$ be the potential posttest responses attached to each individual in the finite population. We assume the following model ξ which links the potential posttest responses and the pretest responses in the following way:

$$
\begin{aligned}
Y_{1i} &= \boldsymbol{Z}_i^T \boldsymbol{\beta}_1 + \epsilon_{1i}, \quad \epsilon_{1i} \sim (0, \sigma_{\epsilon 1}^2), \quad i = 1, \cdots, N; \\
Y_{0i} &= \boldsymbol{Z}_i^T \boldsymbol{\beta}_0 + \epsilon_{0i}, \quad \epsilon_{0i} \sim (0, \sigma_{\epsilon 0}^2), \quad i = 1, \cdots, N.
\end{aligned}
$$

The parameter of interest is $\theta = \mu_1 - \mu_0 = E_\xi(Y_1) - E_\xi(Y_0)$, where $E_\xi(\cdot)$ the the expectation taken with respect to the model. After randomization, the sample s is divided into s_1, the treatment group, and s_0, the control group. Let R_i be the treatment indicator with $R_i = 1$, for $i \in s_1$ and $R_i = 0$, for $i \in s_0$. We assume $P(R_i = 1) = \delta$, $i \in s$. Suppose there are n_1

subjects in s_1 and $n_0 = n - n_1$ subjects in s_0. Our observed data is $\{R_iY_{1i}, d_i, \boldsymbol{Z}_i, i \in s_1\}$ and $\{(1-R_i)Y_{0i}, d_i, \boldsymbol{Z}_i, i \in s_0\}$.

As in Chapter 2, we consider the linear regression imputation approach. Let $\hat{\boldsymbol{\beta}}_{w1}$ and $\hat{\boldsymbol{\beta}}_{w0}$ be the weighted least squares estimators for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$; then,

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{w1} &= \{\sum_{i\in s} d_iR_i\boldsymbol{Z}_i\boldsymbol{Z}_i^T\}^{-1}\sum_{i\in s} d_iR_i\boldsymbol{Z}_iY_{1i},\\ \hat{\boldsymbol{\beta}}_{w0} &= \{\sum_{i\in s} d_i(1-R_i)\boldsymbol{Z}_i\boldsymbol{Z}_i^T\}^{-1}\sum_{i\in s} d_i(1-R_i)\boldsymbol{Z}_iY_{0i}.\end{aligned}$$

Let $Y_{1i}^* = \boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}}_{w1}$, $i \in s_0$, and $Y_{0i}^* = \boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}}_{w0}$, for $i \in s_1$, be respectively the imputed values of Y_1 for the subjects in the control group and the imputed values of Y_0 for the subjects in the treatment group. Define

$$\begin{aligned}\tilde{Y}_{1i} &= R_iY_{1i} + (1-R_i)Y_{1i}^*, \quad i \in s\,,\\ \tilde{Y}_{0i} &= (1-R_i)Y_{0i} + R_iY_{0i}^*, \quad i \in s\,.\end{aligned}$$

The data becomes $\{d_i, \tilde{Y}_{1i}, \tilde{Y}_{0i}, i \in s\}$. A consistent estimator of $\theta = \mu_1 - \mu_0$ is given by

$$\hat{\theta} = \hat{\mu}_1 - \hat{\mu}_0 = \frac{\sum_{i\in s} d_i\tilde{Y}_{1i}}{\sum_{i\in s} d_i} - \frac{\sum_{i\in s} d_i\tilde{Y}_{0i}}{\sum_{i\in s} d_i} = \sum_{i\in s}\tilde{d}_i\tilde{Y}_{1i} - \sum_{i\in s}\tilde{d}_i\tilde{Y}_{0i}\,. \tag{4.16}$$

