DISSERTATION

SOME TOPICS ON SURVEY ESTIMATORS UNDER SHAPE CONSTRAINTS

Submitted by

Xiaoming Xu

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2021

Doctoral Committee:

Advisor: Mary C. Meyer

F. Jay Breidt
Wen Zhou
Edwin K. P. Chong

ABSTRACT

SOME TOPICS ON SURVEY ESTIMATORS UNDER SHAPE CONSTRAINTS

We consider three topics in this dissertation: 1) Nonresponse weighting adjustment using estimated response probability; 2) Improved variance estimation for inequality constrained domain mean estimators in surveys; and 3) One-sided testing of population domain means in surveys.

Weighting by the inverse of the estimated response probabilities is a common type of adjustment for nonresponse in surveys. In the first topic, we propose a new survey estimator under nonresponse where we set the response model in linear form and the parameters are estimated by fitting a constrained least square regression model, with the constraint being a calibration equation. We examine asymptotic properties of Horvitz-Thompson and Hájek versions of the estimators. Variance estimation for the proposed estimators is also discussed. In a limited simulation study, the performances of the estimators are compared with those of the corresponding uncalibrated estimators in terms of unbiasedness, MSE and coverage rate.

In survey domain estimation, *a priori* information can often be imposed in the form of linear inequality constraints on the domain estimators. Wu et al. (2016) formulated the isotonic domain mean estimator, for the simple order restriction, and methods for more general constraints were proposed in Oliva-Avilés et al. (2020). When the assumptions are valid, imposing restrictions on the estimators will ensure that the *a priori* information is respected, and in addition allows information to be pooled across domains, resulting in estimators with smaller variance. In the second topic, we propose a method to further improve the estimation of the covariance matrix for these constrained domain estimators, using a mixture of possible covariance matrices obtained from the inequality constraints. We prove consistency of the improved variance estimator, and simulations demonstrate that the new estimator results in improved coverage probabilities for domain mean confidence intervals, while retaining the smaller confidence interval lengths.

Recent work in survey domain estimation allows for estimation of population domain means under *a priori* assumptions expressed in terms of linear inequality constraints. Imposing the constraints has been shown to provide estimators with smaller variance and tighter confidence intervals. In the third topic, we consider a formal test of the null hypothesis that all the constraints are binding, versus the alternative that at least one constraint is non-binding. The test of constant versus increasing domain means is a special case. The power of the test is substantially better than the test with an unconstrained alternative. The new test is used with data from the National Survey of College Graduates, to show that salaries are positively related to the subject's father's educational level, across fields of study and over several years of cohorts.

ACKNOWLEDGEMENTS

I am very grateful for the support and guidance of many people during my graduate experience.

Firstly, I would like to express my sincere gratitude to my research advisor, Dr. Mary C. Meyer for the continuous support over the years of my Ph.D. study, and her endless patience, encouragement and research insights. It was my fortune to work and study under her guidance. I benefited a lot from the weekly meeting with her and she has been helping me all the time with my research works. Also, I would like to thank my ex-advisor, Dr. Jean D. Opsomer for the mentoring and advice during the first three years of my graduate study. Under his direction, I gained a lot of knowledge in non-parametric statistics and survey statistics, which play an important role in my future research.

Secondly, I would like to thank my committee members: Dr. F. Jay Breidt, Dr. Wen Zhou and Dr. Edwin K. P. Chong, for serving in my committee and providing their insightful feedbacks.

Thanks to my wife Zhuo Zhang for her hard work in taking care of our babies. I could not imagine this challenging journey without all sacrifices she has done for me. Thanks to my three lovely daughters, Rebecca Xu, Isabel Xu and Deborah Xu. They will always be my motivation to work hard and to contribute through my knowledge and actions to build a better world.

Thanks to Kristin Stephens, Katy Jackson and Karena Alons. They are amazing staffs of the statistics department, who have been always ready to help me since my first day as a graduate student.

I also want to thank my fellow graduate students and friends Xin Chen, Jiarui Chi, Robert McAndrew, Ben Zheng, Xiyue Liao, Soo-Young Kim, Josh Hewitt and Cristian M. Oliva Avilés. Working together and becoming friends with them was such an amazing experience in my life. I really appreciate for their support as I pursued my degree.

Lastly, I would like to thank the Department of Statistics of Colorado State University for the financial support during my Ph.D. study. Also, I want to acknowledge the National Science

Foundation for providing funding to allow me to conduct my research. Part of my research works was funded by NSF MMS 1533804.

DEDICATION

I would like to dedicate this dissertation to my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivation

1.1.1 Nonresponse Weighting in Surveys

The goal of conducting a survey is to estimate some characteristics of a target finite population. These characteristics can take many forms. In real applications, quantitative summaries such as means, totals and proportions of the study variable are of the most common interests. Without loss of generality, we focus on the estimation of the finite population mean $\bar{y}_U = \frac{1}{N}\sum_{i \in U} y_i$, where y_i denote the non-random value of a variable of interest for the ith element in the finite population $U = \{1, 2, \cdots, N\}$. Since collecting data from the entire population is infeasible, we usually randomly select sample units to measure, and then construct estimators that rely on the random sampling design. Thus, the estimators we constructed will incorporate design information. Let p be the sampling design and $p(s)$ be the probability of selecting the particular sample s. Define the sample membership indicator $I_i = 1$ if $i \in s$ and $I_i = 0$ otherwise. For $i \in U$, let $\pi_i = E(I_i) = \Pr(i \in s) = \sum_{s \subset U, i \in s} p(s)$ denote the first-order inclusion probabilities of the design p. Then the Horvitz-Thompson estimator (Horvitz and Thompson (1952)) incorporates design information via inverse-probability weighting as follows:

$$\bar{y}_\pi = \frac{1}{N}\sum_{i \in s}\frac{y_i}{\pi_i} = \frac{1}{N}\sum_{i \in U}\frac{y_i I_i}{\pi_i}$$

and $\bar{y}_\pi$ is design unbiased for $\bar{y}_U$.

However, in real surveys, it is often the case that some sampled units do not respond (unit nonresponse). Nonresponse will cause a loss in the precision of survey estimates due to reduced sample size. Also, nonresponse bias generally does not decrease as the sample size increases and thus bias is often the largest component of mean square error of the estimates. To deal with this issue, weighting adjustment is widely used to correct for the potential biasing impact of nonresponse. One way of weight adjustment is to model the response propensities for the sampled units individually, and the adjustment factor is the inverse of the estimated propensities of the respondents. The idea is to estimate the unknown probability of response. In many situations, response propensity modeling may be the main tool to deal with nonresponse problem. Comprehensive descriptions of nonresponse weighting adjustment (NWA) methods in survey sampling are provided by Groves et al. (2002); Sarndal and Lundstrom (2006); and Cassel et al. (1983); Ekholm and Laaksonen (1991); Folsom and Singh (2000); and Iannacchione (2003).

The response probabilities are usually estimated by logistic regression, but probit and non-parametric methods are also used; see Little (1986), Silva and Opsomer (2009), Phipps and Toth (2012) and so on for more details. However, the estimation procedure associated with these methods are complex and thus it is rare to implement these technique in real surveys. Also, we need to specify the response model correctly. Because if the mechanisms that cause unit nonresponse are not adequately reflected in the model specification, survey estimates may be biased even after the weighting adjustments. Another approach is to use calibration estimation for adjustment, see Deville and Särndal (1992). One advantage of calibration method is that it has good performance if the calibrated auxiliary variable is highly correlated to the study variable. Also, it is easy to understand and implement for survey practitioners. Due to the attractive properties of calibration method, we try to develop a new survey estimator under nonresponse where we set the response model in linear form and we estimate the model by using a constrained least square criterion, with the constraint being a calibration equation.

1.1.2 Variance Estimation of Shape Restricted Survey Domain Mean Estimator

In many large-scale surveys, fine-scale domain estimates are of clear interest for many data users, as they provide a lot of useful information. One of the most frequent parameters of interest are the population domain means. For example, National Compensation Survey, conducted by the

U.S. Bureau of Labor Statistics, is designed to provide wage and salary estimates by occupation for many metropolitan areas. The usual domain survey estimators depend only on the domain specific sampled data. These survey estimators are design-based estimator, since the estimation and inference are implemented by using certain survey weights that are determined by a specific probability sampling design.

To establish the notation, a finite population is denoted as $U = \{1, 2, \cdots, N\}$ and let $\{U_d : d = 1, \cdots, D\}$ be a partition of the population U, where D is the number of domains. Let y be the variable of interest and denote by y_i the value for the ith unit in the population. Then, the parameter of interest are $\bar{\boldsymbol{y}}_U = (\bar{y}_{U_1}, \cdots, \bar{y}_{U_D})^\top$, and $\bar{y}_{U_d}$ is given by:

$$\bar{y}_{U_d} = \frac{\sum_{i \in U_d} y_i}{N_d}, \quad d = 1, \cdots, D.$$

where N_d is the population size of domain d.

Based on a sampling design p, a sample s of size n is drawn from U and $p(s)$ is the probability of drawing the sample s. The first order inclusion probability $\pi_i = \text{Pr}(i \in s) = \sum_{i \in s} p(s) = E(I_i)$ and the second order inclusion probability $\pi_{ij} = \text{Pr}(i, j \in s) = \sum_{i,j \in s} p(s) = E(I_i I_j)$ are both assumed to be positive for all $i, j \in U$. We denote by s_d the intersection of s and U_d, and let n_d be the sample size for s_d. Then, two common design-based estimators are the Horvitz-Thompson estimator (Horvitz and Thompson (1952)) and the Hájek estimator (Hájek (1971)). Since the Hájek estimator does not require information about the population domain size N_d and is more popular in practice, we will focus on the Hájek estimator $\tilde{\boldsymbol{y}}_s = (\tilde{y}_{s_1}, \cdots, \tilde{y}_{s_D})^\top$ as an illustration, where the estimator $\tilde{y}_{s_d}$ of domain d is given by:

$$\tilde{y}_{s_d} = \frac{\sum_{i \in s_d} y_i / \pi_i}{\hat{N}_d}$$

and $\hat{N}_d = \sum_{i \in s_d} 1/\pi_i$. The ijth element of the asymptotic covariance matrix $AV(\tilde{\boldsymbol{y}}_s) = \boldsymbol{\Sigma}$ is:

$$\boldsymbol{\Sigma}_{ij} = \frac{1}{N_i N_j} \sum_{k \in U_i} \sum_{l \in U_j} \Delta_{kl} \frac{(y_k - \bar{y}_{U_i})(y_l - \bar{y}_{U_j})}{\pi_k \pi_l}, \quad i, j = 1, 2, \cdots, D.$$

where $\Delta_{kl} = cov(I_k I_l) = \pi_{kl} - \pi_k \pi_l$. The corresponding estimator $\tilde{\boldsymbol{\Sigma}}_{ij}$ for $\boldsymbol{\Sigma}_{ij}$ is given by:

$$\tilde{\boldsymbol{\Sigma}}_{ij} = \frac{1}{\hat{N}_i \hat{N}_j} \sum_{k \in s_i} \sum_{l \in s_j} \frac{\Delta_{kl}}{\pi_{kl}} \frac{(y_k - \tilde{y}_{s_i})(y_l - \tilde{y}_{s_j})}{\pi_k \pi_l}, \quad i, j = 1, 2, \cdots, D.$$

In real large-scale surveys, although the overall sample size might be very large, it is very often that there could be domains of interest with samples sizes that are too small to produce estimates with acceptable precision. There exist several statistical methods to deal with this small area estimation problem. One novel approach is to incorporate the shape-constrained regression techniques into the survey domain estimation and inference. Shape restrictions, that can arise naturally in the survey context, are often expected to be respected by population domain means. For example, younger people are expected to have, on average, lower glucose level than older people, certain jobs might be expected to receive higher salaries than others. The constrained domain mean estimators that respect reasonable shape restrictions have the potential to improve precision and stability of the estimators, while the unconstrained Hájek estimators are very likely to violate those constraints and thus produce unstable and inaccurate estimates, especially when sample size n is small.

Wu et al. (2016) firstly proposed a constrained estimator that respects the monotone assumption along the domains. Such isotonic survey estimators were shown to improve precision, compared with the unconstrained design-based estimators, given that the monotonicity assumption is reasonable. Recently, Oliva-Avilés et al. (2020) proposed the methods for more general constraints in the estimation and inference of population domain means. In particular, for a given sample s, the constrained domain mean estimator $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \cdots, \tilde{\theta}_D)^\top$ has the following explicit form:

$$\tilde{\boldsymbol{\theta}} = \left(\boldsymbol{I}_{D \times D} - \boldsymbol{W}_s^{-1} \boldsymbol{A}_J' (\boldsymbol{A}_J \boldsymbol{W}_s^{-1} \boldsymbol{A}_J')^{-1} \boldsymbol{A}_J \right) \tilde{\boldsymbol{y}}_s, \tag{1.1}$$

where $\boldsymbol{W}_s$ is the diagonal matrix with elements $\hat{N}_1/\hat{N}$, $\hat{N}_2/\hat{N}$, $\cdots, \hat{N}_D/\hat{N}$, $\boldsymbol{A}$ is the $m \times D$ constrained matrix in which each row defines a constraint on the domains, the observed $J \subset \{1, \cdots, m\}$ is determined by the cone projection algorithm and $\boldsymbol{A}_J$ denote the matrix formed by the rows of $\boldsymbol{A}$ indexed by J.

However, one drawback for the work of Oliva-Avilés et al. (2020) is that the variance of the constrained estimator is implicit and hard to implement in practice. As was shown in Oliva-Avilés et al. (2020), for the observed set J, the variance of $\tilde{\theta}_d$ can be approximated by:

$$AV(\tilde{\theta}_d) = \sum_{k \in U} \sum_{l \in U} \Delta_{kl} \frac{u_k}{\pi_k} \frac{u_l}{\pi_l},$$

where

$$u_k = \sum_{i=1}^{D} \alpha_i y_k I(k \in U_i) + \sum_{i=1}^{D} \beta_i I(k \in U_i), \quad k = 1, 2, \cdots, N.$$

and

$$\alpha_i = \frac{\partial \tilde{\theta}_d}{\partial \hat{t}_i}\Big|_{(\hat{t}_1, \cdots, \hat{t}_D, \hat{N}_1, \cdots, \hat{N}_D) = (t_1, \cdots, t_D, N_1, \cdots, N_D)},$$

$$\beta_i = \frac{\partial \tilde{\theta}_d}{\partial \hat{N}_i}\Big|_{(\hat{t}_1, \cdots, \hat{t}_D, \hat{N}_1, \cdots, \hat{N}_D) = (t_1, \cdots, t_D, N_1, \cdots, N_D)},$$

$\hat{t}_d$ is the HT estimator of $t_d = \sum_{k \in U_d} y_k$. Thus, the consistent estimator of the approximated variance of $\tilde{\theta}_d$ is given by:

$$\hat{V}(\tilde{\theta}_d) = \sum_{k \in s} \sum_{l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \frac{\hat{u}_k}{\pi_k} \frac{\hat{u}_l}{\pi_l}, \tag{1.2}$$

where

$$\hat{u}_k = \sum_{i=1}^{D} \hat{\alpha}_i y_k I(k \in s_i) + \sum_{i=1}^{D} \hat{\beta}_i I(k \in s_i), \quad k = 1, 2, \cdots, N.$$

with $\hat{\alpha}_i$, $\hat{\beta}_i$ obtained from α_i, β_i by substituting the appropriate Horvitz-Thompson estimators for each total population.

From (1.2), the expression of the variance estimator involves partial derivatives, which makes it hard to be applied in real practice. To address this issue, we provided a simplified version of the asymptotic variance estimator. The expression of the simplified covariance estimator is quite classical and is preferred to the one in (1.2) from both an intuitive and a computational viewpoint. Furthermore, we proposed a method to improve the estimation of the covariance matrix for the constrained domain estimators. The improved variance estimator recognizes that a different sample s with the same sample size and design might correspond to a different set J in (1.1) and it takes advantage of the mixture of all possible J sets, which better reflects the underlying variance structure. See Chapter 3 for more details.

1.1.3 Validation of Shape Constrained Domain Mean Estimators

Although it has been shown that the constrained domain mean estimator proposed by Wu et al. (2016), Oliva-Avilés et al. (2020) improves the precision of the usual design based survey estimators, it has to be used with caution because invalid population shape assumptions could lead to biased domain mean estimators. Oliva-Avilés et al. (2019) developed a diagnostic method to detect population departures from monotone assumptions. They proposed the Cone Information Criterion for Survey data (CICs) as a data-driven criterion for choosing between the monotone and the unconstrained domain mean estimators. However, a more general shape constrained test needs to be developed. Particularly, we are interested in testing:

$$H_0 : \boldsymbol{A}\bar{\boldsymbol{y}}_U = 0 \quad vs \quad H_1 : \boldsymbol{A}\bar{\boldsymbol{y}}_U \geq 0$$

and $\boldsymbol{A}\bar{\boldsymbol{y}}_U$ has at least one positive element.

This one-sided test has been widely studied outside of the survey context. Under the normal-error model assumption, the null distribution of the likelihood-ratio test statistic for the one-sided test has been derived in many literatures, see Bartholomew (1961), McDermott and Mudholkar (1993), Robertson et al. (1988), Meyer (2003), Silvapulle and Sen (2005) and so on for more details. In summary, when the model variance is known, the null distribution of the likelihood ratio statistic is shown to have a mixture of chi-square distributions. When the model variance is unknown, the test statistic has a mixture of beta distributions under the null. Also, it has been proved that the one-sided test can provide higher power than the test using the unconstrained alternative.

In Chapter 4, we try to extend the techniques of one-sided test into the survey context. The main goal is to formulate a formal testing procedure that can be used to validate the use of the shape constrained domain estimator over the unconstrained estimator.

The following section of this chapter presents some key points on the shape-constrained estimation, which plays an important role for understanding the materials in Chapter 3 and Chapter 4.

1.2 Preliminaries

Let $\boldsymbol{z}$ be an arbitrary vector in $\mathbb{R}^D$ and $\boldsymbol{A}$ be the $m \times D$ *irreducible* constraint matrix (for now assuming $\boldsymbol{A}$ is full row rank). Meyer (1999) defined a matrix as *irreducible* where none of its rows is a positive linear combinations of the other rows, and the origin is not a positive linear combination of its rows. Intuitively, a constraint matrix is *irreducible* when there is no redundant constraints.