4.5.2 Two-Sample Pseudo EL Method

Consider the two-sample pseudo empirical log-likelihood function:

$$\ell_{pel}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^n \tilde{d}_i \log p_i + \sum_{i=1}^n \tilde{d}_i \log q_i, \tag{4.17}$$

where $\boldsymbol{p} = (p_1, \cdots, p_n)$ and $\boldsymbol{q} = (q_1, \cdots, q_n)$ are probability masses that the distributions of $\tilde{Y}_{1i}$ and $\tilde{Y}_{0i}$ respectively put onto the individuals of sample s. We maximize $\ell_{pel}(\boldsymbol{p}, \boldsymbol{q})$ subject to constraints:

$$\sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n q_i = 1, \tag{4.18}$$

and

$$\sum_{i=1}^{n} p_i \tilde{Y}_{1i} - \sum_{i=1}^{n} q_i \tilde{Y}_{0i} = \theta. \tag{4.19}$$

The maximum PEL estimator of θ is computed as $\hat{\theta}_{pel} = \sum_{i=1}^{n} \hat{p}_i \tilde{Y}_{1i} - \sum_{j=1}^{n} \hat{q}_j \tilde{Y}_{0i}$, where the $\hat{p}_i$ and $\hat{q}_j$ maximize $\ell_{pel}(\boldsymbol{p}, \boldsymbol{q})$ subject to constraint (4.18). The resulting estimator is given by

$$\hat{\theta}_{pel} = \hat{\mu}_{1pel} - \hat{\mu}_{0pel} = \sum_{i=1}^{n} \tilde{d}_i \tilde{Y}_{1i} - \sum_{j=1}^{n} \tilde{d}_j \tilde{Y}_{0j},$$

which is the same as $\hat{\theta}$ in (4.16). Let $\hat{\boldsymbol{p}}(\theta) = (\hat{p}_1(\theta), \cdots, \hat{p}_n(\theta))$ and $\hat{\boldsymbol{q}}(\theta) = (\hat{q}_1(\theta), \cdots, \hat{q}_n(\theta))$ be the maximizer of $\ell_{pel}(\boldsymbol{p}, \boldsymbol{q})$ subject to constraints (4.18) and (4.19), for fixed θ. The pseudo empirical log-likelihood ratio statistic for θ is given by:

$$r_{pel}(\theta) = \ell_{pel}\{\hat{\boldsymbol{p}}(\theta), \hat{\boldsymbol{q}}(\theta)\} - \ell_{pel}\{\hat{\boldsymbol{p}}(\hat{\theta}_{pel}), \hat{\boldsymbol{q}}(\hat{\theta}_{pel})\},$$

and note that $\hat{p}_i(\hat{\theta}_{pel}) = \hat{q}_i(\hat{\theta}_{pel}) = \tilde{d}_i$. Next we derive the asymptotic distribution of $r_{pel}(\theta)$ using similar arguments in the proof of Theorem 1 in Chapter 2. First we define a nuisance parameter μ to be $\mu = \mu_0 + o(n^{-1/2})$, and separate the constraint (4.19) into

$$\sum_{i=1}^{n} p_i \tilde{Y}_{1i} = \mu + \theta \quad \text{and} \quad \sum_{i=1}^{n} q_i \tilde{Y}_{0i} = \mu.$$

For fixed values of μ and θ, the solutions to the constrained maximization problem are given by

$$\hat{p}_i(\theta) = \frac{\tilde{d}_i}{1 + \lambda_1(\tilde{Y}_{1i} - \mu - \theta)} \quad \text{and} \quad \hat{q}_i(\theta) = \frac{\tilde{d}_i}{1 + \lambda_0(\tilde{Y}_{0i} - \mu)}.$$

The Lagrange multipliers λ_1 and λ_0 are determined by:

$$\sum_{i=1}^{n} \frac{\tilde{d}_i(\tilde{Y}_{1i} - \mu - \theta)}{1 + \lambda_1(\tilde{Y}_{1i} - \mu - \theta)} = 0 \quad \text{and} \quad \sum_{i=1}^{n} \frac{\tilde{d}_i(\tilde{Y}_{0i} - \mu)}{1 + \lambda_0(\tilde{Y}_{0i} - \mu)} = 0.$$

Let $r_{pel}(\theta, \mu)$ be the pseudo empirical log-likelihood ratio statistic on (θ, μ). We have

$$r_{pel}(\theta, \mu) \quad = \quad -\sum_{i=1}^{n} \tilde{d}_i \log\left[1 + \lambda_1(\tilde{Y}_{1i} - \mu - \theta)\right] - \sum_{i=1}^{n} \tilde{d}_i \log\left[1 + \lambda_0(\tilde{Y}_{0i} - \mu)\right].$$

Let $\hat{\mu} = \hat{\mu}(\theta)$ be the maximizer of $r_{pel}(\theta, \mu)$ for a given θ, which can be obtained through profiling. The solution is obtained by setting

$$\frac{\partial r_{pel}(\theta,\mu)}{\partial \mu} = \sum_{i=1}^{n} \frac{\tilde{d}_i \lambda_1}{1+\lambda_1(\tilde{Y}_{1i}-\mu-\theta)]} + \sum_{i=1}^{n} \frac{\tilde{d}_i \lambda_0}{1+\lambda_0(\tilde{Y}_{0i}-\mu)]} = 0,$$

which yields $\lambda_1 = -\lambda_0$. Moreover, it can be shown that

$$\lambda_1 \doteq \tilde{V}_1^{-1} \sum_{i=1}^{n} \tilde{d}_i (\tilde{Y}_{1i} - \mu - \theta), \tag{4.20}$$

$$\lambda_0 \doteq \tilde{V}_0^{-1} \sum_{i=1}^{n} \tilde{d}_i (\tilde{Y}_{0i} - \mu), \tag{4.21}$$

where $\tilde{V}_1 = \sum_{i=1}^{n} \tilde{d}_i (\tilde{Y}_{1i} - \mu_1)^2$, and $\tilde{V}_0 = \sum_{i=1}^{n} \tilde{d}_i (\tilde{Y}_{0i} - \mu_0)^2$. The profile solution $\hat{\mu} = \hat{\mu}(\theta)$, which satisfies $\lambda_1 = -\lambda_0$, is then given by:

$$\hat{\mu} \doteq \nu(\bar{\tilde{Y}}_1 - \theta) + (1-\nu)\bar{\tilde{Y}}_0\,, \tag{4.22}$$

where $\nu = \tilde{V}_1^{-1}[\tilde{V}_0^{-1} + \tilde{V}_1^{-1}]^{-1}$, $\bar{\tilde{Y}}_1 = \sum_{i=1}^{n} \tilde{d}_i \tilde{Y}_{1i}$, and $\bar{\tilde{Y}}_0 = \sum_{i=1}^{n} \tilde{d}_i \tilde{Y}_{0i}$. The pseudo EL ratio statistic on the parameter of interest, θ, is given by $r_{pel}(\theta) = r_{pel}(\theta, \hat{\mu}(\theta))$. Using the approximations (4.20), (4.21) and (4.22), and the Taylor series expansion, we can show that

$$\begin{aligned} -2r_{pel}(\theta,\hat{\mu}) &= 2\Big\{ \sum_{i=1}^{n} \tilde{d}_i \log(1+\lambda_1(\tilde{Y}_{1i}-\hat{\mu}-\theta)) + \sum_{i=1}^{n} \tilde{d}_i \log(1+\lambda_0(\tilde{Y}_{0i}-\hat{\mu})) \Big\} \\ &\doteq \frac{1}{(\tilde{V}_1+\tilde{V}_0)} (\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta)^2. \end{aligned}$$

It follows that $-2r_{pel}(\theta)/c_{pel}$ converges in distribution to a χ_1^2 random variable when $\theta = \mu_1 - \mu_0$ as $n \to \infty$ and $n_1/n \to \delta \in (0,1)$. The scaling constant c_{pel} is given by

$$c_{pel} = E_\xi E_p (\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta)^2 / (\tilde{V}_1 + \tilde{V}_0).$$