The solution $\hat{\boldsymbol{\phi}}$ to the following constrained least-squares problem:

$$\min_{\boldsymbol{\phi}} ||\boldsymbol{z} - \boldsymbol{\phi}||^2 \quad \text{such that } \boldsymbol{A\phi} \geq \boldsymbol{0}$$

is exactly the projection of $\boldsymbol{z}$ onto the convex cone:

$$\mathcal{C} = \{\boldsymbol{\phi} \in \mathbb{R}^D : \, \boldsymbol{A\phi} \geq \boldsymbol{0}\}.$$

A set is a cone if for every $\boldsymbol{\phi}$ in the set, all positive multiples of $\boldsymbol{\phi}$ are also in the set. If $\mathcal{C}$ is convex cone, then for any $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$ in $\mathcal{C}$, $\alpha\boldsymbol{\phi}_1 + (1-\alpha)\boldsymbol{\phi}_2$ is in $\mathcal{C}$ for all $\alpha \in (0,1)$. The necessary and sufficient conditions for a vector $\hat{\boldsymbol{\phi}}$ to be the projection of $\boldsymbol{z}$ onto $\mathcal{C}$ are

$$\left\langle \boldsymbol{z} - \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\phi}} \right\rangle = 0, \quad \text{and} \quad \left\langle \boldsymbol{z} - \hat{\boldsymbol{\phi}}, \boldsymbol{\phi} \right\rangle \leq 0 \text{ for all } \boldsymbol{\phi} \in \mathcal{C}. \tag{1.3}$$

Define $\boldsymbol{\Omega} = \mathcal{C} \cap V^{\perp}$,where $V^{\perp}$ refers to the orthogonal complement of V. Then it can be shown that $\boldsymbol{\Omega}$ is a closed convex cone and the projection of $\boldsymbol{z}$ onto $\mathcal{C}$ is the sum of the projections onto $\boldsymbol{\Omega}$ and V, respectively. Furthermore, The convex cone $\boldsymbol{\Omega}$ can be specified by a set of generators

$\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_m \in \boldsymbol{\Omega}$; that is, we can express $\boldsymbol{\Omega}$ as:

$$\boldsymbol{\Omega} = \{\boldsymbol{\phi} \in \mathbb{R}^D : \boldsymbol{\phi} = \sum_{j=1}^{m} b_j \boldsymbol{\delta}_j, b_j \geq 0, j = 1, \cdots, m\},$$

where the generators (or edges) of $\boldsymbol{\Omega}$ are the columns of $\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}$ when $\boldsymbol{A}$ is full row rank. Hence, we can write the constraint cone $\mathcal{C}$ as:

$$\mathcal{C} = \{\boldsymbol{\phi} \in \mathbb{R}^D : \boldsymbol{\phi} = \boldsymbol{v} + \sum_{j=1}^{m} b_j \boldsymbol{\delta}_j, b_j \geq 0, j = 1, \cdots, m \quad \text{and} \quad \boldsymbol{v} \in V\}.$$

Now, we define the polar cone $\mathcal{C}^o$ as $\mathcal{C}^o = \{\boldsymbol{\rho} \in \mathbb{R}^D : \langle \boldsymbol{\rho}, \boldsymbol{\phi} \rangle \leq 0, \text{for all } \boldsymbol{\phi} \in \mathcal{C}\}$. It can be shown that the polar cone can be generated by the rows of $-\boldsymbol{A}$, that is, we can express $\mathcal{C}^o$ as:

$$\mathcal{C}^o = \{\boldsymbol{\rho} \in \mathbb{R}^D : \boldsymbol{\rho} = \sum_{j=1}^{m} b_j \boldsymbol{\gamma}_j, b_j \geq 0, j = 1, \cdots, m\},$$

where $\boldsymbol{\gamma}_j$, $j = 1, \cdots, m$, are the rows of $-\boldsymbol{A}$. Let $\hat{\boldsymbol{\rho}}$ be the projection of $\boldsymbol{z}$ onto $\mathcal{C}^o$. Meyer (1999) showed a very important fact that the projection of $\boldsymbol{z}$ onto $\mathcal{C}^o$ is the residual of the projection of $\boldsymbol{z}$ onto $\mathcal{C}$ and vice-versa. This result is quite useful in practice, because it allows us to compute the projection onto $\mathcal{C}$ by finding first the projection onto the polar cone $\mathcal{C}^o$, which has known edges. The necessary and sufficient conditions in (1.3) can be adapted to the polar cone as follows: the vector $\hat{\boldsymbol{\rho}} \in \mathcal{C}^o$ to minimize $||\boldsymbol{z} - \boldsymbol{\rho}||^2$ over $\mathcal{C}^o$ are satisfying

$$\left\langle \boldsymbol{z} - \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\rho}} \right\rangle = 0, \quad \text{and} \quad \left\langle \boldsymbol{z} - \hat{\boldsymbol{\rho}}, \boldsymbol{\gamma}_j \right\rangle \leq 0 \quad \forall j = 1, \cdots, m. \tag{1.4}$$

Based on conditions (1.4), it can be shown that the projection $\hat{\boldsymbol{\rho}}$ of $\boldsymbol{z}$ onto the polar cone $\mathcal{C}^o$ is exactly the projection of $\boldsymbol{z}$ onto the linear space generated by the edges $\boldsymbol{\gamma}_j$ such that $\langle \boldsymbol{z} - \tilde{\boldsymbol{\rho}}, \boldsymbol{\gamma}_j \rangle = 0$. That is, there exists a set $J \subseteq \{1, \ldots, m\}$ such that the projection of $\boldsymbol{z}$ onto $\mathcal{C}^o$ coincides with the projection of $\boldsymbol{z}$ onto the linear space spanned by $\boldsymbol{\gamma}_j$, for $j \in J$. The `R` package `coneproj` (Liao and Meyer (2014)) will perform the cone projection, returning both the projection and the set J. If J is empty, then the constrained estimator coincides with the unconstrained estimator.

Therefore, if we project arbitrary $\boldsymbol{z} \in \mathbb{R}^D$ onto $\mathcal{C}^o$, then there must exist a set J such that we can write $\hat{\boldsymbol{\rho}} = \sum_{j \in J} b_j \boldsymbol{\gamma}_j, b_j > 0$ for $j \in J$, and this representation is unique by the Karush-Kuhn-Tucker (KKT) conditions. Thus, for any $\boldsymbol{z} \in \mathbb{R}^D$, it can be expressed as:

$$\boldsymbol{z} = \boldsymbol{v} + \sum_{j \in J} b_j \boldsymbol{\gamma}_j + \sum_{j \notin J} b_j \boldsymbol{\delta}_j, \tag{1.5}$$

where $b_j > 0$ for $j \in J$ and $b_j \geq 0$ for $j \notin J$. Furthermore, $\boldsymbol{v}$ is the projection of $\boldsymbol{z}$ onto V, $\sum_{j \in J} b_j \boldsymbol{\gamma}_j$ is the projection of $\boldsymbol{z}$ onto $\mathcal{C}^o$ and $\sum_{j \notin J} b_j \boldsymbol{\delta}_j$ is the projection of $\boldsymbol{z}$ onto $\boldsymbol{\Omega}$.

if $\boldsymbol{A}$ is not full row rank, especially when $m > D$, then the set J in (1.5) may not be unique anymore; that is, $\boldsymbol{z}$ might have more than one expression. However, Theorem 3.1 from Oliva-Avilés et al. (2020) guarantees that the projection $\hat{\boldsymbol{\rho}} = \sum_{j \in J} b_j \boldsymbol{\gamma}_j$ is the same for all such J, and that it is always possible to find J^* that is a subset of all such J sets, and the vectors $\boldsymbol{\gamma}_j$, $j \in J^*$ form a linearly independent set. In the following, we assume J is this unique set. Then, we can write the solution $\hat{\boldsymbol{\phi}}$ as follows:

$$\hat{\boldsymbol{\phi}} = \boldsymbol{z} - \hat{\boldsymbol{\rho}} = \boldsymbol{z} - \boldsymbol{A}_J^\top (\boldsymbol{A}_J \boldsymbol{A}_J^\top)^- \boldsymbol{A}_J \boldsymbol{z},$$

where $\boldsymbol{A}_J$ denote the matrix formed by the rows of $\boldsymbol{A}$ indexed by J.

For the techniques regarding the constrained estimation, see Robertson et al. (1988), or Silvapulle and Sen (2005) for more details.

1.3 Overview

In Chapter 2, we propose a new survey estimator under nonresponse and we estimate the propensity function by fitting a constrained least square regression model, with the constraint being a calibration equation. We examine asymptotic properties of the proposed estimator both in

Horvitz-Thompson type version and Hájek type version. The performance of the proposed estimator is demonstrated through simulations.

Chapter 3 is a follow-up of the work in Oliva-Avilés et al. (2020). We first take a brief review of the formulation of the constrained domain means estimator and its variance estimator presented in Oliva-Avilés et al. (2020). Then, we provided a simplified version of the covariance estimator of the constrained domain means estimator, which is practically useful from a computational point of view. Further, a novel mixture variance estimator is proposed. It makes use of a mixture of possible covariance matrices obtained from the inequality constraints. We rigorously proved the consistency of the improved variance estimator. The simulations showed that the new estimator leads to improved coverage probabilities for domain mean confidence intervals, while retaining the smaller confidence interval lengths. Lastly, an application to the California School Data in `survey` package is carried out.

A formal one-sided test procedure for the population domain means is presented in Chapter 4. Here, we consider a test of the null hypothesis that all the constraints are binding, versus the alternative that at least one constraint is non-binding. We formulated the test statistic first and then derived the asymptotic null distribution of the test statistics under design normal assumption. Also, we showed the power of the test goes to 1 as sample size increases. The performance of the proposed test was demonstrated under a variety of simulation scenarios and we applied our test to the 2019 National Survey of College Graduates (NSCG) survey data.

A brief discussion of conclusions and future works is given in Chapter 5. All major proofs are presented in the Appendix.

Chapter 2

Nonresponse Weighting Adjustment Using Estimated Response Probability

2.1 Introduction

Weighting adjustment is widely used to correct for the potential biasing impact of nonresponse. Comprehensive overviews of nonresponse weighting adjustment methods in survey sampling are provided by Groves et al. (2002), Sarndal and Lundstrom (2006). One way to perform weight adjustment is to model the response propensities for the sampled units individually, and the adjustment factor is the inverse of the estimated propensities of the respondents. The idea is to estimate the unknown probability of response. General descriptions of the propensity weighting that adjust survey estimators for nonresponse are provided by Cassel et al. (1983). Applications of the response propensity modeling can be found in Ekholm and Laaksonen (1991), Folsom and Singh (2000) and Iannacchione (2003).

Auxiliary variables are often available in surveys, either at the population or the sample level. Commonly, the response probability is estimated by regressing on the auxiliary information parametrically, with logistic and probit regression models as common choices. See Alho (1990), Folsom (1991), Ekholm and Laaksonen (1991), and Iannacchione et al. (1991) for references. Another approach is to estimate the response propensities through nonparametric methods. Estimation of the response probabilities by kernel smoothing and local polynomial regression are considered by Giommi (1984), Silva and Opsomer (2009).

An interesting characteristic of nonresponse weighting adjustment estimators is that they tend to be more efficient than the unfeasible estimators on which they are based (i.e. those that use the true but unknown probability of responding). A clear justification for reduced variance using estimated response probability from a logistic regression model is given by Beaumont (2005). The estimator that uses estimated response probability is generally more efficient than the estimator using the true response probability as shown by Kim and Kim (2007).

When auxiliary variables are present in surveys, another approach is to use calibration estimation for adjustment. The general concept and techniques on calibration weighting and estimation are formalized by Deville and Särndal (1992). In recent years, using calibration weighting to adjust for nonresponse bias have been investigated by Lundström and Särndal (1999), Kott (2006), Chang and Kott (2008), Kott and Chang (2010). Calibration weighting and estimation are very popular nowadays. The primary reason is efficiency. Calibration over a set of carefully chosen auxiliary variables has proven to be an effective way of using known auxiliary information. If the study variable is highly correlated with the set of auxiliary variables, the gain of efficiency in estimation can be substantial.

In this paper, we proposed a new nonresponse weighting adjustment estimator using the estimated response probability by fitting a least square regression model that incorporates a calibration equation as a constraint. In theory, we show that the proposed estimators, both in Horvitz-Thompson and Hájek type, are asymptotically unbiased for the population parameter and are asymptotically normally distributed. From simulation study, we found that the performance of the Horvitz-Thompson type estimator is better than the corresponding estimator without calibration equation in terms of MSE in some situations, depending on the specification of the response model. On the other hand, the proposed Hájek type estimator is less affected by the specific settings of the response model and thus is more robust in winning over its corresponding Hájek estimator without calibration equation.

In Section 2.2, we introduce the new proposed estimator. In Section 2.3, the assumptions and some preliminary results are given. The asymptotic properties of the Horvitz-Thompson type estimator and Hájek estimator are provided in Section 2.4 and Section 2.5, respectively. In Section 2.6, we perform a simulation study to evaluate the finite sample properties of the proposed estimator. Conclusions are given in Section 2.7.

2.2 Notations and the Proposed Estimator

Let the finite population be $U = \{1, 2, ..., N\}$, where N is assumed to be known. Let $\mathcal{F}_N = \{\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_N\}$ be the population variable, where $\boldsymbol{u}_i = ({\boldsymbol{x}_i}^T, y_i)^T$ and $\boldsymbol{x}_i$ is the vector of auxiliary variables for unit i, which is known over the population. Our parameter of interest is $\bar{y}_U = \frac{1}{N}\sum_{i \in U} y_i$.

Given a particular probability sampling design, the inclusion of an element i in a sample S is a random event indicated by the binary random variable I_i, with:

$$I_i = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{if } i \in S. \end{cases}$$

The simplest design-based estimator is the Horvitz-Thompson estimator, defined as:

$$\bar{y}_\pi = \frac{1}{N}\sum_{i \in S}\frac{y_i}{\pi_i} = \frac{1}{N}\sum_{i \in U}\frac{y_i}{\pi_i} I_i, \tag{2.1}$$

where $\pi_i = \Pr(i \in S) = E_p(I_i)$ and ${\pi_i}^{-1}$ is called design weight of unit i. Obviously, $\bar{y}_\pi$ is an unbiased estimator for $\bar{y}_U$ with respect to sampling design p.

Under nonresponse, the study variable y_i may not be obtained for the entire set of element in S. In order to describe the response mechanism, we define the response indicator variable of y_i as:

$$R_i = \begin{cases} 1 & \text{if unit } i \text{ in the sample responds} \\ 0 & \text{if unit } i \text{ in the sample does not respond} \end{cases}$$

and denote $E(R_i) = \Pr(R_i = 1 | i \in S) = p_i$ be the response probability of sampled unit i. If we know the true response probability p_i, then the Horvitz-Thompson type estimator

$$\bar{y}_{e^*} = \frac{1}{N}\sum_{i \in S}\frac{y_i R_i}{\pi_i p_i} \tag{2.2}$$

will be unbiased for $\bar{y}_U$. Instead of Horvitz-Thompson estimator, we can also use the Hájek estimator, which is useful when the population size is unknown and shown to be more efficient in many situations. So we also consider Hájek type estimator in this paper. The Hájek type estimator is in the form:

$$\bar{y}_{H^*} = \frac{\sum_{i\in S} \frac{y_i R_i}{\pi_i p_i}}{\sum_{i\in S} \frac{R_i}{\pi_i p_i}}$$

and $\bar{y}_{H^*}$ is asymptotically unbiased for $\bar{y}_U$. However, p_i is unknown in practice, so we have to estimate it based on some specified model.

Here, we specify the response model as $E(R_i) = p_i = p(\boldsymbol{x}_i) = \boldsymbol{B}_i\boldsymbol{\nu}^*$, where $\boldsymbol{B}_i$ is a vector in which each component is a function of $\boldsymbol{x}_i$ and $\boldsymbol{\nu}^*$ is the true unknown population parameter. More specifically, we define $\boldsymbol{B}_i = (1, x_{1i}, f_2(\boldsymbol{x}_i), \cdots, f_p(\boldsymbol{x}_i))^T$, where x_1 is the covariate to be calibrated in the following criterion and supposed to be correlated with the survey variables, $f_k(\boldsymbol{x}_i)$ $(k = 2, \cdots, p)$ can be additional uncalibrated covariates or spline basis function of the uncalibrated covariates. For simplicity, we denote x, instead of x_1, as the calibrated variable in the following context. Furthermore, instead of using the usual logistic model to estimate the response probability, we specify the unknown response probability in linear form. Setting the response probability in linear form is tolerable for several reasons. First, instead of trying to interpret the response model, we just want to adjust for nonresponse. So response model specification is not that critical as long as the specified response model has decent predictive value for the true response probability. Secondly, we are cautious to apply logistic method since the logistic model may not reach convergence in some instances such as multicollinearity. As multicollinearity increases, coefficients remain unbiased but standard errors increase and the likelihood of model convergence decreases. Another reason for linearity is that it is easier to calibrate. Mathematically, linear model is much easier to deal with than logistic regression.

Now, we estimate $\boldsymbol{\nu}^*$ by $\hat{\boldsymbol{\nu}}$ using the following criterion:

$$\min_{\boldsymbol{\nu}} \sum_{i\in S} \frac{1}{\pi_i}(R_i - \boldsymbol{B}_i\boldsymbol{\nu})^2$$

subject to calibration equation:

$$\sum_{i\in S}\frac{x_i}{\pi_i}=\sum_{i\in S}\frac{x_i}{\pi_i}\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}},$$

where x_i (short for x_{1i}) is a variable that is an element of $\boldsymbol{B}_i$ and on which the calibration is implemented. In Hájek type scenario, the calibration equation is given by:

$$\frac{\sum_{i\in S}\frac{x_i}{\pi_i}}{\sum_{i\in S}\frac{1}{\pi_i}}=\frac{\sum_{i\in S}\frac{x_i}{\pi_i}\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}}{\sum_{i\in S}\frac{1}{\pi_i}\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}}.$$

However, for simplicity, we only derive the asymptotic results for Hájek estimator $\bar{y}_H$, in which the parameter $\boldsymbol{\nu}$ is estimated from above least square criterion subject to $\sum_{i\in S}\frac{x_i}{\pi_i}=\sum_{i\in S}\frac{x_i}{\pi_i}\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}$, not from calibration $\frac{\sum_{i\in S}\frac{x_i}{\pi_i}}{\sum_{i\in S}\frac{1}{\pi_i}}=\frac{\sum_{i\in S}\frac{x_i}{\pi_i}\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}}{\sum_{i\in S}\frac{1}{\pi_i}\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}}$.

Lemma 1. *Based on the criterion, the estimate $\hat{\boldsymbol{\nu}}$ is the solution to:*

$$\begin{aligned}\boldsymbol{S}(\boldsymbol{\nu})&=\frac{\partial L(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}}\\&=\sum_{i\in S}\frac{2}{\pi_i}(R_i-\boldsymbol{B}_i\boldsymbol{\nu})(-\boldsymbol{B}_i)+\frac{\sum_{i\in S}\frac{x_i}{\pi_i}[\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}-R_i-1+\boldsymbol{B}_i\boldsymbol{\nu}]}{\sum_{i\in S}\frac{x_i{}^2R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}}\sum_{i\in S}\frac{x_iR_i\boldsymbol{B}_i}{\pi_i[-(\boldsymbol{B}_i\boldsymbol{\nu})^2]}=0.\end{aligned}$$

Proof of Lemma 1. Based on the criterion, the Lagrange function is given by:

$$L(\boldsymbol{\nu})=\sum_{i\in S}\frac{1}{\pi_i}(R_i-\boldsymbol{B}_i\boldsymbol{\nu})^2+\lambda\sum_{i\in S}\frac{x_i}{\pi_i}(\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}-1).$$

Taking derivative with respect to $\boldsymbol{\nu}$ and λ respectively and setting the derivatives equal to 0, we have:

$$\frac{\partial L(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}}=\sum_{i\in S}\frac{2}{\pi_i}(R_i-\boldsymbol{B}_i\boldsymbol{\nu})(-\boldsymbol{B}_i)+\lambda\sum_{i\in S}\frac{x_iR_i\boldsymbol{B}_i}{\pi_i[-(\boldsymbol{B}_i\boldsymbol{\nu})^2]}=0 \tag{2.3}$$

and

$$\sum_{i\in S}\frac{x_i}{\pi_i}=\sum_{i\in S}\frac{x_i}{\pi_i}\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}. \tag{2.4}$$

Let S_R denote the set that the selected individual i in the sample responds. From (2.4), we have:

$$\sum_{i\in S_R}\frac{x_i}{\pi_i}\left(\frac{1}{\boldsymbol{B}_i\boldsymbol{\nu}}-1\right)=\sum_{i\notin S_R}\frac{x_i}{\pi_i}. \tag{2.5}$$

Focus on the row of $\frac{\partial L(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}}$ in which the predictor x is calibrated, we get $\sum_{i\in S}\frac{2}{\pi_i}(R_i-\boldsymbol{B}_i\boldsymbol{\nu})x_i=-\sum_{i\in S}\frac{\lambda {x_i}^2R_i}{\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}$, which, through some algebra, implies:

$$\sum_{i\notin S_R}\frac{x_i}{\pi_i}=\sum_{i\in S}\frac{x_i}{\pi_i}+\sum_{i\in S}\frac{\lambda {x_i}^2R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}-\sum_{i\in S}\frac{x_i\boldsymbol{B}_i\boldsymbol{\nu}}{\pi_i}. \tag{2.6}$$

Plugging (2.6) into (2.5), we have:

$$\sum_{i\in S_R}\frac{x_i}{\pi_i}(\frac{1}{\boldsymbol{B}_i\boldsymbol{\nu}}-1)=\sum_{i\in S}\frac{x_iR_i}{\pi_i}\left(\frac{1}{\boldsymbol{B}_i\boldsymbol{\nu}}-1\right)=\lambda\sum_{i\in S}\frac{{x_i}^2R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}+\sum_{i\in S}\frac{x_i}{\pi_i}(1-\boldsymbol{B}_i\boldsymbol{\nu}).$$

Solving for λ, we get:

$$\lambda=\frac{\sum_{i\in S}\frac{x_iR_i}{\pi_i}(\frac{1}{\boldsymbol{B}_i\boldsymbol{\nu}}-1)-\sum_{i\in S}\frac{x_i}{\pi_i}(1-\boldsymbol{B}_i\boldsymbol{\nu})}{\sum_{i\in S}\frac{{x_i}^2R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}}=\frac{\sum_{i\in S}\frac{x_i}{\pi_i}[\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}}-R_i-1+\boldsymbol{B}_i\boldsymbol{\nu}]}{\sum_{i\in S}\frac{{x_i}^2R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}}.$$

Plugging λ into (2.3), we get the desired result. $\square$

Thus, p_i is estimated by $\hat{p}_i=\boldsymbol{B}_i\hat{\boldsymbol{\nu}}$ and plugging the $\hat{p}_i$ into $\bar{y}_{e^*}=\frac{1}{N}\sum_{i\in S}\frac{y_iR_i}{\pi_ip_i}$, we get our Horvitz-Thompson type proposed estimator:

$$\bar{y}_e=\frac{1}{N}\sum_{i\in S}\frac{y_iR_i}{\pi_i\hat{p}_i}. \tag{2.7}$$

The Hájek type estimator is given by:

$$\bar{y}_H=\frac{\sum_{i\in S}\frac{y_iR_i}{\pi_i\hat{p}_i}}{\sum_{i\in S}\frac{R_i}{\pi_i\hat{p}_i}}. \tag{2.8}$$

2.3 Assumptions and Preliminary Results

First, we state the assumptions we will use in obtaining the theoretical results. The assumptions on the probability sampling design and population distribution of $\boldsymbol{u}_i$ are listed as follows:

(D.1) We assume that the sequence of finite populations of $\boldsymbol{u}_i = (\boldsymbol{x}_i{}^T, y_i)^T$ have bounded fourth moments.

(D.2) We assume the sample size n is non-random and as $N \to \infty$, $\frac{n}{N} \to \pi^* \in (0, 1)$. For all N, $\min_{i \in U} \pi_i \geq \lambda_1 > 0$, $\min_{i,j \in U} \pi_{ij} \geq \lambda_2 > 0$ and we have:

$$\lim_{N \to \infty} \sup n \max_{i,j \in U, i \neq j} |\Delta_{ij}| < \infty,$$

where $\Delta_{ij} = \mathrm{cov}(I_i, I_j) = \pi_{ij} - \pi_i \pi_j$.

(D.3) We assume the Horvitz-Thompson estimator is asymptotically normally distributed. That is,

$$\frac{\bar{y}_\pi - \bar{y}_U}{\sqrt{\mathrm{Var}(\bar{y}_\pi)}} \xrightarrow{d} N(0, 1).$$

In addition to above assumptions, we also need the following assumptions on the response model.

(R.1) The response indicator variables R_i and R_j are independent for $i \neq j$ and

$$E(R_i) = p_i,$$

$$\mathrm{Var}(R_i) = p_i(1 - p_i).$$

(R.2) The inverse true response probability is bounded. That is:

$$p_i^{-1} < K$$

for all $i \in U$, where K is a fixed constant.

(R.3) We specify the response model in the form of:

$$p_i = \boldsymbol{B}_i \boldsymbol{\nu}^*,$$

where $\boldsymbol{B}_i$ is observable over the population and the parameter is evaluated at $\boldsymbol{\nu} = \boldsymbol{\nu}^*$. Also, we assume that the response model is continuously differentiable with respect to $\boldsymbol{\nu}$.

(R.4) We assume the matrices $\mathcal{I}(\boldsymbol{\nu}^*)$, $\boldsymbol{I}(\boldsymbol{\nu}^*)$, $\boldsymbol{J}(\boldsymbol{\nu}^*)$ and $\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})$, defined in the following context, are nonsingular and thus invertible.

The assumption D.1 is a mild condition. Usually, bounded fourth moments of the study variable are required to show the variance consistency of the Horvitz-Thompson estimator.

By assuming the ratio $\frac{n}{N} \to \pi^* \in (0,1)$, we are excluding vanishing sampling fraction to stay within the finite population framework, which is often done in the design-based context. It is reasonable to say that the ratio $\frac{n}{N}$ is bounded below by π^* since usually, the ratio is decreasing in N. We need this condition to prove the asymptotic normality of our proposed estimator later. The condition $\min_{i \in U} \pi_i \geq \lambda_1 > 0$ implies that the design is a probability sampling design. The condition $\min_{i,j \in U} \pi_{ij} \geq \lambda_2 > 0$ indicates that the design is measurable, which ensures that $\hat{V}(\bar{y}_\pi) = \frac{1}{N^2} \sum_{i,j \in S} \frac{\Delta_{ij} y_i y_j}{\pi_{ij} \pi_i \pi_j}$ is unbiased for $\mathrm{Var}(\bar{y}_\pi) = \frac{1}{N^2} \sum_{i,j \in U} \frac{\Delta_{ij} y_i y_j}{\pi_i \pi_j}$. The assumption on the Δ_{ij} states that the covariance between sample membership indicators is sufficiently small. Assumptions in D.2 are satisfied for many classical sampling designs, including simple random sampling with and without replacement, and also holds for some unequal probability samplings.

Under assumption D.1 and D.2, we can derive the following result as presented in Breidt and Opsomer (2017):

$$\mathrm{Var}(\bar{y}_\pi) \leq \frac{1}{N\lambda_1} \sum_{i \in U} \frac{y_i^2}{N} + \frac{\max_{i,j \in U, i \neq j} |\Delta_{ij}|}{\lambda_1^2} \left(\sum_{i \in U} \frac{|y_i|}{N} \right)^2 = O\left(\frac{1}{N}\right) + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right).$$

Taken together with the unbiasedness of $\bar{y}_\pi$, this implies that:

$$\frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i} - \frac{1}{N}\sum_{i\in U} y_i = O_p(n^{-\frac{1}{2}}) \tag{2.9}$$

by applying Corollary 5.1.1.1 in Fuller (1996). Using the similar bounding argument, we can show the variance consistency of Horvitz-Thompson estimator. This argument for showing (2.9) will be used heavily later in proving the asymptotic variance consistency of the proposed estimator.

D.3 is usually assumed explicitly and is satisfied for many specific sampling designs, including simple random sampling without replacement, Poisson sampling and unequal probability sampling with replacement. The design asymptotic normality assumption, together with the variance consistency of the Horvitz-Thompson estimator, implies that:

$$\frac{\bar{y}_\pi - \bar{y}_U}{\sqrt{\hat{V}(\bar{y}_\pi)}} \xrightarrow{d} N(0,1).$$

In terms of the assumptions for response mechanism, R.1 and R.2 are assumed for tractability and these two conditions ensure that the order of sampling and response mechanism are interchangeable. For R.3, we specify the response model in linear form. Though the assumption for the linear expression of the response probability may not be appropriate in some cases, it is easy to implement the least square criterion in practice and we could have a closed-form solution for the parameters. Also, even though we mis-specify the response model to some degree, the estimate will be adjusted towards the “correct" value by the calibration equation, leading to an efficient estimate. R.4 is used to ensure existence of the estimator.

Before stating the main theorems of the paper, we present some preliminary results that will be used in the next section in deriving the properties of the proposed estimator.

Theorem 1. *Under the assumption D.1-D.2 and R.1-R.4, the estimator $\hat{\boldsymbol{\nu}}$ satisfies:*

$$\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^* = [\boldsymbol{I}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}),$$

where the $\boldsymbol{S}_L(\boldsymbol{\nu}^*)$ is the linearized version of $\boldsymbol{S}(\boldsymbol{\nu}^*)$, given by:

$$\boldsymbol{S}_L(\boldsymbol{\nu}^*) = \sum_{i\in S}\frac{2}{\pi_i}(R_i - \boldsymbol{B}_i\boldsymbol{\nu}^*)(-\boldsymbol{B}_i) + \frac{\sum_{i\in U}\frac{x_i\boldsymbol{B}_i}{(-\boldsymbol{B}_i\boldsymbol{\nu}^*)}}{\sum_{i\in U}\frac{x_i{}^2}{2(\boldsymbol{B}_i\boldsymbol{\nu}^*)}}\sum_{i\in S}\frac{x_i}{\pi_i}\left[\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}^*} - R_i - 1 + \boldsymbol{B}_i\boldsymbol{\nu}^*\right]$$

and $\boldsymbol{I}(\boldsymbol{\nu}^*)$ is the matrix evaluated at $\boldsymbol{\nu}^*$, defined by:

$$\boldsymbol{I}(\boldsymbol{\nu}^*) = -E\left(\left.\frac{\partial\boldsymbol{S}(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}^T}\right|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*}\right).$$

Proof of Theorem 1. Apply Taylor expansion to function $\boldsymbol{S}(\boldsymbol{\nu})$, yielding:

$$\boldsymbol{S}(\boldsymbol{\nu}) = \boldsymbol{S}(\boldsymbol{\nu}^*) + \left(\left.\frac{\partial\boldsymbol{S}(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}^T}\right|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*}\right)(\boldsymbol{\nu} - \boldsymbol{\nu}^*) + o_p(\boldsymbol{\nu} - \boldsymbol{\nu}^*).$$

Plugging in $\hat{\boldsymbol{\nu}}$ and using the fact that $\boldsymbol{S}(\hat{\boldsymbol{\nu}}) = 0$, we have $0 = \boldsymbol{S}(\boldsymbol{\nu}^*) + \left(\left.\frac{\partial\boldsymbol{S}(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}^T}\right|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*}\right)(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) + o_p(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)$, which can be written as:

$$\begin{aligned}\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^* &= \left(-\left.\frac{\partial\boldsymbol{S}(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}^T}\right|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*}\right)^{-1}\boldsymbol{S}(\boldsymbol{\nu}^*) + o_p(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)\\ &= \left(-\frac{1}{N}\left.\frac{\partial\boldsymbol{S}(\boldsymbol{\nu})}{\partial\boldsymbol{\nu}^T}\right|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*}\right)^{-1}\left(\frac{1}{N}\boldsymbol{S}(\boldsymbol{\nu}^*)\right) + o_p(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*). \qquad (2.10)\end{aligned}$$

Note that the term $\frac{1}{N}\boldsymbol{S}(\boldsymbol{\nu}^*)$ is not in linearized form, so we apply Taylor linearization to it, yielding:

$$\begin{aligned}\frac{1}{N}\boldsymbol{S}(\boldsymbol{\nu}^*) &= \frac{1}{N}\Bigg\{\sum_{i\in S}\frac{2}{\pi_i}(R_i - \boldsymbol{B}_i\boldsymbol{\nu}^*)(-\boldsymbol{B}_i) + \frac{E[\sum_{i\in S}\frac{x_iR_i\boldsymbol{B}_i}{\pi_i[-(\boldsymbol{B}_i\boldsymbol{\nu}^*)^2]}]}{E[\sum_{i\in S}\frac{x_i{}^2R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu}^*)^2}]}\\ &\qquad\times\sum_{i\in S}\frac{x_i}{\pi_i}\left[\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}^*} - R_i - 1 + \boldsymbol{B}_i\boldsymbol{\nu}^*\right]\Bigg\} + o_p(n^{-\frac{1}{2}})\\ &= \frac{1}{N}\Bigg\{\sum_{i\in S}\frac{2}{\pi_i}(R_i - \boldsymbol{B}_i\boldsymbol{\nu}^*)(-\boldsymbol{B}_i) + \frac{\sum_{i\in U}\frac{x_i\boldsymbol{B}_i}{[-\boldsymbol{B}_i\boldsymbol{\nu}^*]}}{\sum_{i\in U}\frac{x_i{}^2}{2(\boldsymbol{B}_i\boldsymbol{\nu}^*)}}\\ &\qquad\times\sum_{i\in S}\frac{x_i}{\pi_i}\left[\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}^*} - R_i - 1 + \boldsymbol{B}_i\boldsymbol{\nu}^*\right]\Bigg\} + o_p(n^{-\frac{1}{2}})\\ &= \frac{1}{N}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}).\end{aligned}$$

Define $\boldsymbol{S}_{LU}(\boldsymbol{\nu}^*) = \sum_{i\in U} 2(R_i - \boldsymbol{B}_i\boldsymbol{\nu}^*)(-\boldsymbol{B}_i) + \frac{\sum_{i\in U}\frac{x_i\boldsymbol{B}_i}{(-\boldsymbol{B}_i\boldsymbol{\nu}^*)}}{\sum_{i\in U}\frac{{x_i}^2}{2(\boldsymbol{B}_i\boldsymbol{\nu}^*)}} \sum_{i\in U} x_i \left[\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}^*} - R_i - 1 + \boldsymbol{B}_i\boldsymbol{\nu}^*\right]$, then we have:

$$
\begin{aligned}
\frac{1}{N}\boldsymbol{S}_L(\boldsymbol{\nu}^*) &= \left[\frac{1}{N}\boldsymbol{S}_L(\boldsymbol{\nu}^*) - \frac{1}{N}\boldsymbol{S}_{LU}(\boldsymbol{\nu}^*)\right] + \frac{1}{N}\boldsymbol{S}_{LU}(\boldsymbol{\nu}^*) \\
&= O_p(n^{-\frac{1}{2}}) + O_p(N^{-\frac{1}{2}}) \\
&= O_p(n^{-\frac{1}{2}}), \qquad (2.11)
\end{aligned}
$$

where $\left(\frac{1}{N}\boldsymbol{S}_L(\boldsymbol{\nu}^*) - \frac{1}{N}\boldsymbol{S}_{LU}(\boldsymbol{\nu}^*)\right) = O_p(n^{-\frac{1}{2}})$ by using a analogous argument for showing (2.9). Now, let $\mathcal{I}(\boldsymbol{\nu}^*) = -\frac{\partial \boldsymbol{S}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^T}\Big|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*}$ and denote $\mathcal{I}_U(\boldsymbol{\nu}^*)$ to be the population version of $\mathcal{I}(\boldsymbol{\nu}^*)$. That is, the sampling design is a census. Then:

$$
\begin{aligned}
\frac{1}{N}\mathcal{I}(\boldsymbol{\nu}^*) - \frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*) &= \left(\frac{1}{N}\mathcal{I}(\boldsymbol{\nu}^*) - \frac{1}{N}\mathcal{I}_U(\boldsymbol{\nu}^*)\right) + \left(\frac{1}{N}\mathcal{I}_U(\boldsymbol{\nu}^*) - \frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*)\right) \\
&= O_p(n^{-\frac{1}{2}}) + O_p(N^{-\frac{1}{2}}) \\
&= O_p(n^{-\frac{1}{2}}),
\end{aligned}
$$

where $\left(\frac{1}{N}\mathcal{I}(\boldsymbol{\nu}^*) - \frac{1}{N}\mathcal{I}_U(\boldsymbol{\nu}^*)\right) = O_p(n^{-\frac{1}{2}})$ by the similar argument for showing (2.9). Assume $\mathcal{I}(\boldsymbol{\nu}^*)$ is continuous, so applying the Taylor expansion to above result, we will have:

$$
\left\{\frac{1}{N}\mathcal{I}(\boldsymbol{\nu}^*)\right\}^{-1} = \left\{\frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*)\right\}^{-1} + O_p(n^{-\frac{1}{2}}). \qquad (2.12)
$$

Plugging (2.11) and (2.12) into (2.10), we have:

$$
\begin{aligned}
\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^* &= \left\{\left[\frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*)\right]^{-1} + O_p(n^{-\frac{1}{2}})\right\} O_p(n^{-\frac{1}{2}}) + o_p(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) \\
&= O_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

Finally, we have:

$$
\begin{aligned}
&\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^* \\
&= \left\{ \frac{1}{N} \mathcal{I}(\boldsymbol{\nu}^*) \right\}^{-1} \frac{1}{N} \boldsymbol{S}(\boldsymbol{\nu}^*) + o_p(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) \\
&= \left\{ \left[\frac{1}{N} \boldsymbol{I}(\boldsymbol{\nu}^*) \right]^{-1} + O_p(n^{-\frac{1}{2}}) \right\} \left\{ \frac{1}{N} \boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) \right\} + o_p(n^{-\frac{1}{2}}) \\
&= \left\{ \frac{1}{N} \boldsymbol{I}(\boldsymbol{\nu}^*) \right\}^{-1} \left\{ \frac{1}{N} \boldsymbol{S}_L(\boldsymbol{\nu}^*) \right\} + \left\{ \frac{1}{N} \boldsymbol{I}(\boldsymbol{\nu}^*) \right\}^{-1} o_p(n^{-\frac{1}{2}}) + O_p(n^{-\frac{1}{2}}) \left\{ \frac{1}{N} \boldsymbol{S}_L(\boldsymbol{\nu}^*) \right\} + o_p(n^{-\frac{1}{2}}) \\
&= \left\{ \frac{1}{N} \boldsymbol{I}(\boldsymbol{\nu}^*) \right\}^{-1} \left\{ \frac{1}{N} \boldsymbol{S}_L(\boldsymbol{\nu}^*) \right\} + o_p(n^{-\frac{1}{2}}) + O_p(n^{-1}) + o_p(n^{-\frac{1}{2}}) \\
&= \left\{ \boldsymbol{I}(\boldsymbol{\nu}^*) \right\}^{-1} \left\{ \boldsymbol{S}_L(\boldsymbol{\nu}^*) \right\} + o_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

□

In order to compute the asymptotic variance of $(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)$, we have to calculate the variance of $\boldsymbol{S}_L(\boldsymbol{\nu}^*)$. Before giving its variance in explicit form, we need to rewrite the form of $\boldsymbol{S}_L(\boldsymbol{\nu}^*)$ first. Let's denote $\boldsymbol{B}_i \boldsymbol{\nu}^* = p_i$ for simplicity, then $\boldsymbol{S}_L(\boldsymbol{\nu}^*)$ can be written as follows:

$$
\begin{aligned}
\boldsymbol{S}_L(\boldsymbol{\nu}^*) &= \sum_{i \in S} \frac{2}{\pi_i} (R_i - \boldsymbol{B}_i \boldsymbol{\nu}^*)(-\boldsymbol{B}_i) + \frac{\sum_{i \in U} \frac{x_i \boldsymbol{B}_i}{(-\boldsymbol{B}_i \boldsymbol{\nu}^*)}}{\sum_{i \in U} \frac{{x_i}^2}{2(\boldsymbol{B}_i \boldsymbol{\nu}^*)}} \sum_{i \in S} \frac{x_i}{\pi_i} \left[\frac{R_i}{\boldsymbol{B}_i \boldsymbol{\nu}^*} - R_i - 1 + \boldsymbol{B}_i \boldsymbol{\nu}^* \right] \\
&= \sum_{i \in S} \frac{2}{\pi_i} (R_i - p_i)(-\boldsymbol{B}_i) + \frac{\sum_{i \in U} \frac{x_i \boldsymbol{B}_i}{(-p_i)}}{\sum_{i \in U} \frac{{x_i}^2}{2p_i}} \sum_{i \in S} \frac{x_i}{\pi_i} \left[(R_i - p_i)(\frac{1}{p_i} - 1) \right] \\
&= \sum_{i \in S} \frac{1}{\pi_i} (R_i - p_i) \left\{ -2\boldsymbol{B}_i + 2x_i(1 - \frac{1}{p_i}) \frac{\sum_{k \in U} x_k \boldsymbol{B}_k / p_k}{\sum_{k \in U} {x_k}^2 / p_k} \right\}.
\end{aligned}
$$

Lemma 2. *Under the assumption D.1-D.2 and R.1-R.4, we have:*

$$
\mathrm{Var}(\boldsymbol{S}_L(\boldsymbol{\nu}^*)) = \sum_{i \in U} \frac{p_i(1 - p_i)}{\pi_i} \boldsymbol{W}_i {\boldsymbol{W}_i}^T,
$$

where $\boldsymbol{W}_i = 2x_i(1 - \frac{1}{p_i}) \frac{\sum_{k \in U} x_k \boldsymbol{B}_k / p_k}{\sum_{k \in U} x_k^2 / p_k} - 2\boldsymbol{B}_i$.

Proof of Lemma 2. $\mathrm{Var}(\boldsymbol{S}_L(\boldsymbol{\nu}^*))$ can be expressed as:

$$\begin{aligned}\mathrm{Var}(\boldsymbol{S}_L(\boldsymbol{\nu}^*)) &= \mathrm{Cov}\left(\boldsymbol{S}_L(\boldsymbol{\nu}^*), \boldsymbol{S}_L(\boldsymbol{\nu}^*)\right) \\ &= \mathrm{E}\left\{\mathrm{Cov}\left(\boldsymbol{S}_L(\boldsymbol{\nu}^*), \boldsymbol{S}_L(\boldsymbol{\nu}^*)|S\right)\right\} + \mathrm{Cov}\left\{\mathrm{E}(\boldsymbol{S}_L(\boldsymbol{\nu}^*)|S), \mathrm{E}(\boldsymbol{S}_L(\boldsymbol{\nu}^*)|S)\right\},\end{aligned}$$

where S denotes the selected sample. Since $\mathrm{E}(\boldsymbol{S}_L(\boldsymbol{\nu}^*)|S) = \mathrm{E}(\sum_{i\in S}\frac{(R_i-p_i)}{\pi_i}\boldsymbol{W}_i)|S) = 0$, so the second term of above expression is zero. Thus;

$$\begin{aligned}\mathrm{Var}(\boldsymbol{S}_L(\boldsymbol{\nu}^*)) &= \mathrm{E}\left[\mathrm{Cov}\left(\sum_{i\in S}\frac{(R_i-p_i)}{\pi_i}\boldsymbol{W}_i, \sum_{i\in S}\frac{(R_i-p_i)}{\pi_i}\boldsymbol{W}_i\middle| S\right)\right] \\ &= \mathrm{E}\left[\sum_{i\in S}\mathrm{Cov}\left(\frac{(R_i-p_i)}{\pi_i}\boldsymbol{W}_i, \frac{(R_i-p_i)}{\pi_i}\boldsymbol{W}_i\right)\right] \\ &= \mathrm{E}\left[\sum_{i\in S}\frac{\boldsymbol{W}_i}{\pi_i}p_i(1-p_i)\frac{\boldsymbol{W}_i{}^T}{\pi_i}\right] \\ &= \mathrm{E}\left[\sum_{i\in S}\frac{p_i(1-p_i)}{\pi_i{}^2}\boldsymbol{W}_i\boldsymbol{W}_i{}^T\right] \\ &= \sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\boldsymbol{W}_i\mathbf{W}_i{}^T.\end{aligned}$$

$\square$

2.4 Main Results for Horvitz-Thompson Type Estimator

In this section, we list several properties of the proposed Horvitz-Thompson type estimator.