Let $\theta_N = E_p(\hat{\theta}_{pel})$. Then,

$$\begin{aligned} E_\xi E_p (\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta)^2 &= E_\xi E_p \{ (\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta_N)^2 + (\theta_N - \theta)^2 + 2(\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0 - \theta_N)(\theta_N - \theta) \}, \\ &= E_\xi V_p (\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0) + \text{Var}_\xi(\theta_N) + 0\,, \\ &= E_\xi V_p (\bar{\tilde{Y}}_1 - \bar{\tilde{Y}}_0) + o_p(1). \end{aligned}$$

4.6 Concluding Remarks

In this chapter, we studied the empirical likelihood methods for pretest-posttest studies with complex survey data. We first considered the setting based on the ITC 4C Wave 7-8 data. The potential confounding problem in the observational study was adjusted using the propensity score method. We developed the estimators based on both propensity score stratification and the pseudo EL method with propensity score weighting. We also derived the variance estimators for our proposed estimators. For the pseudo EL estimator, we developed the asymptotic distribution for the pseudo EL ratio statistic, which can be used for constructing a confidence interval for the parameter of interest. As an alternative to the normal based confidence interval, the confidence interval based on the pseudo EL ratio statistic has nice properties stemming from its use of the EL method. We applied our proposed methods to the ITC 4C Wave 7-8 data. For the combined sample of the email invitation group and the no email invitation group, the difference of the answers to question hFR245v ("cigarettes per day" at Wave 8) is not statistically significant between different survey modes. In the data analysis separately for the email invitation group and the no email invitation group, while we found a significant difference of the answers to hFR245v between the two survey modes for the email invitation group, the difference was not significant for the no email invitation group. Later in the chapter, we considered a simpler setting where we have a randomized pretest-posttest study with survey data, and our proposed method is based on an imputation approach and the two-sample pseudo EL approach in Wu and Yan (2012).

Chapter 5

Summary and Future Research

5.1 Summary and Future Research Topics

We conclude this thesis with a brief summary and discussion of possible future research topics. In Chapter 2, we proposed an imputation based two-sample EL approach to efficiently estimate the treatment effect of pretest-posttest studies. We derived the asymptotic properties of our proposed estimators and compared their efficiency with that of one of the methods proposed by Huang et al. (2008) (HQF). We demonstrated both in theory and in simulation studies that our imputation based EL estimators are as efficient as the HQF estimator under a correctly specified working model. Moreover, the kernel regression imputation approach provides a robust alternative against model misspecification. The kernel regression imputation based EL estimator is more efficient than the HQF estimator under a misspecified working model. The materials in this chapter formulate the paper Chen et al. (2014a).

In Chapter 3, we studied the problem of testing the difference of distributions of the treatment group and control group for pretest-posttest studies. We proposed an EL based Mann-Whitney test using the HQF estimators and derived the asymptotic distribution for the test statistic. We also considered the imputation based two-sample EL and the jackknife EL method for the Mann-Whitney test. Due to the technical difficulty of deriving the asymptotic distribution of the two-sample EL and the jackknife EL ratio statistics, we applied a bootstrap calibration method. We further proposed a two-sample jackknife EL method for the Mann-Whitney test with less demanding computational procedures and with flexibility in incorporating baseline information. In the future work, it is of our great interest to derive the asymptotic properties of the two-sample jackknife EL ratio statistic for the Mann-Whitney test, and develop a theoretical justification of the bootstrap procedure which we used to approximate the asymptotic distribution of the two-sample jackknife EL ratio statistic. More specifically, we will pursue two research problems. The first is on the weighted EL method in conduction with the jackknife EL method, with theoretical development on the asymptotic distributions of the empirical likelihood ratio statistics. The second is on a potential adjustment to the imputation-based method. Our simulation studies showed that the current form of the method described in Chapter 3 has test sizes below the nominal values under the null hypothesis but also has very large powers under the alternatives. If we could make adjustment for the test to have the right test size, the imputation-based test would become a potentially very powerful method. The materials presented in Chapter 3 formulate the paper Chen et al. (2014b).