Theorem 2. *Under the assumption D.1-D.2 and R.1-R.3, the Horvitz-Thompson type non-response weighting adjustment estimator* $\bar{y}_e = \frac{1}{N}\sum_{i\in S}\frac{y_iR_i}{\pi_i\hat{p}_i}$ *is root-*n *consistent for* $\bar{y}_U$ *with respect to both the sampling mechanism and the response mechanism. That is,*

$$\bar{y}_e = \bar{y}_U + O_p(n^{-\frac{1}{2}}).$$

Proof of Theorem 2. By definition in (2.1) and (2.2), we can express $\bar{y}_e - \bar{y}_U$ as:

$$\begin{aligned}
&\bar{y}_e - \bar{y}_U \\
=&(\bar{y}_e - \bar{y}_{e^*}) + (\bar{y}_{e^*} - \bar{y}_\pi) + (\bar{y}_\pi - \bar{y}_U) \\
=&\left(\frac{1}{N}\sum_{i\in S}\frac{y_i R_i}{\pi_i \hat{p}_i} - \frac{1}{N}\sum_{i\in S}\frac{y_i R_i}{\pi_i p_i}\right) + \left(\frac{1}{N}\sum_{i\in S}\frac{y_i R_i}{\pi_i p_i} - \frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i}\right) + \left(\frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i} - \frac{1}{N}\sum_{i\in U} y_i\right) \\
=&\frac{1}{N}\sum_{i\in S}\frac{y_i R_i}{\pi_i}\left(\frac{1}{\hat{p}_i} - \frac{1}{p_i}\right) + \frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i p_i}(R_i - p_i) + \left(\frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i} - \frac{1}{N}\sum_{i\in U} y_i\right) \\
=&A + B + C.
\end{aligned}$$

Apply Taylor expansion to $\left(\frac{1}{\hat{p}_i} - \frac{1}{p_i}\right)$, we have:

$$\begin{aligned}
\left(\frac{1}{\hat{p}_i} - \frac{1}{p_i}\right) &= \left.\frac{\partial {p_i}^{-1}}{\partial \boldsymbol{\nu}}\right|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) + 0.5(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)^T\left(\left.\frac{\partial^2 {p_i}^{-1}}{\partial \boldsymbol{\nu}\partial \boldsymbol{\nu}^T}\right|_{\boldsymbol{\nu}=\ddot{\boldsymbol{\nu}}}\right)(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) \\
&= \left(-\frac{1}{p_i^2}\boldsymbol{B}_i\right)(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) + 0.5(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)^T\left(\frac{2}{p_i^3(\ddot{\boldsymbol{\nu}})}\boldsymbol{B}_i {\boldsymbol{B}_i}^T\right)(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*),
\end{aligned}$$

where $\ddot{\boldsymbol{\nu}}$ is on the line segment joining $\hat{\boldsymbol{\nu}}$ and $\boldsymbol{\nu}^*$. Thus, we have:

$$\begin{aligned}
A =& \left[\frac{1}{N}\sum_{i\in S}\frac{y_i R_i}{\pi_i}\left(-\frac{1}{p_i^2}\boldsymbol{B}_i\right)\right](\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) \\
&+ 0.5(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)^T\left[\frac{1}{N}\sum_{i\in S}\frac{y_i R_i}{\pi_i}\left(\frac{2}{p_i^3(\ddot{\boldsymbol{\nu}})}\boldsymbol{B}_i {\boldsymbol{B}_i}^T\right)\right](\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) \\
&= O_p(1)O_p(n^{-\frac{1}{2}}) + O_p(n^{-\frac{1}{2}})O_p(1)O_p(n^{-\frac{1}{2}}) \\
&= O_p(n^{-\frac{1}{2}}).
\end{aligned}$$

For B, it's easy to see that $\mathrm{E}(\frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i p_i}(R_i - p_i)) = 0$, its variance is given by:

$$
\begin{aligned}
&\mathrm{Var}\left(\frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i p_i}(R_i - p_i)\right)\\
=&\mathrm{E}\left[\mathrm{Var}\left(\frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i p_i}(R_i - p_i)\middle| S\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(\frac{1}{N}\sum_{i\in S}\frac{y_i}{\pi_i p_i}(R_i - p_i)\middle| S\right)\right]\\
=&\mathrm{E}\left[\frac{1}{N^2}\sum_{i\in S}\mathrm{Var}\left(\frac{y_i R_i}{\pi_i p_i}\middle| S\right)\right] + 0\\
=&\frac{1}{N^2}\mathrm{E}\left(\sum_{i\in S}\frac{p_i(1-p_i)y_i^2}{\pi_i^2 p_i^2}\right)\\
=&\frac{1}{N^2}\sum_{i\in U}\frac{p_i(1-p_i)y_i^2}{\pi_i p_i^2}\\
=&O\left(\frac{1}{N}\right).
\end{aligned}
$$

So by Corollary 5.1.1.1 in Fuller (1996), $B = O_p(N^{-\frac{1}{2}})$. By (2.9), $C = O_p(n^{-\frac{1}{2}})$, so overall we have:

$$\bar{y}_e - \bar{y}_U = O_p(n^{-\frac{1}{2}}) + O_p(N^{-\frac{1}{2}}) + O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}).$$

$\square$

In order to perform inference for $\bar{y}_e$, we derive an expression for the variance of its linearized approximation. By the definition of $\bar{y}_{e^*}$ in (2.2), we can write $\bar{y}_e - \bar{y}_{e^*}$ as:

$$
\begin{aligned}
\bar{y}_e - \bar{y}_{e^*} &= \left[\frac{1}{N}\sum_{i\in S}\frac{y_i R_i}{\pi_i}\left(-\frac{1}{p_i^2}\boldsymbol{B}_i\right)\right](\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) + O_p(n^{-1})\\
&= \left[-\frac{1}{N}\sum_{i\in S}\frac{y_i(R_i - p_i + p_i)\boldsymbol{B}_i}{\pi_i p_i^2}\right](\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) + O_p(n^{-1})\\
&= \left[-\frac{1}{N}\sum_{i\in S}\frac{\boldsymbol{B}_i y_i}{\pi_i p_i} - \frac{1}{N}\sum_{i\in S}\frac{R_i - p_i}{\pi_i p_i^2}\boldsymbol{B}_i y_i\right](\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) + O_p(n^{-1}).
\end{aligned}
$$

Since $\boldsymbol{B}_i$ satisfies the same population moments as x_i itself, by the analogous argument for showing (2.9), we have $\frac{1}{N}\sum_{i\in S}\frac{\boldsymbol{B}_i y_i}{\pi_i p_i} = \frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i y_i}{p_i} + O_p(n^{-\frac{1}{2}})$. Also, using the same procedure for proving $B = O_p(N^{-\frac{1}{2}})$, it's easy to show that $\frac{1}{N}\sum_{i\in S}\frac{R_i - p_i}{\pi_i p_i^2}\boldsymbol{B}_i y_i = O_p(N^{-\frac{1}{2}})$. Hence:

$$
\begin{aligned}
\bar{y}_e - \bar{y}_{e^*} &= \left[-\frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i y_i}{p_i} + O_p(n^{-\frac{1}{2}}) + O_p(N^{-\frac{1}{2}})\right](\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) + O_p(n^{-1}) \\
&= \left[-\frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i y_i}{p_i}\right](\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*) + O_p(n^{-1}) \\
&= \left[-\frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i y_i}{p_i}\right]\{[\boldsymbol{I}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}})\} + O_p(n^{-1}) \\
&= \left[-\frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i y_i}{p_i}\right][\boldsymbol{I}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) + O_p(n^{-1}) \\
&= \left[-\frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i y_i}{p_i}\right][\boldsymbol{I}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

Direct and careful computation of $\frac{\partial \boldsymbol{S}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^T}$ yields:

$$
\begin{aligned}
\frac{\partial \boldsymbol{S}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^T} = &\sum_{i\in S}\frac{2}{\pi_i}\boldsymbol{B}_i{\boldsymbol{B}_i}^T + \sum_{i\in S}\frac{x_i R_i \boldsymbol{B}_i}{\pi_i[-(\boldsymbol{B}_i\boldsymbol{\nu})^2]}\times \\
&\left\{\frac{\sum_{i\in S}\frac{x_i}{\pi_i}[(1-\frac{R_i}{(\boldsymbol{B}_i\boldsymbol{\nu})^2}){\boldsymbol{B}_i}^T]}{\sum_{i\in S}\frac{{x_i}^2 R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}} + \frac{\sum_{i\in S}\frac{x_i}{\pi_i}[\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}} - R_i - 1 + \boldsymbol{B}_i\boldsymbol{\nu}]\times\sum_{i\in S}\frac{{x_i}^2 R_i {\boldsymbol{B}_i}^T}{\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^3}}{[\sum_{i\in S}\frac{{x_i}^2 R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}]^2}\right\} \\
&+ \frac{\sum_{i\in S}\frac{x_i}{\pi_i}[\frac{R_i}{\boldsymbol{B}_i\boldsymbol{\nu}} - R_i - 1 + \boldsymbol{B}_i\boldsymbol{\nu}]}{\sum_{i\in S}\frac{{x_i}^2 R_i}{2\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^2}}\times\sum_{i\in S}\frac{2x_i R_i}{\pi_i(\boldsymbol{B}_i\boldsymbol{\nu})^3}\boldsymbol{B}_i{\boldsymbol{B}_i}^T.
\end{aligned}
$$

After plugging the $\boldsymbol{\nu}^*$, we linearize the $\frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*) = -\frac{1}{N}E(\frac{\partial \boldsymbol{S}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^T}|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*})$, resulting:

$$
\frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*) = -\frac{1}{N}\sum_{i\in U}2\boldsymbol{B}_i{\boldsymbol{B}_i}^T - \frac{1}{N}\sum_{i\in U}\frac{x_i\boldsymbol{B}_i}{p_i}\times\frac{\sum_{i\in U}x_i(\frac{1}{p_i}-1){\boldsymbol{B}_i}^T}{\sum_{i\in U}\frac{{x_i}^2}{2p_i}} + O\left(\frac{1}{N}\right).
$$

Let $\boldsymbol{J}(\boldsymbol{\nu}^*) = (-\sum_{i\in U}2\boldsymbol{B}_i{\boldsymbol{B}_i}^T - \sum_{i\in U}\frac{x_i\boldsymbol{B}_i}{p_i}\times\frac{\sum_{i\in U}x_i(\frac{1}{p_i}-1){\boldsymbol{B}_i}^T}{\sum_{i\in U}\frac{{x_i}^2}{2p_i}})$ and apply Taylor expansion to $\frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*)$, we have:

$$
\left\{\frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*)\right\}^{-1} = \left\{\frac{1}{N}\boldsymbol{J}(\boldsymbol{\nu}^*)\right\}^{-1} + O\left(\frac{1}{N}\right).
$$

Denote $\boldsymbol{A} = -\frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i y_i}{p_i}$, then we can rewrite $\bar{y}_e - \bar{y}_{e^*}$ as:

$$\begin{aligned}
\bar{y}_e - \bar{y}_{e^*} &= \boldsymbol{A}^T\left\{\frac{1}{N}\boldsymbol{I}(\boldsymbol{\nu}^*)\right\}^{-1}\frac{1}{N}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) \\
&= \boldsymbol{A}^T\left\{\left\{\frac{1}{N}\boldsymbol{J}(\boldsymbol{\nu}^*)\right\}^{-1} + O\left(\frac{1}{N}\right)\right\}\frac{1}{N}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) \\
&= \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}).
\end{aligned}$$

Thus, the $\bar{y}_e$ can be expresses as:

$$\bar{y}_e = \bar{y}_{e^*} + \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) = \bar{y}_{el} + o_p(n^{-\frac{1}{2}}), \tag{2.13}$$

where $\bar{y}_{el}$ is the linearized $\bar{y}_e$, and the variance of the proposed estimator can be approximated by variance of $\bar{y}_{el}$.

Lemma 3. *Under assumptions D.1, D.2 and R.1-R.4, the variance of* $\mathrm{Var}(\bar{y}_{el})$ *is given by:*

$$\begin{aligned}
\mathrm{Var}(\bar{y}_{el}) = &\frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i} - 1\right)y_i^2 + \mathrm{Var}(\bar{y}_\pi) \\
&+ \sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}C_i^2 \\
&+ \frac{2}{N}\sum_{i\in U}\frac{C_i(1-p_i)y_i}{\pi_i},
\end{aligned}$$

where $C_i = \boldsymbol{A}^T\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*)\boldsymbol{W}_i$*.*

Proof of Lemma 3. Direct computation of the variance of $\bar{y}_{el}$ yields:

$$\begin{aligned}
\mathrm{Var}(\bar{y}_{el}) &= \mathrm{Var}(\bar{y}_{e^*} + \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)) \\
&= \mathrm{Var}(\bar{y}_{e^*}) + \mathrm{Var}(\boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)) + 2\mathrm{Cov}(\bar{y}_{e^*}, \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)).
\end{aligned}$$

We will compute each term in above expression.

$$\begin{aligned}
\mathrm{Var}(\bar{y}_{e^*}) &= \mathrm{Var}\left(\frac{1}{N}\sum_{i\in S}\frac{R_i y_i}{\pi_i p_i}\right)\\
&= \frac{1}{N^2}\left\{\mathrm{E}\left[\mathrm{Var}\left(\sum_{i\in S}\frac{R_i y_i}{\pi_i p_i}\middle| S\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(\sum_{i\in S}\frac{R_i y_i}{\pi_i p_i}\middle| S\right)\right]\right\}\\
&= \frac{1}{N^2}\left\{\mathrm{E}\left[\sum_{i\in S}\frac{p_i(1-p_i)y_i^2}{\pi_i^2 p_i^2}\right] + \mathrm{Var}\left[\sum_{i\in S}\frac{y_i}{\pi_i}\right]\right\}\\
&= \frac{1}{N^2}\left\{\sum_{i\in U}\frac{(1-p_i)y_i^2}{\pi_i p_i} + \mathrm{Var}\left[\sum_{i\in S}\frac{y_i}{\pi_i}\right]\right\}\\
&= \frac{1}{N^2}\sum_{i\in U}\frac{(1-p_i)y_i^2}{\pi_i p_i} + \mathrm{Var}(\bar{y}_\pi).
\end{aligned}$$

Denote $C_i = \boldsymbol{A}^T\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*)\boldsymbol{W}_i$, by Lemma 2, we have:

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)) &= \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\mathrm{Var}(\boldsymbol{S}_L(\boldsymbol{\nu}^*))[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}\\
&= \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\boldsymbol{W}_i\boldsymbol{W}_i^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}\\
&= \sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}C_i^2.
\end{aligned}$$

Now we compute the covariance between $\bar{y}_e$ and $\boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)$ as follows:

$$\begin{aligned}
\mathrm{Cov}(\bar{y}_{e^*}, \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)) &= \mathrm{Cov}\left(\frac{1}{N}\sum_{i\in S}\frac{R_i y_i}{\pi_i p_i}, \sum_{i\in S}\frac{1}{\pi_i}(R_i-p_i)C_i\right)\\
&= \frac{1}{N}\mathrm{E}\left[\mathrm{Cov}\left(\sum_{i\in S}\frac{R_i y_i}{\pi_i p_i}, \sum_{i\in S}\frac{1}{\pi_i}(R_i-p_i)C_i\middle| S\right)\right]\\
&\quad + \frac{1}{N}\mathrm{Cov}\left[\mathrm{E}\left(\sum_{i\in S}\frac{R_i y_i}{\pi_i p_i}\middle| S\right), \mathrm{E}\left(\sum_{i\in S}\frac{1}{\pi_i}(R_i-p_i)C_i\middle| S\right)\right]\\
&= \frac{1}{N}\mathrm{E}\left[\sum_{i\in S}\frac{p_i(1-p_i)C_i y_i}{\pi_i^2 p_i}\right] + 0\\
&= \frac{1}{N}\sum_{i\in U}\frac{(1-p_i)C_i y_i}{\pi_i}.
\end{aligned}$$

Combining the above results, we get the desired variance. $\square$

By plugging in the $\hat{\boldsymbol{\nu}}$ to the response model, we can obtain the estimated response probability $\hat{p}_i$ for each individual i in the sample. Then, we are able to define the variance estimator $\hat{\mathrm{V}}(\bar{y}_{el})$ appropriately. The following theorem states that the defined variance estimator is consistent for $\mathrm{Var}(\bar{y}_{el})$ under both the sampling mechanism and the response mechanism. The proof is in the appendix.

Theorem 3. *Under assumption D.1, D.2 and R.1-R.4, the variance estimator $\hat{\mathrm{V}}(\bar{y}_{el})$ is consistent for $\mathrm{Var}(\bar{y}_{el})$. That is:*

$$n(\hat{\mathrm{V}}(\bar{y}_{el}) - \mathrm{Var}(\bar{y}_{el})) = o_p(1),$$

where $\mathrm{Var}(\bar{y}_{el})$ is given in Lemma 3 and is expressed as:

$$\begin{aligned}
\mathrm{Var}(\bar{y}_{el}) &= \frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)y_i^2 + \mathrm{Var}(\bar{y}_\pi) \\
&\quad + \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\boldsymbol{W}_i\boldsymbol{W}_i{}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A} \\
&\quad + \frac{2}{N}\boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{(1-p_i)y_i}{\pi_i}\boldsymbol{W}_i \\
&= V_1 + V_2 + V_3 + V_4
\end{aligned}$$

and $\hat{\mathrm{V}}(\bar{y}_{el}) = \hat{V}_1 + \hat{V}_2 + \hat{V}_3 + \hat{V}_4$, where:

$$\hat{V}_1 = \frac{1}{N^2}\sum_{i\in S}\frac{1}{\pi_i^2}\left(\frac{1}{\hat{p}_i}-1\right)y_i^2\frac{R_i}{\hat{p}_i}, \quad \hat{V}_2 = \frac{1}{N^2}\sum_{i\in S}\frac{(1-\pi_i)y_i^2}{\pi_i^2}\frac{R_i}{\hat{p}_i} + \frac{1}{N^2}\sum_{i\neq j\, i,j\in S}\frac{\Delta_{ij}y_iy_j}{\pi_{ij}\pi_i\pi_j}\frac{R_iR_j}{\hat{p}_i\hat{p}_j},$$

$$\hat{V}_3 = \hat{\boldsymbol{A}}^T[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\sum_{i\in S}\frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\hat{\boldsymbol{W}}_i\hat{\boldsymbol{W}}_i^T[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{A}},$$

$$\hat{V}_4 = \frac{2}{N}\hat{\boldsymbol{A}}^T[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\sum_{i\in S}\frac{(1-\hat{p}_i)y_i}{\pi_i^2}\frac{R_i}{\hat{p}_i}\hat{\boldsymbol{W}}_i \quad \textit{and} \quad \hat{\boldsymbol{A}} = -\frac{1}{N}\sum_{i\in S}\frac{\boldsymbol{B}_iy_iR_i}{\pi_i\hat{p}_i^2},$$

$$\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}}) = -\sum_{i\in S}\frac{2}{\pi_i}\boldsymbol{B}_i\boldsymbol{B}_i{}^T - \sum_{i\in S}\frac{x_i\boldsymbol{B}_iR_i}{\pi_i\hat{p}_i^2}\times\frac{\sum_{i\in S}\frac{x_i}{\pi_i}(\frac{1}{\hat{p}_i}-1)\frac{R_i}{\hat{p}_i}\boldsymbol{B}_i{}^T}{\sum_{i\in S}\frac{x_i^2R_i}{2\pi_i\hat{p}_i^2}},$$

$$\hat{\boldsymbol{W}}_i = 2x_i \left(1 - \frac{1}{\hat{p}_i}\right) \frac{\sum_{k \in S} \frac{x_k \boldsymbol{B}_k R_k}{\pi_k \hat{p}_k^2}}{\sum_{k \in S} \frac{x_k^2 R_k}{\pi_k \hat{p}_k^2}} - 2\boldsymbol{B}_i = 2x_i \left(1 - \frac{1}{\hat{p}_i}\right) \hat{\boldsymbol{D}} - 2\boldsymbol{B}_i,$$

where $\boldsymbol{D} = \frac{\sum_{k \in U} \frac{x_k \boldsymbol{B}_k}{p_k}}{\sum_{k \in U} \frac{x_k^2}{p_k}}$ *is estimated by* $\hat{\boldsymbol{D}} = \frac{\sum_{k \in S} \frac{x_k \boldsymbol{B}_k R_k}{\pi_k \hat{p}_k^2}}{\sum_{k \in S} \frac{x_k^2 R_k}{\pi_k \hat{p}_k^2}}$.

The following theorem gives the asymptotic normality of the proposed Horvitz-Thompson type estimator. The proof is in the appendix.

Theorem 4. *Under the condition D.1-D.3 and R.1-R.4, we obtain the following asymptotic normality, jointly with respect to the sampling design and the response mechanism,*

$$\frac{\bar{y}_e - \bar{y}_U}{\sqrt{\hat{V}(\bar{y}_{el})}} \xrightarrow{d} N(0, 1),$$

where $\bar{y}_e = \frac{1}{N} \sum_{i \in S} \frac{y_i R_i}{\pi_i \hat{p}_i}$ *and* $\hat{V}(\bar{y}_{el})$ *is defined in Theorem 3.*

2.5 Main Results for Hájek Type Estimator

In this section, we list several properties of the proposed Hájek type estimator.

Theorem 5. *Under the assumption D.1-D.2 and R.1-R.3, the Hájek type nonresponse weighting adjustment estimator* $\bar{y}_H = \frac{\sum_{i \in S} \frac{y_i R_i}{\pi_i \hat{p}_i}}{\sum_{i \in S} \frac{R_i}{\pi_i \hat{p}_i}}$ *is root-n consistent for $\bar{y}_U$ with respect to both sampling mechanism and response mechanism. That is,*

$$\bar{y}_H = \bar{y}_U + O_p(n^{-\frac{1}{2}}).$$

Proof of Theorem 5. Apply Taylor expansion to $\bar{y}_H$ around $\frac{\sum_{i\in U} y_i}{N}$, we have:

$$\begin{aligned}
\bar{y}_H - \bar{y}_U &= \frac{1}{N}\left(\sum_{i\in S}\frac{y_i R_i}{\pi_i \hat{p}_i} - \sum_{i\in U} y_i\right) - \frac{\sum_{i\in U} y_i}{N^2}\left(\sum_{i\in S}\frac{R_i}{\pi_i \hat{p}_i} - N\right) + O_p(n^{-1}) \\
&= \frac{1}{N}\left(\sum_{i\in S}\frac{y_i R_i}{\pi_i \hat{p}_i} - \bar{y}_U \sum_{i\in S}\frac{R_i}{\pi_i \hat{p}_i}\right) + O_p(n^{-1}) \\
&= (\bar{y}_e - \bar{y}_U) - \bar{y}_U\left(\frac{1}{N}\sum_{i\in S}\frac{R_i}{\pi_i \hat{p}_i} - 1\right) + O_p(n^{-1}).
\end{aligned}$$

By Theorem 2, $\bar{y}_e - \bar{y}_U = O_p(n^{-\frac{1}{2}})$, and by a similar proof of Theorem 2, we have:

$$\frac{1}{N}\sum_{i\in S}\frac{R_i}{\pi_i \hat{p}_i} - 1 = O_p(n^{-\frac{1}{2}}).$$

Hence, we have:

$$\begin{aligned}
\bar{y}_H - \bar{y}_U &= (\bar{y}_e - \bar{y}_U) - \bar{y}_U\left(\frac{1}{N}\sum_{i\in S}\frac{R_i}{\pi_i \hat{p}_i} - 1\right) + O_p(n^{-1}) \\
&= O_p(n^{-\frac{1}{2}}) + O(1)O_p(n^{-\frac{1}{2}}) + O_p(n^{-1}) \\
&= O_p(n^{-\frac{1}{2}}).
\end{aligned}$$

□

Note that we can write $\bar{y}_H = \bar{y}_U + \frac{1}{N}\sum_{i\in S}\frac{(y_i - \bar{y}_U)R_i}{\pi_i \hat{p}_i} + O_p(n^{-1})$. In order to do inference for $\bar{y}_H$, we need to obtain the linearized approximation of the term $\frac{1}{N}\sum_{i\in S}\frac{(y_i - \bar{y}_U)R_i}{\pi_i \hat{p}_i}$ first. We have the following:

$$
\begin{aligned}
&\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i\hat{p}_i}-\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i}\\
&=\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i}\left(\frac{1}{\hat{p}_i}-\frac{1}{p_i}\right)\\
&=\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i}\left[\left(-\frac{1}{p_i^2}\boldsymbol{B}_i\right)(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}^*)+0.5(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}^*)^T\left(\frac{2}{p_i^3(\ddot{\boldsymbol{\nu}})}\boldsymbol{B}_i\boldsymbol{B}_i^T\right)(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}^*)\right]\\
&=\left[-\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i\boldsymbol{B}_i}{\pi_i p_i^2}\right](\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}^*)+O_p(n^{-1})\\
&=\left[-\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{\pi_i p_i}-\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)(R_i-p_i)\boldsymbol{B}_i}{\pi_i p_i^2}\right](\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}^*)+O_p(n^{-1}).
\end{aligned}
$$

Since $\boldsymbol{B}_i$ satisfies the same population moments as x_i itself, by a similar argument for showing (2.9), we have that $\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{\pi_i p_i}=\frac{1}{N}\sum_{i\in U}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{p_i}+O_p(n^{-\frac{1}{2}})$. Also, by a similar argument of showing $B=O_p(N^{-\frac{1}{2}})$ in Theorem 2, we have $\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)(R_i-p_i)\boldsymbol{B}_i}{\pi_i p_i^2}=O_p(N^{-\frac{1}{2}})$, thus:

$$
\begin{aligned}
&\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i\hat{p}_i}-\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i}\\
&=\left[-\frac{1}{N}\sum_{i\in U}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{p_i}+O_p(n^{-\frac{1}{2}})+O_p(N^{-\frac{1}{2}})\right](\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}^*)+O_p(n^{-1})\\
&=\left[-\frac{1}{N}\sum_{i\in U}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{p_i}\right](\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}^*)+O_p(n^{-1})\\
&=\left[-\frac{1}{N}\sum_{i\in U}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{p_i}\right]\boldsymbol{I}^{-1}(\boldsymbol{\nu}^*)\boldsymbol{S}_L(\boldsymbol{\nu}^*)+o_p(n^{-\frac{1}{2}})+O_p(n^{-1})\\
&=\left[-\frac{1}{N}\sum_{i\in U}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{p_i}\right]\left\{\left\{\frac{1}{N}\boldsymbol{J}(\boldsymbol{\nu}^*)\right\}^{-1}+O\left(\frac{1}{N}\right)\right\}\frac{1}{N}\boldsymbol{S}_L(\boldsymbol{\nu}^*)+o_p(n^{-\frac{1}{2}})\\
&=\left[-\frac{1}{N}\sum_{i\in U}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{p_i}\right][\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)+o_p(n^{-\frac{1}{2}})\\
&=\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)+o_p(n^{-\frac{1}{2}}),
\end{aligned}
$$

where $\boldsymbol{A}_H^T = -\frac{1}{N}\sum_{i\in U}\frac{(y_i-\bar{y}_U)\boldsymbol{B}_i}{p_i}$ and $\boldsymbol{J}(\boldsymbol{\nu}^*)$, $\boldsymbol{S}_L(\boldsymbol{\nu}^*)$ are as defined previously. Therefore, we can write $\bar{y}_H$ as:

$$
\begin{aligned}
\bar{y}_H &= \bar{y}_U + \frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i} + \boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) + O_p(n^{-1}) \\
&= \bar{y}_U + \frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i} + \boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) \\
&= \bar{y}_U + \bar{y}_{HL} + o_p(n^{-\frac{1}{2}}), \qquad (2.14)
\end{aligned}
$$

where $\bar{y}_{HL} = \frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i} + \boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)$, so the variance of $\bar{y}_H$ can be approximated by the variance of $\bar{y}_{HL}$.

Lemma 4. *Under assumption D.1, D.2 and R.1-R.4, the variance of $\bar{y}_{HL}$ is given by:*

$$
\begin{aligned}
\mathrm{Var}(\bar{y}_{HL}) = &\frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)(y_i-\bar{y}_U)^2 + \frac{1}{N^2}\sum_{i,j\in U}\Delta_{ij}\frac{(y_i-\bar{y}_U)(y_j-\bar{y}_U)}{\pi_i\pi_j} \\
&+ \sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}C_{Hi}^2 \\
&+ \frac{2}{N}\sum_{i\in U}\frac{C_{Hi}(1-p_i)(y_i-\bar{y}_U)}{\pi_i},
\end{aligned}
$$

where $C_{Hi} = \boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{W}_i$.

Proof of Lemma 4. Direct computation of $\mathrm{Var}(\bar{y}_{HL})$ yields:

$$
\begin{aligned}
\mathrm{Var}(\bar{y}_{HL}) &= \mathrm{Var}\left(\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i} + \boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)\right) \\
&= \mathrm{Var}\left(\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i}\right) + \mathrm{Var}(\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)) \\
&+ 2\mathrm{Cov}\left(\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i}, \boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)\right),
\end{aligned}
$$

where we have:

$$
\begin{aligned}
&\mathrm{Var}\left(\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i}\right)\\
=&\frac{1}{N^2}\left\{\mathrm{E}\left[\mathrm{Var}\left(\sum_{i\in S}\frac{R_i(y_i-\bar{y}_U)}{\pi_i p_i}\middle| S\right)\right]+\mathrm{Var}\left[\mathrm{E}\left(\sum_{i\in S}\frac{R_i(y_i-\bar{y}_U)}{\pi_i p_i}\middle| S\right)\right]\right\}\\
=&\frac{1}{N^2}\left\{\mathrm{E}\left[\sum_{i\in S}\frac{p_i(1-p_i)(y_i-\bar{y}_U)^2}{\pi_i^2 p_i^2}\right]+\mathrm{Var}\left[\sum_{i\in S}\frac{(y_i-\bar{y}_U)}{\pi_i}\right]\right\}\\
=&\frac{1}{N^2}\left\{\sum_{i\in U}\frac{(1-p_i)(y_i-\bar{y}_U)^2}{\pi_i p_i}+\sum_{i,j\in U}\Delta_{ij}\frac{(y_i-\bar{y}_U)(y_j-\bar{y}_U)}{\pi_i\pi_j}\right\}\\
=&\frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)(y_i-\bar{y}_U)^2+\frac{1}{N^2}\sum_{i,j\in U}\Delta_{ij}\frac{(y_i-\bar{y}_U)(y_j-\bar{y}_U)}{\pi_i\pi_j}.
\end{aligned}
$$

By Lemma 2, we have:

$$
\begin{aligned}
&\mathrm{Var}(\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*))=\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\mathrm{Var}(\boldsymbol{S}_L(\boldsymbol{\nu}^*))[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}_H\\
=&\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\boldsymbol{W}_i\boldsymbol{W}_i{}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}_H=\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}C_{Hi}^2,
\end{aligned}
$$

where $C_{Hi}=\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{W}_i$. Now, the covariance is given by:

$$
\begin{aligned}
&\mathrm{Cov}\left(\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i},\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*)\right)\\
=&\ \mathrm{Cov}\left(\frac{1}{N}\sum_{i\in S}\frac{R_i(y_i-\bar{y}_U)}{\pi_i p_i},\sum_{i\in S}\frac{1}{\pi_i}(R_i-p_i)C_{Hi}\right)\\
=&\ \frac{1}{N}\mathrm{E}\left[\mathrm{Cov}\left(\sum_{i\in S}\frac{R_i(y_i-\bar{y}_U)}{\pi_i p_i},\sum_{i\in S}\frac{1}{\pi_i}(R_i-p_i)C_{Hi}\middle| S\right)\right]\\
&+\frac{1}{N}\mathrm{Cov}\left[\mathrm{E}\left(\sum_{i\in S}\frac{R_i(y_i-\bar{y}_U)}{\pi_i p_i}\middle| S\right),\mathrm{E}\left(\sum_{i\in S}\frac{1}{\pi_i}(R_i-p_i)C_{Hi}\middle| S\right)\right]\\
=&\ \frac{1}{N}\mathrm{E}\left[\sum_{i\in S}\frac{p_i(1-p_i)C_{Hi}(y_i-\bar{y}_U)}{\pi_i^2 p_i}\right]+0\\
=&\ \frac{1}{N}\sum_{i\in U}\frac{(1-p_i)C_{Hi}(y_i-\bar{y}_U)}{\pi_i}.
\end{aligned}
$$

Combining the above results, we get the variance of $\bar{y}_{HL}$. $\square$

In the following, we showed that the variance estimator $\hat{V}(\bar{y}_{HL})$ is consistent for $\mathrm{Var}(\bar{y}_{HL})$. Also, we derived the asymptotic normality of the Hájek type estimator. The proofs are in the appendix.

Theorem 6. *Under assumption D.1, D.2 and R.1-R.4, the variance estimator $\hat{V}(\bar{y}_{HL})$ is consistent for $\mathrm{Var}(\bar{y}_{HL})$. That is:*

$$n(\hat{V}(\bar{y}_{HL}) - \mathrm{Var}(\bar{y}_{HL})) = o_p(1),$$

where $\mathrm{Var}(\bar{y}_{HL})$ is given in Lemma 4 and $\hat{V}(\bar{y}_{HL}) = \hat{V}_{H1} + \hat{V}_{H2} + \hat{V}_{H3} + \hat{V}_{H4}$, in which each term is defined as follows:

$$\hat{V}_{H1} = \frac{1}{\hat{N}^2} \sum_{i \in S} \frac{1}{\pi_i^2} \left(\frac{1}{\hat{p}_i} - 1 \right) (y_i - \bar{y}_H)^2 \frac{R_i}{\hat{p}_i},$$

$$\hat{V}_{H2} = \frac{1}{\hat{N}^2} \sum_{i \in S} \frac{(1-\pi_i)(y_i - \bar{y}_H)^2}{\pi_i^2} \frac{R_i}{\hat{p}_i} + \frac{1}{\hat{N}^2} \sum_{i \neq j\, i,j \in S} \frac{\Delta_{ij}(y_i - \bar{y}_H)(y_j - \bar{y}_H)}{\pi_{ij}\pi_i\pi_j} \frac{R_i R_j}{\hat{p}_i \hat{p}_j},$$

$$\hat{V}_{H3} = \hat{\boldsymbol{A}}_H^T [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1} \sum_{i \in S} \frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2} \frac{R_i}{\hat{p}_i} \hat{\boldsymbol{W}}_i \hat{\boldsymbol{W}}_i^T [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1} \hat{\boldsymbol{A}}_H^T,$$

$$\hat{V}_{H4} = \frac{2}{\hat{N}} \hat{\boldsymbol{A}}_H^T [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1} \sum_{i \in S} \frac{(1-\hat{p}_i)(y_i - \bar{y}_H)}{\pi_i^2} \frac{R_i}{\hat{p}_i} \hat{\boldsymbol{W}}_i,$$

where $\hat{N} = \sum_{i \in S} \frac{R_i}{\pi_i \hat{p}_i}$, $\hat{\boldsymbol{A}}_H = -\frac{1}{\hat{N}} \sum_{i \in S} \frac{\boldsymbol{B}_i (y_i - \bar{y}_H) R_i}{\pi_i \hat{p}_i^2}$, $\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})$ and $\hat{\boldsymbol{W}}_i$ are defined in Theorem 3.

Theorem 7. *Under the condition D.1-D.3 and R.1-R.4, we obtain the following asymptotic normality for the Hájek type estimator, jointly with respect to the sampling design and the response mechanism,*

$$\frac{\bar{y}_H - \bar{y}_U}{\sqrt{\hat{V}(\bar{y}_{HL})}} \xrightarrow{d} N(0,1),$$

where $\bar{y}_H = \frac{\sum_{i \in S} \frac{y_i R_i}{\pi_i \hat{p}_i}}{\sum_{i \in S} \frac{R_i}{\pi_i \hat{p}_i}}$ and $\hat{V}(\bar{y}_{HL})$ is defined in Theorem 6.

2.6 Simulation Study

A simulation study was conducted to evaluate the finite-sample performance of the proposed estimators. Here, we set the population size $N = 10000$. The response model is generated by linear model $p_i = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i}$, where x_1 and x_2 are independently and identically distributed uniform (0,1) variables. Letting $\gamma_0 = 0.1$ and setting the values of γ_1 and γ_2 differently, we considered the following 7 response models:

$(1)\, p_i = 0.1 + 0x_{1i} + 0.9x_{2i};$ $(2)\, p_i = 0.1 + 0.1x_{1i} + 0.8x_{2i};$

$(3)\, p_i = 0.1 + 0.25x_{1i} + 0.65x_{2i};$ $(4)\, p_i = 0.1 + 0.45x_{1i} + 0.45x_{2i};$

$(5)\, p_i = 0.1 + 0.65x_{1i} + 0.25x_{2i};$ $(6)\, p_i = 0.1 + 0.8x_{1i} + 0.1x_{2i};$

$(7)\, p_i = 0.1 + 0.9x_{1i} + 0x_{2i}.$

First note that by setting the parameter value appropriately, we make the $p_i \in (0.1, 1), \forall i$. Also, from (1) to (7), we make the contribution of covariate x_1 to the response model increase and thus the contribution of x_2 to the response model decreases accordingly. In doing so, we want to see whether the importance of the calibrated variable towards the response model affect the performance of the proposed estimator.

In terms of the outcome model, we consider the simple linear models. Since we want to calibrate on different covariates in the response model, so we set the outcome model as follows:

$(1) \quad y_i = 4x_{1i} + \epsilon_i;$ $(2) \quad y_i = 10x_{1i} + \epsilon_i;$

$(3) \quad y_i = 4x_{2i} + \epsilon_i;$ $(4) \quad y_i = 10x_{2i} + \epsilon_i;$

where ϵ_i are generated from $N(0, 1)$ independently. Here, study variable is thought to be weakly correlated with $x_i (i = 1, 2)$ when coefficient is 4 and strongly correlated with $x_i (i = 1, 2)$ when coefficient equals 10. Intuitively, strong correlation between calibrated variable and study variable implies good effect of the calibration equation, which in turn indicate the good performance of the proposed estimator.

From each of the realized finite populations, a simple random sample of size $n = 400$ is generated without replacement. The response indicator variable R_i are generated from the Bernoulli distribution with probability $p_i = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i}$. The finite populations of $(y_i; x_{1i}; x_{2i})$ are fixed and R_i is random in the Monte Carlo sampling. The study variable y_i is observed if and only if $R_i = 1$. The auxiliary variables x_1 and x_2 are observed throughout the sample. The Monte Carlo sample sizes are all set to be $B = 10,000$. In terms of the algorithm, since we are faced with a nonlinear constrained optimization problem, we use the algorithms from “NLopt" package to get the estimates of the parameters in the response model. In detail, we used the global optimization algorithm “ISRES" to find the global optimum. Then, setting the global optimum as a starting point, we applied the local optimization algorithm “LBFGS" to "polish" the optimum to a greater accuracy. The details for algorithms “ISRES" and “LBFGS" can be found in Runarsson and Yao (2000) and Liu and Nocedal (1989), respectively.

Using the Monte Carlo samples generated above, we computed: (1) Relative bias of $\bar{y}_{\tilde{e}}$, $\bar{y}_e$, $\bar{y}_{\tilde{H}}$, $\bar{y}_H$ and $\bar{y}_{\hat{H}}$. The estimators $\bar{y}_e$ and $\bar{y}_H$ are defined in (2.7) and (2.8), respectively. In terms of the estimator $\bar{y}_{\tilde{e}}$, the response probability is estimated only by least square criterion without calibration. The estimator $\bar{y}_{\tilde{H}}$ is the “Hájek type" of $\bar{y}_{\tilde{e}}$. We define the estimator $\bar{y}_{\hat{H}} = \frac{\sum_{i\in S} \frac{y_i R_i}{\pi_i \hat{p}_H}}{\sum_{i\in S} \frac{R_i}{\pi_i \hat{p}_H}}$, where $\hat{p}_H = \boldsymbol{B}_i \hat{\boldsymbol{\nu}}_H$ and $\hat{\boldsymbol{\nu}}_H$ is estimated by the least square criterion subject to $\frac{\sum_{i\in S} \frac{x_i}{\pi_i}}{\sum_{i\in S} \frac{1}{\pi_i}} = \frac{\sum_{i\in S} \frac{x_i}{\pi_i} \frac{R_i}{\boldsymbol{B}_i \boldsymbol{\nu}}}{\sum_{i\in S} \frac{1}{\pi_i} \frac{R_i}{\boldsymbol{B}_i \boldsymbol{\nu}}}$. In practice, we prefer this “Hájek type" calibration equation since it may be more efficient in real applications; (2) The ratio of $\frac{MSE(\bar{y}_e)}{MSE(\bar{y}_{\tilde{e}})}$, $\frac{MSE(\bar{y}_H)}{MSE(\bar{y}_{\tilde{H}})}$ and $\frac{MSE(\bar{y}_{\hat{H}})}{MSE(\bar{y}_{\tilde{H}})}$; (3) Relative biases of the variance estimator for $\bar{y}_e$ and $\bar{y}_H$, compared with corresponding theoretical asymptotic variance; (4) Coverage of 95 percent confidence interval for $\bar{y}_e$, $\bar{y}_H$ and $\bar{y}_{\hat{H}}$. All the simulation results are reported in the following tables.