In Chapter 4, we extended our discussion of the empirical likelihood method for pretest-posttest studies to the context of complex survey data. Our motivating problem is to analyze the "effect" of the web survey mode for the ITC 4C survey. The methods we developed in this chapter address the complex survey design, the confounding problem from an observational study and the design feature of pretest-posttest studies. We considered the estimators based on propensity score stratification and propensity score weighting. We applied the two-sample pseudo EL method to construct the confidence intervals of our proposed propensity score weighting estimator. In the data analysis, the confidence intervals constructed using the two-sample pseudo EL ratio statistic are very close to the ones based on the normal approximation. In the future work, we will conduct simulation studies to evaluate the finite sample performance of our proposed estimators. For our proposed methods, the baseline information was not incorporated into the estimation process except for modelling the propensity score. Therefore, a possible future research topic is to investigate methods which effectively incorporate the baseline information for pretest-posttest studies with complex survey data.

5.2 Discussion on Recommendations

In this section, we discuss some recommendations for the real data applications of our proposed methods in this thesis. When the goal is to make inference of the treatment effect, we consider the methods discussed in Chapter 2. If the number of baseline covariates is small, we recommend the kernel regression imputation based EL method, since the kernel imputation method effectively incorporates baseline information and is robust against model misspecification. When the number of baseline covariates is large, the linear regression imputation based EL method is a good choice since it is most efficient and easy to implement. However, it is important to conduct model diagnosis of the validity of linear model assumptions before using the linear regression imputation method.

When the objective is to test the difference of the distributions of the posttest responses, we consider the methods studied in Chapter 3. The adjusted Mann-Whitney test based on the HQF estimators has the following advantages: (1) asymptotic normality of the test statistic, (2) fast computation, (3) effective incorporation of the baseline information, and (4) higher power in the simulation studies. The imputation based EL and JEL tests have great potential to be more powerful after further adjustments to the size of the tests. We will investigate this topic in our future work.

Moreover in our future research, we plan to write the proposed methods in this thesis into R functions so that our methods are ready for use by statisticians and applied science researchers.

References

Austin, P. C. (2008). A critical appraisal of propensity score matching in the medical literature from 1996-2003. *Statistics in Medicine*, 27:2037–2049.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28:3083–3107.

Brogan, D. R. and Kutner, M. H. (1980). Comparative analyses of pretest-posttest research designs. *The American Statistician*, 34(4):229–232.

Cao, R. and van Keilegom, I. (2009). Empirical likelihood tests for two-sample problems via nonparametric density estimation. *Canadian Journal of Statistics*, 34(1):64–77.

Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 80(1):107–116.

Chen, J. and Sitter, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, 9:385–406.

Chen, M., Wu, C., and Thompson, M. E. (2014a). An imputation based empirical likelihood approach to pretest-posttest studies. *The Canadian Journal of Statistics*, under minor revision.

Chen, M., Wu, C., and Thompson, M. E. (2014b). Mann-Whitney test with empirical likelihood methods for pretest-posttest studies. *Journal of Nonparametric Statistics*, submitted.

Chen, S., Leung, D. H. Y., and Qin, J. (2003). Information recovery in a study with surrogate endpoints. *Journal of the American Statistical Association*, 98(464):1052–1062.

Chen, S., Qin, J., and Tang, C. (2013). Mann-Whitney test with adjustments to pretreatment variables for missing values and observational study. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)*, 75(1):81–102.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Jounral of the American Statistical Association*, 89:81–87.

Crager, M. R. (1987). Analysis of covariance in parallel-group clinical trials with pretreatment baselines. *Biometrics*, 43(4):895–901.

Davidian, M., Tsiatis, A. A., and Leon, S. (2005). Semiparametric estimation of treatment effect in a pretest-posttest study with missing data. *Statistical Science*, 20(3):261–301.

Follmann, D. A. (1991). The effect of screening on some pretest-posttest test variances. *Biometrics*, 47(2):763–771.

Fu, Y., Wang, X., and Wu, C. (2009). Weighted empirical likelihood inference for multiple samples. *Journal of Statistical Planning and Inference*, 139:1462–1473.