The tables for the 7 response models present the Monte Carlo relative biases of the 5 nonresponse weighting adjustment estimators obtained from the simulation study. The Monte Carlo

Table 2.1: Relative biases of the 5 nonresponse weighting adjustment estimator, based on 10,000 samples. For response model (1): $p_i = 0.1 + 0x_{1i} + 0.9x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_{\tilde{e}}$	-0.1808	-0.2218	0.0474	0.0063	-0.1808	-0.2218	0.0474	0.0063
$\bar{y}_e$	-0.2994	-0.3333	0.1386	0.0981	-0.1456	-0.1847	0.0596	0.0197
$\bar{y}_{\tilde{H}}$	0.1561	0.1160	0.4307	0.3905	0.1561	0.1160	0.4307	0.3905
$\bar{y}_H$	0.0793	0.0384	0.5199	0.4782	0.1559	0.1147	0.4188	0.3765
$\bar{y}_{\hat{H}}$	0.1072	0.0732	0.8183	0.7789	0.1527	0.1066	0.4999	0.4573

Table 2.2: Relative biases of the 5 nonresponse weighting adjustment estimator, based on 10,000 samples. For response model (2): $p_i = 0.1 + 0.1x_{1i} + 0.8x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_{\tilde{e}}$	0.1296	0.1044	-0.0044	-0.0297	0.1296	0.1044	-0.0044	-0.0297
$\bar{y}_e$	0.0395	0.0143	0.0399	0.0149	0.1598	0.1349	0.0076	-0.0175
$\bar{y}_{\tilde{H}}$	0.1943	0.1695	0.0790	0.0541	0.1943	0.1695	0.0790	0.0541
$\bar{y}_H$	0.1568	0.1311	0.1512	0.1275	0.1930	0.1682	0.0697	0.0457
$\bar{y}_{\hat{H}}$	0.1769	0.1530	0.3585	0.3336	0.2596	0.2358	0.1319	0.1033

Table 2.3: Relative biases of the 5 nonresponse weighting adjustment estimator, based on 10,000 samples. For response model (3): $p_i = 0.1 + 0.25x_{1i} + 0.65x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_{\tilde{e}}$	0.0868	0.0667	-0.0065	-0.0267	0.0868	0.0667	-0.0065	-0.0267
$\bar{y}_e$	0.0781	0.0581	0.0195	-0.0012	0.0934	0.0729	-0.0055	-0.0257
$\bar{y}_{\tilde{H}}$	0.1337	0.1139	0.0447	0.0249	0.1337	0.1139	0.0447	0.0249
$\bar{y}_H$	0.1302	0.1099	0.0659	0.0470	0.1343	0.1142	0.0453	0.0252
$\bar{y}_{\hat{H}}$	0.1460	0.1254	0.2443	0.2221	0.2458	0.2230	0.055	0.0341

relative bias is computed by the following:

$$\text{Percentage Relative Bias} = \frac{E(\cdot) - \bar{y}_U}{\bar{y}_U} \times 100\%.$$

Table 2.4: Relative biases of the 5 nonresponse weighting adjustment estimator, based on 10,000 samples. For response model (4) $p_i = 0.1 + 0.45x_{1i} + 0.45x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_{\tilde{e}}$	0.0745	0.0635	-0.0290	-0.0399	0.0745	0.0635	-0.0290	-0.0399
$\bar{y}_e$	0.0717	0.0619	-0.0242	-0.0334	0.0859	0.0745	-0.0251	-0.0354
$\bar{y}_{\tilde{H}}$	0.1109	0.0998	0.0077	-0.0033	0.1109	0.0998	0.0077	-0.0033
$\bar{y}_H$	0.1129	0.1020	0.0143	0.0034	0.1145	0.1045	0.0088	-0.003
$\bar{y}_{\hat{H}}$	0.1277	0.1147	0.1574	0.1385	0.2574	0.2513	0.0307	0.0170

Table 2.5: Relative biases of the 5 nonresponse weighting adjustment estimator, based on 10,000 samples. For response model (5) $p_i = 0.1 + 0.65x_{1i} + 0.25x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_{\tilde{e}}$	0.0511	0.0559	-0.0480	-0.0433	0.0511	0.0559	-0.0480	-0.0433
$\bar{y}_e$	0.0511	0.0560	-0.0427	-0.0384	0.0764	0.0830	-0.0588	-0.0526
$\bar{y}_{\tilde{H}}$	0.1113	0.1161	0.0076	0.0125	0.1113	0.1161	0.0076	0.0125
$\bar{y}_H$	0.1115	0.1169	0.0086	0.0134	0.1327	0.1358	0.0002	0.0050
$\bar{y}_{\hat{H}}$	0.1167	0.1171	0.1186	0.1206	0.310	0.3113	0.0112	0.0204

Table 2.6: Relative biases of the 5 nonresponse weighting adjustment estimator, based on 10,000 samples. For response model (6) $p_i = 0.1 + 0.8x_{1i} + 0.1x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_{\tilde{e}}$	0.0620	0.0608	-0.0241	-0.0254	0.062	0.0608	-0.0241	-0.0254
$\bar{y}_e$	0.0676	0.0655	-0.0019	-0.0026	0.1234	0.1125	-0.0906	-0.0889
$\bar{y}_{\tilde{H}}$	0.1591	0.1588	0.0554	0.0551	0.1591	0.1588	0.0554	0.0551
$\bar{y}_H$	0.1516	0.1512	0.0560	0.0541	0.2144	0.2196	0.0193	0.019
$\bar{y}_{\hat{H}}$	0.1859	0.1892	0.1237	0.1274	0.4459	0.4438	0.0411	0.0417

The results from Table 2.1 to Table 2.7 reveal that the relative biases of the five estimators are all very small with absolute values less than 1 percent, across all 7 settings of the response model.

In term of the ratio of MSEs, from Table 2.8 to Table 2.14, we found that the performance of $\bar{y}_e$ wins over $\bar{y}_{\tilde{e}}$ in some scenario, depending both on response model specification and outcome

Table 2.7: Relative biases of the 5 nonresponse weighting adjustment estimator, based on 10,000 samples. For response model (7): $p_i = 0.1 + 0.9x_{1i} + 0x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_{\tilde{e}}$	0.0882	0.0866	-0.3976	-0.3993	0.0882	0.0866	-0.3976	-0.3993
$\bar{y}_e$	0.0917	0.0904	-0.3735	-0.3728	0.2022	0.2077	-0.4615	-0.4522
$\bar{y}_{\tilde{H}}$	0.5082	0.5087	-0.0223	-0.0218	0.5082	0.5087	-0.0223	-0.0218
$\bar{y}_H$	0.4989	0.4968	-0.0246	-0.0213	0.5662	0.5666	-0.0896	-0.087
$\bar{y}_{\hat{H}}$	0.5448	0.5475	-0.0232	-0.0239	0.9057	0.8998	-0.0543	-0.0553

model. While, the performance of $\bar{y}_{\hat{H}}$, compared with $\bar{y}_{\tilde{H}}$, only depends on the outcome model, regardless of the simple linear form response model. More specifically, for the Horvitz-Thompson type estimator $\bar{y}_e$, when the calibrated variable contributes less to the linear response model and correlates more with the study variable, the performance of the $\bar{y}_e$, compared with $\bar{y}_{\tilde{e}}$, will be better. If there is a strong correlation between calibrated variable and the response probability, the performance of $\bar{y}_e$ is close (or slightly better) to the performance of $\bar{y}_{\tilde{e}}$ when study variable depends on the calibrated variable and slightly weaker than the performance of $\bar{y}_{\tilde{e}}$ when study variable does not depends on the calibrated variable. The reason may be that the less the calibrated variable contributes to the linear response model, the more the calibration equation will twist the estimate from the least square criterion, which eventually makes our proposed estimator perform better. For the Hájek type estimator $\bar{y}_{\hat{H}}$, no matter how the simple linear response model is specified, $\bar{y}_{\hat{H}}$ has much smaller MSE, compared with $\bar{y}_{\tilde{H}}$, when the study variable correlates with the calibrated variable and $\bar{y}_{\hat{H}}$ performs close to $\bar{y}_{\tilde{H}}$ when study variable does not depend on the calibrated variable.

We also computed the relative bias of the variance estimator of $\bar{y}_e$ and $\bar{y}_H$, compared with the corresponding theoretical asymptotic variance, from the Monte Carlo samples. The relative bias is calculated by the following formula:

$$\text{Relative Bias} = \frac{mean(\hat{V}(\cdot)) - V(\cdot)}{V(\cdot)} \times 100\%.$$

Table 2.8: The ratio of MSEs, based on 10,000 samples. For response model (1): $p_i = 0.1 + 0x_{1i} + 0.9x_{2i}$

	calibrate on x_1				calibrate on x_2			
	model1	model2	model3	model4	model1	model2	model3	model4
$\frac{MSE(\bar{y}_e)}{MSE(\bar{y}_{\tilde{e}})}$	0.79	0.68	1.06	1.14	1.06	1.09	0.99	0.98
$\frac{MSE(\bar{y}_H)}{MSE(\bar{y}_{\tilde{H}})}$	0.96	0.92	0.92	0.87	1.00	1.00	1.01	1.02
$\frac{MSE(\bar{y}_{\hat{H}})}{MSE(\bar{y}_{\tilde{H}})}$	0.87	0.76	1.01	1.04	1.00	1.00	0.85	0.72

Table 2.9: The ratio of MSEs, based on 10,000 samples. For response model (2): $p_i = 0.1 + 0.1x_{1i} + 0.8x_{2i}$

	calibrate on x_1				calibrate on x_2			
	model1	model2	model3	model4	model1	model2	model3	model4
$\frac{MSE(\bar{y}_e)}{MSE(\bar{y}_{\tilde{e}})}$	0.87	0.76	1.04	1.09	1.06	1.11	0.99	0.97
$\frac{MSE(\bar{y}_H)}{MSE(\bar{y}_{\tilde{H}})}$	0.97	0.96	0.95	0.92	1.00	1.00	1.01	1.01
$\frac{MSE(\bar{y}_{\hat{H}})}{MSE(\bar{y}_{\tilde{H}})}$	0.89	0.80	0.99	1.00	1.00	1.00	0.87	0.76

Table 2.10: The ratio of MSEs, based on 10,000 samples. For response model (3): $p_i = 0.1 + 0.25x_{1i} + 0.65x_{2i}$

	calibrate on x_1				calibrate on x_2			
	model1	model2	model3	model4	model1	model2	model3	model4
$\frac{MSE(\bar{y}_e)}{MSE(\bar{y}_{\tilde{e}})}$	0.96	0.92	1.02	1.04	1.03	1.06	0.99	0.98
$\frac{MSE(\bar{y}_H)}{MSE(\bar{y}_{\tilde{H}})}$	0.99	0.98	0.99	0.98	1.00	1.0	1.00	1.0
$\frac{MSE(\bar{y}_{\hat{H}})}{MSE(\bar{y}_{\tilde{H}})}$	0.91	0.83	0.97	0.96	0.98	0.97	0.90	0.82

Table 2.11: The ratio of MSEs, based on 10,000 samples. For response model (4): $p_i = 0.1 + 0.45x_{1i} + 0.45x_{2i}$

	calibrate on x_1				calibrate on x_2			
	model1	model2	model3	model4	model1	model2	model3	model4
$\frac{MSE(\bar{y}_e)}{MSE(\bar{y}_{\tilde{e}})}$	0.98	0.97	1.01	1.03	1.01	1.03	0.98	0.97
$\frac{MSE(\bar{y}_H)}{MSE(\bar{y}_{\tilde{H}})}$	0.99	0.99	1.00	1.00	1.0	1.0	0.99	0.99
$\frac{MSE(\bar{y}_{\hat{H}})}{MSE(\bar{y}_{\tilde{H}})}$	0.91	0.84	0.97	0.94	0.97	0.95	0.91	0.84

Table 2.12: The ratio of MSEs, based on 10,000 samples. For response model (5): $p_i = 0.1 + 0.65x_{1i} + 0.25x_{2i}$

	calibrate on x_1				calibrate on x_2			
	model1	model2	model3	model4	model1	model2	model3	model4
$\frac{MSE(\bar{y}_e)}{MSE(\bar{y}_{\tilde{e}})}$	0.99	0.98	1.03	1.06	1.02	1.05	0.95	0.91
$\frac{MSE(\bar{y}_H)}{MSE(\bar{y}_{\tilde{H}})}$	1.00	1.00	1.0	0.99	0.99	0.98	0.99	0.98
$\frac{MSE(\bar{y}_{\hat{H}})}{MSE(\bar{y}_{\tilde{H}})}$	0.90	0.82	0.98	0.96	0.98	0.97	0.91	0.84

Table 2.13: The ratio of MSEs, based on 10,000 samples. For response model (6): $p_i = 0.1+0.8x_{1i}+0.1x_{2i}$

	calibrate on x_1				calibrate on x_2			
	model1	model2	model3	model4	model1	model2	model3	model4
$\frac{MSE(\bar{y}_e)}{MSE(\bar{y}_{\tilde{e}})}$	0.99	0.98	1.06	1.10	1.09	1.08	0.87	0.77
$\frac{MSE(\bar{y}_H)}{MSE(\bar{y}_{\tilde{H}})}$	1.00	1.01	0.99	1.00	0.96	0.93	0.97	0.96
$\frac{MSE(\bar{y}_{\hat{H}})}{MSE(\bar{y}_{\tilde{H}})}$	0.88	0.77	1.00	0.99	1.00	1.00	0.89	0.79

Table 2.14: The ratio of MSEs, based on 10,000 samples. For response model (7): $p_i = 0.1+0.9x_{1i}+0x_{2i}$

	calibrate on x1				calibrate on x2			
	model1	model2	model3	model4	model1	model2	model3	model4
$\frac{MSE(\bar{y}_e)}{MSE(\bar{y}_{\tilde{e}})}$	0.99	0.98	1.05	1.09	1.06	1.21	0.78	0.65
$\frac{MSE(\bar{y}_H)}{MSE(\bar{y}_{\tilde{H}})}$	1.01	1.02	1.00	1.0	0.94	0.89	0.95	0.92
$\frac{MSE(\bar{y}_{\hat{H}})}{MSE(\bar{y}_{\tilde{H}})}$	0.86	0.74	1.00	1.01	1.02	1.05	0.85	0.74

Based on the results from Table 2.15 to Table 2.21, we see that the relative biases of the variance estimator for all combinations of response model and outcome model are less than 9%, with most of the relative biases less than 2%, which coincide with Theorem 3 and Theorem 6.

We also computed interval estimators for 95% nominal coverage for $\bar{y}_e$, $\bar{y}_H$ and $\bar{y}_{\hat{H}}$. The Table 2.22 to Table 2.28 displays the actual coverages of 95% confidence intervals. The confidence intervals are calculated by $(\hat{\theta} - 1.96\sqrt{\hat{V}}, \hat{\theta} + 1.96\sqrt{\hat{V}})$, where $\hat{\theta}$ a point estimate and $\hat{V}$ is its estimated variance. Note that here we use the variance estimate of $\bar{y}_H$ as a substitution for the vari-

Table 2.15: Relative biases of the variance estimator of $\bar{y}_e$ and $\bar{y}_H$, compared with its corresponding theoretical asymptotic variance, based on 10,000 samples. For response model (1): $p_i = 0.1 + 0x_{1i} + 0.9x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\hat{V}(\bar{y}_e)$	-1.49	0.88	-2.79	-1.70	2.63	4.51	-0.56	0.14
$\hat{V}(\bar{y}_H)$	-1.16	0.59	-1.76	-0.27	0.43	1.69	0.12	1.64

Table 2.16: Relative biases of the variance estimator of $\bar{y}_e$ and $\bar{y}_H$, compared with its corresponding theoretical asymptotic variance, based on 10,000 samples. For response model (2): $p_i = 0.1 + 0.1x_{1i} + 0.8x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\hat{V}(\bar{y}_e)$	0.29	1.00	-0.14	0.21	4.89	8.25	0.62	0.5334
$\hat{V}(\bar{y}_H)$	1.00	2.30	0.63	1.70	1.60	2.71	3.14	5.69

Table 2.17: Relative biases of the variance estimator of $\bar{y}_e$ and $\bar{y}_H$, compared with its corresponding theoretical asymptotic variance, based on 10,000 samples. For response model (3): $p_i = 0.1 + 0.25x_{1i} + 0.65x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\hat{V}(\bar{y}_e)$	0.69	0.86	0.64	0.83	1.94	3.04	0.64	0.57
$\hat{V}(\bar{y}_H)$	1.75	2.97	1.44	2.44	1.60	2.50	2.48	4.25

Table 2.18: Relative biases of the variance estimator of $\bar{y}_e$ and $\bar{y}_H$, compared with its corresponding theoretical asymptotic variance, based on 10,000 samples. For response model (4) $p_i = 0.1 + 0.45x_{1i} + 0.45x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\hat{V}(\bar{y}_e)$	0.54	0.59	0.94	1.44	0.97	1.40	0.53	0.62
$\hat{V}(\bar{y}_H)$	2.04	3.49	1.56	2.62	1.61	2.64	2.05	3.54

ance estimate of $\bar{y}_{\hat{H}}$. There are two reasons for this substitution. First, the theoretical asymptotic variance of $\bar{y}_{\hat{H}}$ is very hard to derive and thus we don't have the explicit expression for variance estimator of $\bar{y}_{\hat{H}}$. Second, the simulation study shows that the $\hat{\boldsymbol{\nu}}$ is very close to $\hat{\boldsymbol{\nu}}_H$, so it's safe to

Table 2.19: Relative biases of the variance estimator of $\bar{y}_e$ and $\bar{y}_H$, compared with its corresponding theoretical asymptotic variance, based on 10,000 samples. For response model (5) $p_i = 0.1 + 0.65x_{1i} + 0.25x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\hat{V}(\bar{y}_e)$	0.42	0.44	1.80	3.08	0.49	0.76	0.50	0.82
$\hat{V}(\bar{y}_H)$	2.17	4.03	1.29	2.32	1.18	2.24	1.47	2.82

Table 2.20: Relative biases of the variance estimator of $\bar{y}_e$ and $\bar{y}_H$, compared with its corresponding theoretical asymptotic variance, based on 10,000 samples. For response model (6): $p_i = 0.1 + 0.8x_{1i} + 0.1x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\hat{V}(\bar{y}_e)$	0.46	0.41	4.77	8.16	-0.08	0.21	0.22	1.01
$\hat{V}(\bar{y}_H)$	2.93	5.51	1.36	2.50	0.53	1.53	0.82	2.15

Table 2.21: Relative biases of the variance estimator of $\bar{y}_e$ and $\bar{y}_H$, compared with its corresponding theoretical asymptotic variance, based on 10,000 samples. For response model (7): $p_i = 0.1 + 0.9x_{1i} + 0x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\hat{V}(\bar{y}_e)$	-0.71	0.05	1.97	4.05	-2.61	-1.55	-1.60	0.88
$\hat{V}(\bar{y}_H)$	0.31	1.68	0.09	1.42	-1.68	-0.31	-1.31	0.46

make this substitution. In fact, in most cases, the Monte Carlo variance of $\bar{y}_H$ is larger than that of $\bar{y}_{\hat{H}}$, so we are "conservative" to replace the variance estimate of $\bar{y}_{\hat{H}}$ with variance estimate of $\bar{y}_H$. From the results in Table 2.22 through Table 2.28, for all combinations of response model and outcome model, the coverage probabilities are all above 91.5%, with most coverage probabilities around 95%, which agree with the asymptotic normality of $\bar{y}_e$ and $\bar{y}_H$ in Theorem 4 and Theorem 7.

Table 2.22: Coverage of 95 percent confidence interval for proposed estimators, based on 10,000 samples. For response model (1): $p_i = 0.1 + 0x_{1i} + 0.9x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimate	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_e$	94.08	93.16	93.97	94.11	94.71	95.02	94.67	94.74
$\bar{y}_H$	94.52	94.53	93.85	93.89	94.76	94.8	94.1	94.15
$\bar{y}_{\hat{H}}$	95.48	96.58	92.67	91.22	94.66	94.62	95.87	97.25

Table 2.23: Coverage of 95 percent confidence interval for proposed estimators, based on 10,000 samples. For response model (2): $p_i = 0.1 + 0.1x_{1i} + 0.8x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimator	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_e$	94.86	94.99	94.61	94.61	94.98	95.21	94.92	94.92
$\bar{y}_H$	94.99	95.22	94.83	94.64	95.02	95.29	94.93	94.96
$\bar{y}_{\hat{H}}$	95.72	96.59	94.1	93.69	94.89	95.03	96.18	97.29

Table 2.24: Coverage of 95 percent confidence interval for proposed estimators, based on 10,000 samples. For response model (3): $p_i = 0.1 + 0.25x_{1i} + 0.65x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimate	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_e$	95	95.16	94.6	94.8	94.93	94.99	94.73	94.95
$\bar{y}_H$	94.79	95.02	94.74	94.88	94.81	94.94	94.72	94.99
$\bar{y}_{\hat{H}}$	95.67	96.47	94.66	94.83	95.09	95.26	95.74	96.69

Table 2.25: Coverage of 95 percent confidence interval for proposed estimators, based on 10,000 samples. For response model (4) $p_i = 0.1 + 0.45x_{1i} + 0.45x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimate	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_e$	95.06	95.02	94.56	94.73	94.94	95.07	94.53	94.8
$\bar{y}_H$	94.94	94.99	94.27	94.89	94.85	94.92	94.43	94.92
$\bar{y}_{\hat{H}}$	95.74	96.44	94.61	95.16	95.2	95.15	95.35	96.15

Table 2.26: Coverage of 95 percent confidence interval for proposed estimators, based on 10,000 samples. For response model (5) $p_i = 0.1 + 0.65x_{1i} + 0.25x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimate	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_e$	94.81	95.01	94.29	94.56	94.79	95	94.35	94.68
$\bar{y}_H$	94.81	94.7	94.35	94.71	94.69	94.58	94.37	94.76
$\bar{y}_{\hat{H}}$	95.66	96.77	94.47	95.02	94.63	94.61	95.25	96.3

Table 2.27: Coverage of 95 percent confidence interval for proposed estimators, based on 10,000 samples. For response model (6): $p_i = 0.1 + 0.8x_{1i} + 0.1x_{2i}$

	calibrate on x_1				calibrate on x_2			
estimate	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_e$	94.71	95.25	94.18	94.76	94.5	94.81	94.32	94.77
$\bar{y}_H$	94.62	94.8	94.33	94.45	94.53	94.65	94.36	94.63
$\bar{y}_{\hat{H}}$	96.01	97.38	94.22	94.56	93.82	93.51	95.17	96.65

Table 2.28: Coverage of 95 percent confidence interval for proposed estimators, based on 10,000 samples. For response model (7): $p_i = 0.1 + 0.9x_{1i} + 0x_{2i}$.

	calibrate on x_1				calibrate on x_2			
estimate	model1	model2	model3	model4	model1	model2	model3	model4
$\bar{y}_e$	94.72	95.11	94.17	94.41	94.4	94.43	93.75	93.34
$\bar{y}_H$	94.61	94.45	94.31	94.58	94.13	93.85	94.28	94.58
$\bar{y}_{\hat{H}}$	96.28	97.6	94.19	94.29	93.03	91.55	95.51	96.43

2.7 Conclusion

In this chapter, we studied the properties of the nonresponse weighting adjustment estimators where the response probability is modeled in linear form and the parameters are estimated by fitting a constrained least square regression model, with the constraint being a calibration equation. Both Horvitz-Thompson type estimator and Hájek type estimator are considered in this article. In theory, both estimators are shown to be asymptotically unbiased for the population mean and variance estimators of $\bar{y}_e$ and $\bar{y}_H$ are also shown to be consistent to the corresponding asymptotic variance. Furthermore, under the regular design assumptions, we proved that both estimators are asymptotically normally distributed. All the asymptotic properties are supported by the simulation study, where we set the response probability model to be a simple linear model.

From simulation study, our proposed Horvitz-Thompson type estimator works significantly better than its corresponding unconstrained estimator when the calibrated variable is highly correlated with the study variable and contribute less to the response model. In contrast, under the simple linear response model setting, the Hájek type estimator perform better, as long as the calibrated variable is correlated with the study variable to some degree.

The method we proposed is simple and easy to be implemented in real surveys, compared with other complex estimation technique. Also, the response model can be specified very flexibly. Yet, response model selection and how to bound the estimated response probability may be the topics we need to look into in the future.

Chapter 3

Improved Variance Estimation for Inequality Constrained Domain Mean Estimators Using Survey Data

3.1 Statement of the Problem and Literature Review

We begin by reviewing the classical domain means estimator and establishing notation. Let $U = \{1, 2, \cdots, N\}$ denote the finite population of size N. For example, this might represent all salaried employees in the United States. A sample $s \subset U$ is to be drawn from the population according to probability sampling design p, where $p(s)$ is the probability of drawing the sample s. The design determines the first order inclusion probability $\pi_i = \Pr(i \in s) = \sum_{i \in s} p(s)$, which is assumed to be positive for all $i = 1, \ldots, N$. The second order inclusion probability is $\pi_{ij} = \Pr(i, j \in s) = \sum_{i,j \in s} p(s)$, which is assumed to be positive for all $i, j \in U$. We denote the sample membership indicator $I_i(s) = 1$ if $i \in s$ and $I_i(s) = 0$ otherwise.

To define the domains of interest, let $\{U_d : d = 1, \cdots, D\}$ be a partition of the population U and N_d be the population size of domain U_d, where D is the fixed number of domains. These domains might be formed by a single variable, such as job level, or might be a grid of domains formed by several variables, such as job level, job type, and location. We denote by s_d the intersection of s and U_d. Let n be the sample size, and let n_d be the sample size for s_d.

Next we consider a study variable y, for which we are interested in estimating the population domain means. For example, y might be salary, and we denote by y_i the salary of the ith individual in the population. The population domain means are $\bar{\boldsymbol{y}}_U = (\bar{y}_{U_1}, \cdots, \bar{y}_{U_D})'$, where $\bar{y}_{U_d}$ is given by:

$$\bar{y}_{U_d} = \frac{\sum_{i \in U_d} y_i}{N_d} \quad d = 1, \cdots, D.$$

The goal is to estimate the $\bar{y}_{U_d}$ and provide inference such as confidence intervals. When no qualitative information such as ordering is available about the population domain means, the Horvitz-Thompson (HT) type (Horvitz and Thompson (1952)) estimator $\hat{y}_{s_d}$ or the Hájek estimator $\tilde{y}_{s_d}$ (Hájek (1971)) may be used. These are

$$\hat{y}_{s_d} = \frac{\sum_{i\in s_d} y_i/\pi_i}{N_d}, \text{ and } \tilde{y}_{s_d} = \frac{\sum_{i\in s_d} y_i/\pi_i}{\hat{N}_d}$$

respectively, where $\hat{N}_d = \sum_{i\in s_d} 1/\pi_i$. The Hájek estimator is often more useful in practice because it does not require information about the population domain size N_d, so we will focus only on properties based on the Hájek type estimator in this paper, which we refer to as the unconstrained estimators of $\bar{y}_{U_d}$. The results for HT estimator, however, can be derived analogously.

The unconstrained estimators might have large variance, especially when the sample sizes n_d are small. In practice, some of the n_d can be small even with large surveys, if there is a grid of many domains. It is helpful to use *a priori* knowledge regarding the population domain means. For example, the population domain means might be expected to increasing with respect to a given ordering based on job level. Wu et al. (2016) gave a derivation of this isotonic estimator and showed that it has smaller variance, compared with the unconstrained estimator. A diagnostic procedure was given by Oliva-Avilés et al. (2019); this can be used to verify that the imposed constraints are indeed satisfied by the population domain means.

More recently, estimation and inference with more general shape constraints was proposed by Oliva-Avilés et al. (2020). They considered assumptions that can be represented by a constraint matrix $\boldsymbol{A}$, where each of its rows defines a linear constraint on the domain means. For example, suppose we may assume that salaries in major metropolitan areas are higher than salaries in rural areas, for the same job type and level, but we do not have a complete ordering of areas. At the same time, we can impose some inequality restrictions on salaries by job type, or we can impose monotonicity constraints on a two-dimensional grid of domains. We assume that $\boldsymbol{A}$ is an $m \times D$ *irreducible* constraint matrix, which was defined by Meyer (1999). Intuitively, a constraint matrix is irreducible if the constraints are not redundant; see Oliva-Avilés et al. (2020) for more details.

The constrained estimator $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \cdots, \tilde{\theta}_D)'$ is the solution to the following constrained weighted least squares problem

$$\min_{\boldsymbol{\theta}} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta})' \boldsymbol{W}_s (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}) \quad \text{such that } \boldsymbol{A}\boldsymbol{\theta} \geq \mathbf{0} \tag{3.1}$$

where $\tilde{\boldsymbol{y}}_s = (\tilde{y}_{s_1}, \cdots, \tilde{y}_{s_D})'$ and $\boldsymbol{W}_s$ is the diagonal matrix with elements $\hat{N}_1/\hat{N}$, $\hat{N}_2/\hat{N}, \cdots, \hat{N}_D/\hat{N}$. If the constraints are satisfied by the Hájek estimator $\tilde{\boldsymbol{y}}_s$, then the minimizer of (3.1) coincides with $\tilde{\boldsymbol{y}}_s$.

Oliva-Avilés et al. (2020) gave an expression for the estimated covariance of $\tilde{\boldsymbol{\theta}}$; we provide an equivalent but simpler expression in Section 3.2. This estimated covariance matrix is used to construct confidence intervals for the population domain means, which are typically tighter than those for the unconstrained estimator. We use the simplified expression to derive a new estimated covariance matrix that is a mixture of possible covariance matrices of the previous type. This new covariance estimator takes into account the fact that the set of binding constraints, i.e. the elements of $\boldsymbol{A}\tilde{\boldsymbol{\theta}}$ that are zero, are random and for different samples the set of binding constraints may be different. In contrast, the classical constrained covariance estimator uses the observed set of binding constraints only. Simulations in Section 3.3 show that the new covariance estimator results in improved confidence interval coverage without sacrificing improvements in interval length.

3.2 Mixture Variance Estimation for the Domain Means Estimator

3.2.1 Review of the Formulation of the Constrained Domain Means Estimator

The following is a summary of the work in Oliva-Avilés et al. (2020). The first step is to transform the weighted constrained least-squares problem (3.1) into an unweighted projection, by letting $\tilde{\boldsymbol{z}}_s = \boldsymbol{W}_s^{1/2}\tilde{\boldsymbol{y}}_s$, $\boldsymbol{\phi} = \boldsymbol{W}_s^{1/2}\boldsymbol{\theta}$ and $\boldsymbol{A}_s = \boldsymbol{A}\boldsymbol{W}_s^{-1/2}$. Define $\tilde{\boldsymbol{\phi}}$ to be the unique vector that solves the least-squares problem:

$$\min_{\boldsymbol{\phi}} ||\tilde{\boldsymbol{z}}_s - \boldsymbol{\phi}||^2 \quad \text{such that } \boldsymbol{A}_s\boldsymbol{\phi} \geq \boldsymbol{0},$$

and subsequently $\tilde{\boldsymbol{\theta}} = \boldsymbol{W}_s^{-1/2}\tilde{\boldsymbol{\phi}}$.

The `R` package `coneproj` (Liao and Meyer (2014)) finds $\tilde{\boldsymbol{\phi}}$ given $\tilde{\boldsymbol{z}}_s$ and $\boldsymbol{A}_s$, and also returns the set of binding constraints. Let $J_s \subseteq \{1, \ldots, m\}$ indicate the zero elements of $\boldsymbol{A}_s\tilde{\boldsymbol{\phi}}$. If J_s is empty, then the constrained estimator coincides with the unconstrained estimator.

Let $\boldsymbol{A}_{J_s}$ denote the matrix formed by the rows of $\boldsymbol{A}$ indexed by J_s, where $\boldsymbol{A}_{J_s}$ is a zero matrix if J_s is empty. Otherwise, the rows of $\boldsymbol{A}_{J_s}$ will form a linearly independent set of vectors in $\mathbb{R}^D$ even if $\boldsymbol{A}$ is not full row rank. Let $\mathcal{I}_J(s) = 1$ if $J = J_s$ and $\mathcal{I}_J(s) = 0$ otherwise, and write the constrained estimator as:

$$\tilde{\boldsymbol{\theta}} = \sum_J [\big(\boldsymbol{I}_{D\times D} - \boldsymbol{W}_s^{-1}\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_s^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\big)\,\tilde{\boldsymbol{y}}_s]\mathcal{I}_J(s), \tag{3.2}$$

where the sum is over all subsets $J \subseteq \{1, \ldots, m\}$ such that the rows of $\boldsymbol{A}_J$ form a linearly independent set. For each sample s, there is only one subset J for which $\mathcal{I}_J(s) = 1$; this is J_s. The brief introduction of the formulation of the expression in (3.2) is in Appendix and refer to Oliva-Avilés et al. (2020) for more details. This expression of the constrained estimator is used to derive the improved covariance matrix estimator.

3.2.2 Assumptions

Before we present our theoretical results, we list and discuss our assumptions on the probability sampling design:

(A1) The number D of domains is a fixed integer. For $d = 1, 2, \cdots, D$, $\liminf_{N\to\infty} \frac{N_d}{N} > 0$ and $\limsup_{N\to\infty} \frac{N_d}{N} < 1$.

(A2) For the study variable $\{y_i\}_{i\in U}$, we have

$$\limsup_{N\to\infty} N^{-1} \sum_{i\in U} {y_i}^4 < \infty.$$

(A3) The sample size is non-random and there is a $\pi \in (0,1)$ such that $\min_d \frac{n_d}{N_d} \geq \pi$.

(A4) For all N, $\min_{i\in U} \pi_i \geq \lambda_1 > 0$ and $\min_{i,j\in U} \pi_{ij} \geq \lambda_2 > 0$, and

$$\limsup_{N\to\infty} n \max_{i,j\in U, i\neq j} |\Delta_{ij}| < \infty$$

where $\Delta_{ij} = \mathrm{cov}(I_i, I_j) = \pi_{ij} - \pi_i\pi_j$.

(A5) Let $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_D)'$ be a vector of limiting domain means, where $\bar{y}_{U_d} - \mu_d = O(N^{-\frac{1}{2}})$ for $d = 1, \ldots, D$. We assume $\boldsymbol{A\mu} \geq 0$.

(A6) The assumption involving higher-order inclusion probability:

$$\lim_{N\to\infty} \max_{(i_1,i_2,i_3,i_4)\in D_{4,N}} |\mathrm{E}[(I_{i_1}I_{i_2} - \pi_{i_1}\pi_{i_2})(I_{i_3}I_{i_4} - \pi_{i_3}\pi_{i_4})]| = 0,$$

where $D_{t,N}$ denotes the set of all distinct t-tuples $i_1, i_2, \cdots, i_t$ from U.

(A7) For any vector of D variables $\boldsymbol{x}$ with finite fourth population moment, we have:

$$\mathrm{var}(\hat{\boldsymbol{x}}_s)^{-\frac{1}{2}}(\hat{\boldsymbol{x}}_s - \bar{\boldsymbol{x}}_U) \xrightarrow{d} N(0, \boldsymbol{I}_D)$$

where $\hat{\boldsymbol{x}}_s$ is the HT domain mean estimator of $\bar{\boldsymbol{x}}_U = (N_1^{-1}\sum_{k\in U_1} x_k, \cdots, N_D^{-1}\sum_{k\in U_D} x_k)'$, $\boldsymbol{I}_D$ is the identity matrix of dimension D, the design covariance matrix $\mathrm{var}(\hat{\boldsymbol{x}}_s)$ is positive definite.

The assumption (A1) states that the number of domains remains constant when the population size N changes and guarantees that there is no asymptotically vanishing domains. Assumption (A2) is one of the conditions for showing the unbiasedness and variance consistency of the Horvitz-Thompson estimator and this can be applied to most types of survey data. In (A3) we are excluding vanishing sampling fraction to stay within the finite population framework, which is common in the design-based context. The assumption in (A3) may exclude small area estimation, if n_d/N_d becomes negligible.

Assumption (A4) ensures that the design is both a probability sampling design and a measurable design. The assumption on the Δ_{ij} states that the covariance between sample membership indicators is sufficiently small. These are satisfied for classical sampling designs, including simple random sampling with and without replacement, and also holds for rejective sampling (Hájek (1964)). This condition does not generally hold for multistage sampling designs. Under (A2) and (A4), using the same bounding arguments presented in Section 3 in Breidt and Opsomer (2017), together with the unbiasedness of HT type estimator, we can show the sample moments converge to population moments. That is, for example in terms of study variable y, we have:

$$\frac{1}{N}\sum_{i\in s}\frac{y_i}{\pi_i} - \frac{1}{N}\sum_{i\in U} y_i = O_p(n^{-\frac{1}{2}}).$$

Assumptions in (A5) ensure that the limiting domain means satisfy the shape constraints, although the population domain means may deviate slightly from the constraints.

Assumptions in (A6) involve conditions on correlations up to order four, which are difficult to check for complex sampling designs. They are similar to the higher order assumptions considered by Breidt and Opsomer (2000), which are needed for proving the consistency and the asymptotic normality of some complex estimators. Boistard et al. (2012) proved that these assumptions hold for rejective sampling, which is an unequal probability sampling design with fixed sample size. Simple random sampling without replacement is a particular case of rejective sampling.

Assumption (A7) is satisfied for many specific sampling designs, including simple random sampling with or without replacement, Poisson sampling and unequal probability sampling with replacement. This ensures asymptotic normality for a general finite fourth moment vector of variables $\boldsymbol{x}$. The design asymptotic normality assumption, together with the variance consistency of the Horvitz-Thompson estimator, can be used to obtain the asymptotic distribution of the shape-restricted estimator. This assumption is used when estimating the mixture probabilities in our proposed mixture variance estimator; see Section 3.2.4 and the proof of Theorem 9.

Overall, the assumptions in this paper are almost the same as the ones in Oliva-Avilés et al. (2020), except that Oliva-Avilés et al. (2020) explicitly assumed the variance consistency of the HT estimator, while in this paper, we relaxed that assumption. Instead, we formally proved the consistency of the variance estimator of the HT estimator by using the higher-order inclusion probability assumption in (A6), which is a required condition in proving the variance consistency of the HT estimator.

3.2.3 Linearized Variance Estimation of the Domain Mean Estimator

The following is proved in Appendix B.

Proposition 1. *The following expression for the asymptotic covariance matrix of $\tilde{\boldsymbol{\theta}}$ is equivalent to the expression in Oliva-Avilés et al. (2020):*

$$AV(\tilde{\boldsymbol{\theta}}) = \sum_{J} \left[(\boldsymbol{I} - \boldsymbol{P}_J) \boldsymbol{\Sigma}_J (\boldsymbol{I} - \boldsymbol{P}_J)' \right] \mathcal{I}_J(s) \tag{3.3}$$

where $\boldsymbol{P}_J = \boldsymbol{W}_U^{-1} \boldsymbol{A}_J' [\boldsymbol{A}_J \boldsymbol{W}_U^{-1} \boldsymbol{A}_J']^{-1} \boldsymbol{A}_J$ and the ijth element of $\boldsymbol{\Sigma}_J$ is given by

$$\{\boldsymbol{\Sigma}_J\}_{ij} = \frac{1}{N_i N_j} \sum_{k \in U_i} \sum_{l \in U_j} \Delta_{kl} \frac{(y_k - \theta_{i,J})(y_l - \theta_{j,J})}{\pi_k \pi_l}, \quad i, j = 1, 2, \cdots, D.$$

and $\theta_{j,J} = \bar{y}_{U_j} - \frac{N}{N_j} \left\{ \boldsymbol{A}_J' (\boldsymbol{A}_J \boldsymbol{W}_U^{-1} \boldsymbol{A}_J')^{-1} \boldsymbol{A}_J \bar{\boldsymbol{y}}_U \right\}_j$.