Hajducek, D. M., Boudreau, C., and Thompson, M. E. (2012). Propensity score weight adjustment for dual-mode or dual-frame longitudinal surveys. In *JSM Proceedings, Section on Survey Research Methods*, pages 4459–4473. American Statistical Association, Alexandria, VA.

Hájek, J. (1971). Contribution to the discussion of "An essay on the logical foundations of survey sampling, Part One" by D. Basu. In Godambe, V. P. and Sprott, D. A., editors, *Foundations of Statistical Inference*, page 236. Holt, Rinehart and Winston, Toronto.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., and et al. for the AIDS Clinical Trials Group Study 175 Study Team (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD 4 cell counts from 200 to 500 per cubic millimetre. *The New England Jounral of Medicine*, 335:1081–1089.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685.

Huang, C. Y., Qin, J., and Follmann, D. A. (2008). Empirical likelihood-based estimation of the treatment effect in a pretest-posttest study. *Journal of the American Statistical Association*, 103(483):1270–1280.

Jing, B.-Y. (1995). Two-sample empirical likelihood method. *Statistics and Probability Letters*, 24:315–319.

Jing, B.-Y., Yuan, J., and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association*, 104:1224–1232.

Laird, N. (1983). Further comparative analyses of pretest-posttest research designs. *The American Statistician*, 37(4a):329–330.

Leon, S., Tsiatis, A. A., and Davidian, M. (2003). Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics*, 59(4):1046–1055.

Liang, H., Wang, S., and Carroll, R. J. (2007). Partially linear models with missing response variables and error-prone covariates. *Biometrika*, 94(1):185–198.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley and Sons.

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23:2937–2960.

Mann, H. and Whitney, D. (1947). On a test whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18:50–60.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*, 18(1):90–120.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC.

Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, 85(3):619–630.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, 22(1):300–325.

Qin, J. and Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 84(3):609–618.

Qin, J., Zhang, B., and Leung, D. H. Y. (2009). Empirical likelihood in missing data problems. *Journal of the American Statistical Association*, 104(488):1492–1503.

Rao, J. N. K. and Wu, C. (2009). Empirical likelihood methods. In Pfeffermann, D. and Rao, C., editors, *Handbook of Statistics: Sample Surveys: Theory, Methods, and Inference*, volume 29, pages 189–207.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866.

Rosenbaum, P. R. (1998). Propensity score. In Armitage, P. and Colton, T., editors, *Encyclopedia of Biostatistics*, volume 5, pages 3551–3555. Wiley: New York.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, 6:34–58.

Shi, X. (1984). The approximate independence of jackknife pseudo-values and the bootstrap methods. *Journal of Wuhan Institute Hydra-Electric Engineering*, 2:83–90.

Stanek, E. J. (1988). Choosing a pretest-posttest analysis. *The American Statistician*, 42(3):178–183.

Tsao, M. and Wu, C. (2006). Empirical likelihood inference for a common mean in the presence of heteroscedasticity. *The Canadian Journal of Statistics*, 34:45–59.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Wang, D. and Chen, S. X. (2009). Empirical likelihood for estimating equations with missing values. *Annals of Statistics*, 37(1):490–517.

Wang, Q. and Rao, J. N. K. (2002). Empirical likelihood-based inference under imputation for missing response data. *Annals of Statistics*, 30:896–924.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1:80–83.

Wu, C. (2004). Some algorithmic aspects of the empirical likelihood method in survey sampling. *Statistica Sinica*, 14:1057–1067.

Wu, C. and Rao, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics*, 34:359–375.

Wu, C. and Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96(453):185–193.

Wu, C. and Yan, Y. (2012). Empirical likelihood inference for two-sample problems. *Statistics and Its Interface*, 5(3):345–354.

Yang, L. and Tsiatis, A. A. (2001). Efficiency study of estimators for a treatment effect in a pretest-posttest trial. *The American Statistician*, 55(4):314–321.