If the observed $J_s = \emptyset$, then $AV(\tilde{\boldsymbol{\theta}})$ reduces to $AV(\tilde{\boldsymbol{y}}_s) = \boldsymbol{\Sigma}$, and the ijth element of $\boldsymbol{\Sigma}$ is:

$$\boldsymbol{\Sigma}_{ij} = \frac{1}{N_i N_j} \sum_{k \in U_i} \sum_{l \in U_j} \Delta_{kl} \frac{(y_k - \bar{y}_{U_i})(y_l - \bar{y}_{U_j})}{\pi_k \pi_l}, \quad i, j = 1, 2, \cdots, D.$$

Expression (3.3) is preferred to (B.1), the expression from Oliva-Avilés et al. (2020), from both an intuitive and a computational viewpoint. From (3.3), we can define the estimator of the asymptotic covariance matrix of $\tilde{\boldsymbol{\theta}}$ as follows:

$$\hat{V}_J(\tilde{\boldsymbol{\theta}}) = \sum_J \left[(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}_J(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)'\right] \mathcal{I}_J(s)$$

where $\hat{\boldsymbol{P}}_J = \boldsymbol{W}_s^{-1}\boldsymbol{A}'_J[\boldsymbol{A}_J\boldsymbol{W}_s^{-1}\boldsymbol{A}'_J]^{-1}\boldsymbol{A}_J$ and the elements of $\tilde{\boldsymbol{\Sigma}}_J$ are given by:

$$\{\tilde{\boldsymbol{\Sigma}}_J\}_{ij} = \frac{1}{\hat{N}_i\hat{N}_j}\sum_{k\in s_i}\sum_{l\in s_j}\frac{\Delta_{kl}}{\pi_{kl}}\frac{(y_k - \tilde{\theta}_{i,J})(y_l - \tilde{\theta}_{j,J})}{\pi_k\pi_l}, \quad i,j = 1,2,\cdots,D.$$

An even simpler estimator is

$$\hat{V}(\tilde{\boldsymbol{\theta}}) = \sum_J \left[(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)'\right] \mathcal{I}_J(s) \tag{3.4}$$

where the ijth element of $\tilde{\boldsymbol{\Sigma}}$ is given by:

$$\{\tilde{\boldsymbol{\Sigma}}\}_{ij} = \frac{1}{\hat{N}_i\hat{N}_j}\sum_{k\in s_i}\sum_{l\in s_j}\frac{\Delta_{kl}}{\pi_{kl}}\frac{(y_k - \tilde{y}_{s_i})(y_l - \tilde{y}_{s_j})}{\pi_k\pi_l}, \quad i,j = 1,2,\cdots,D.$$

That is, $\tilde{\boldsymbol{\Sigma}}$ is the covariance estimator of the unconstrained estimator $\tilde{\boldsymbol{y}}_s$. The proof of the following is in Appendix B.

Theorem 8. *Under assumptions (A1)-(A6), the covariance estimator $\hat{V}(\tilde{\boldsymbol{\theta}})$ is consistent for AV$(\tilde{\boldsymbol{\theta}})$ in the sense that:*

$$n(\hat{V}(\tilde{\boldsymbol{\theta}}) - \mathit{AV}(\tilde{\boldsymbol{\theta}})) = o_p(1) \tag{3.5}$$

where AV$(\tilde{\boldsymbol{\theta}})$, $\hat{V}(\tilde{\boldsymbol{\theta}})$ are given in (3.3), (3.4), respectively.

3.2.4 The Proposed Mixture Covariance Estimator

From (3.2), the constrained domain mean estimator can be expressed as:

$$\begin{aligned}\tilde{\boldsymbol{\theta}} &= \sum_J \left(\boldsymbol{I}_{D\times D} - \boldsymbol{W}_s^{-1}\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_s^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\right)\tilde{\boldsymbol{y}}_s\mathcal{I}_J(s)\\ &= \sum_J \left(\boldsymbol{I}_{D\times D} - \hat{\boldsymbol{P}}_J\right)\tilde{\boldsymbol{y}}_s\mathcal{I}_J(s)\end{aligned}$$

Instead of using only the observed J to compute the estimated covariance matrix, we propose the following mixture covariance of the constrained domain mean estimator.

$$AV^m(\tilde{\boldsymbol{\theta}}) = \sum_J(\boldsymbol{I} - \boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I} - \boldsymbol{P}_J)'P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J) \tag{3.6}$$

where we define the $\mathcal{C}_J$ to be the set of points $\tilde{\boldsymbol{y}}_s \in \mathbb{R}^D$ such that, based on those points, the corresponding unweighted projection algorithm will return the set of binding constraints J.

This estimator recognizes that a different sample with the same sample size and design might correspond to a different J, and an improved variance estimator uses a mixture of the possible Js in roughly the proportions corresponding the probabilities of observing the Js. In theory, there are a large number of J sets. Asymptotically, the probability of $\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J$ goes to zero for most of the J sets, and for moderate-to-large sample sizes we will observe only a few J with substantial probabilities. For example, if the elements of $\boldsymbol{A\mu}$ are strictly positive, the consistency of the estimator guarantees that as n and N increase without bound, all of the constraints must become unbinding, and the probability that J is the empty set goes to one. If one or more of the elements of $\boldsymbol{A\mu}$ are zero, the J sets that correspond to these constraints being binding or not become the only J sets with nonzero probability. By constructing our estimator of $AV^m(\tilde{\boldsymbol{\theta}})$ in (3.6), we are no longer conditioning on the J_s. This will tend to increase the estimated variance but this better reflects the underlying variance of the constrained estimator.

To estimate the mixture probabilities, we generate many $\boldsymbol{y}^{(i)}$'s identically and independently from a multivariate normal distribution with mean $\tilde{\boldsymbol{\theta}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$, i.e.,

$$\boldsymbol{y}^{(i)} \stackrel{iid}{\sim} \mathrm{MVN}(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}).$$

Thus the $\boldsymbol{y}^{(i)}$ have approximately the distribution of $\tilde{\boldsymbol{y}}_s$, and can be used to simulate the distribution of the set of binding constraints J. From each simulated $\boldsymbol{y}^{(i)}$, we observe the corresponding set J, and with repeated sampling we can tally the number of times each set J is observed. Specifically, if B is the number of simulations, then for a particular J set, we use $B^{-1}\sum_{i=1}^{B} I(\boldsymbol{y}^{(i)} \in \mathcal{C}_J)$ to estimate the mixture probability $P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J)$.

The use of the normal distribution for the $\boldsymbol{y}^{(i)}$s is motivated by the asymptotic normality of $\tilde{\boldsymbol{y}}_s$. Hence, this simulated distribution is an approximation that improves as n and N increase. Since $\tilde{\boldsymbol{y}}_s$ has asymptotic multivariate normal distribution with mean $\bar{\boldsymbol{y}}_U$ and covariance $\boldsymbol{\Sigma}$ by assumption (A7), taken together with the fact that $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\Sigma}}$ are consistent for $\bar{\boldsymbol{y}}_U$ and $\boldsymbol{\Sigma}$ respectively, $P(\boldsymbol{y}^{(i)} \in \mathcal{C}_J)$ should approach $P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J)$ as the sample size increases. As we set B to be large, $B^{-1}\sum_{i=1}^{B} I(\boldsymbol{y}^{(i)} \in \mathcal{C}_J)$ approaches $P(\boldsymbol{y}^{(i)} \in \mathcal{C}_J)$ by the law of large numbers.

Finally, the proposed mixture covariance estimator for $AV^m(\tilde{\boldsymbol{\theta}})$ is expressed as:

$$\begin{aligned}
\hat{V}^m(\tilde{\boldsymbol{\theta}}) &= \sum_J (\boldsymbol{I} - \hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)' \frac{1}{B}\sum_{i=1}^{B} I(\boldsymbol{y}^{(i)} \in \mathcal{C}_J) \\
&= \frac{1}{B}\sum_{i=1}^{B}\left(\sum_J (\boldsymbol{I} - \hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)' I(\boldsymbol{y}^{(i)} \in \mathcal{C}_J)\right) \\
&= \frac{1}{B}\sum_{i=1}^{B} (\boldsymbol{I} - \hat{\boldsymbol{P}}_{J^{(i)}})\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_{J^{(i)}})
\end{aligned} \tag{3.7}$$

The following result is proved in the Appendix B.

Theorem 9. *If $\boldsymbol{A\mu} > 0$, then under assumptions (A1)-(A6), the proposed mixture covariance estimator is consistent for the asymptotic mixture covariance of $\tilde{\boldsymbol{\theta}}$ in the sense that, as $B \to \infty$:*

$$n(\hat{V}^m(\tilde{\boldsymbol{\theta}}) - AV^m(\tilde{\boldsymbol{\theta}})) = o_p(1).$$

where $AV^m(\tilde{\boldsymbol{\theta}})$, $\hat{V}^m(\tilde{\boldsymbol{\theta}})$ are given in (3.6), (3.7), respectively.

It is useful to note that inference based on the mixture variance estimator remains design-based. Only the design variability is accounted for by the asymptotic variance in Theorem 9. The consistent estimator of the design variance uses a parametric bootstrap approach, with the asymptotic normal distribution of the estimator serving as bootstrap distribution. While the qualitative constraints can be viewed as "model-like" assumptions, they do not imply a random structure for the population and the inference does not involve any type of model variability.

3.3 Simulation Studies

We compare the length and coverage probabilities of domain mean confidence intervals, using three methods: the unconstrained Hájek domain means estimator with its covariance estimator, the constrained domain means estimator with the covariance estimate based on the observed J, as in Oliva-Avilés et al. (2020), and the constrained estimator with the proposed mixture covariance estimator. The simulation scenarios involve one- or two-dimensional grids, with various constraints and μ values. We report results from three scenarios: for each, we generate a population, then we draw 10,000 samples from the population according to a sampling design. For each sample we compute the three confidence intervals; we present the coverage rates (proportion of intervals capturing the population mean, for each domain) and average interval lengths in graphical form. The results show that the proposed estimator has coverage probabilities that meet or exceed the target 95%, and indeed have higher coverage than those for the unconstrained estimator, while retaining the smaller confidence interval length of the intervals of Oliva-Avilés et al. (2020).

3.3.1 Isotonic in One Variable

As in Wu et al. (2016) and Oliva-Avilés et al. (2020), we choose the limiting domain means for generating the population elements to be in a sigmoidal shape across the domains:

$$\mu_d = \frac{\exp(20d/D - 10)}{1 + \exp(20d/D - 10)} \text{ for } d = 1, 2, \cdots, D,$$

where $D = 20$ is the number of domains. The population size is set to be $N = 8000$, with domain population size $N_d = N/D$. The study variables $y_1, \ldots, y_N$ are generated by adding independent and identically distributed $N(0, 1)$ errors to the μ_d values.

Samples are drawn without replacement from a stratified simple random sampling design, with $H = 4$ strata that cut across the D domains. The strata are determined by a variable z, which is correlated with y. The values of z are generated by adding standard normal errors to (d/D). Then, the stratum membership is determined by sorting the population based on the corresponding ranked z, so that there are N/H elements for each stratum. Finally, the total sample size $n = 200$ is assigned to each strata with sample size $(25, 50, 50, 75)$ in each strata accordingly.

Simulation results are summarized in Figure 3.1, along with a typical sample. The population means are shown as black dots; due to randomness in generating the population, these do not quite satisfy the constraints. The coverage is highest for the intervals computed using the mixture covariance estimator. The average length of the confidence interval for the constrained estimators, either based on non-mixture variance or mixture variance, are narrower than that of the unconstrained estimator.

3.3.2 Block Isotonic in One Variable

We consider the scenario in which only a partial ordering is known *a priori*. For a "block isotonic" ordering, we assume the population means are ordered among blocks, but no ordering is assumed within blocks. In particular, we have $D = 20$ domains, organized in four blocks of five domains, and the population mean for each of the domains in block i is assumed to be at least as

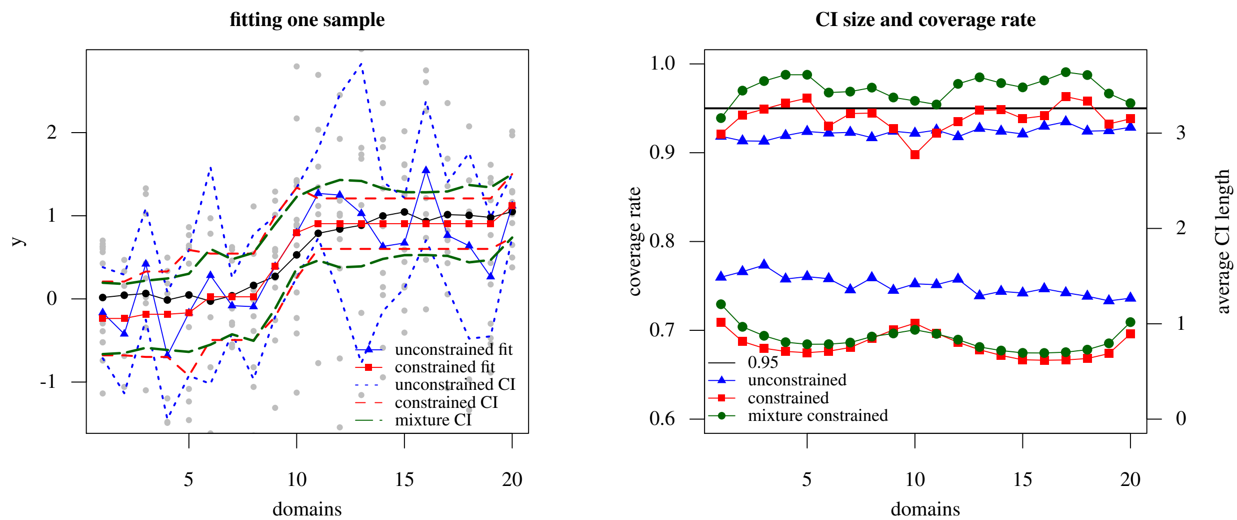

Figure 3.1: Isotonic in one variable: On the left is a typical sample, with the population means shown as black dots. On the right are the coverage probabilities and average lengths for 10,000 samples, for a target 95%.

large as those in block $i-1$, for $i = 2, 3, 4$. In particular,

$$\boldsymbol{\mu} = \left(-.1 \quad 0 \quad .1 \quad 0 \quad -.1 \;\middle|\; .2 \quad .1 \quad .2 \quad .3 \quad .2 \;\middle|\; .3 \quad .4 \quad .3 \quad .4 \quad .5 \;\middle|\; .5 \quad .6 \quad .5 \quad .7 \quad .6 \right)$$

where the blocks are separated by the vertical lines. We use the same stratified sampling design as in the previous example.

The results in Figure 3.2 show that again the coverage is highest for the mixture constrained estimator. The confidence interval lengths are smallest for the middle two blocks, because the estimators for these domains are able to use information from all four blocks.

3.3.3 Isotonic in Two Variables

Here we consider a grid of domains, which represent two variables such as job type and job level, with the *a priori* assumption that the μ values are non-decreasing in each. In particular, we consider that there are six levels of one variable and five of the other, so there are $D = 30$ domains.

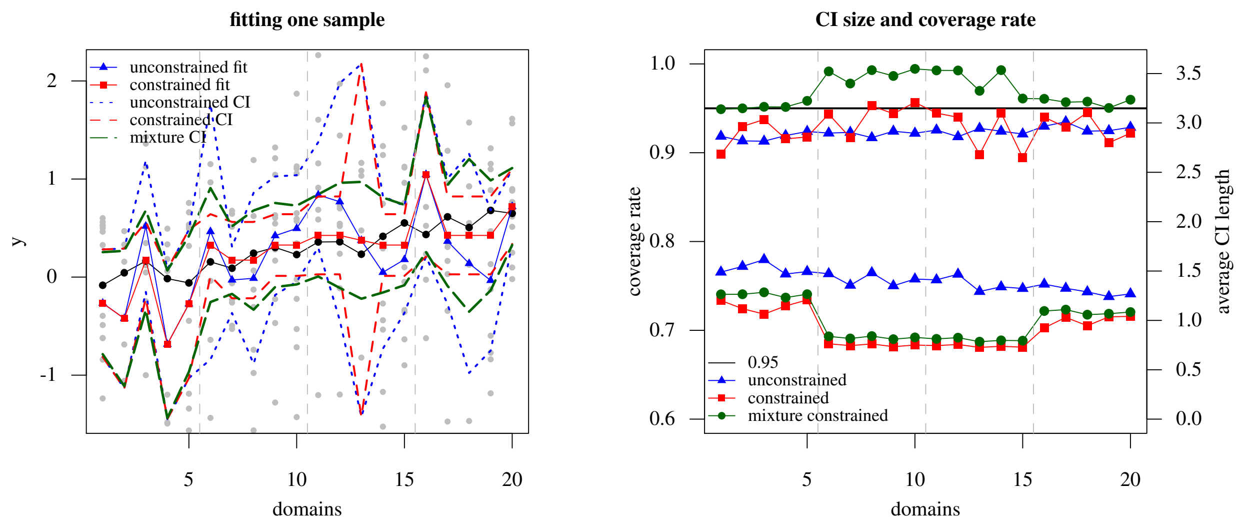

Figure 3.2: Block isotonic in one variable: On the left is a typical sample, with the population means shown as black dots. On the right are the coverage probabilities and average lengths for 10,000 samples, for a target 95%.

We choose μ values according to the array

$$\boldsymbol{\mu} = \begin{pmatrix} .2 & .4 & .8 & 1.1 & 1.4 & 1.8 \\ .1 & .3 & .5 & .8 & 1.0 & 1.2 \\ .1 & .2 & .3 & .5 & .7 & .8 \\ .1 & .2 & .3 & .4 & .5 & .6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The procedure for generating y and the sampling mechanism are the same as that in one dimensional case, except that we set $N = 12000$ and $n = 240$, which is divided among the strata as $(30, 60, 60, 90)$. A typical sample is displayed on the left in Figure 3.3, where the numbering for the domains is "by row" of the μ matrix; that is, the first six domains have means in the top row of the the μ matrix, the next six have means from the second row, and the last six domains have limiting mean zero.

The simulation results in the right plot of the figure demonstrate similar properties as the previous scenarios: the lengths of the confidence intervals of the constrained estimator are smaller while the coverage is the highest when using the mixture variance estimator.

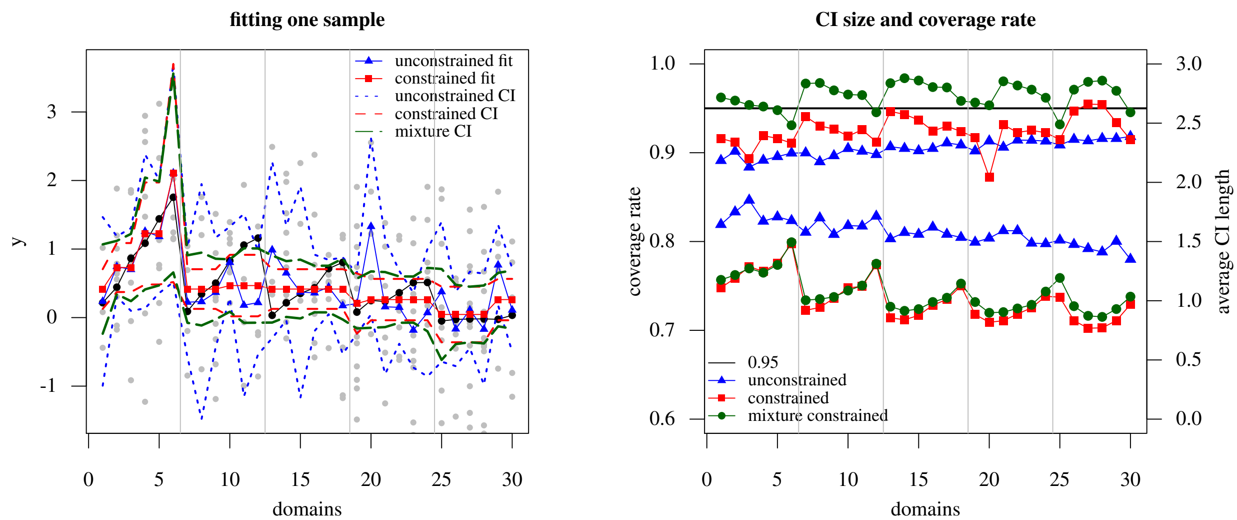

Figure 3.3: Isotonic in two variables: On the left is a typical sample, with the population means shown as black dots. On the right are the coverage probabilities and average lengths for 10,000 samples, for a target 95%.

3.4 Application to California School Data

The `survey` package in `R` provides the data set `apipop`, with information about the population of California schools, including average standardized test scores and demographic variables. This data set is convenient for demonstrating survey methods, because we can sample from the population of over 6000 schools, apply our methods to the sample, and at the same time we can compute the true population values. For our purpose, we use the variables `api00` and `meals`; the former is the average standardized test score at the school, and the latter is the percent of students who are eligible for subsided or free meals (breakfast and lunch). We categorize the `meals` variable into 20 domains, where domain one indicates 0-4% free meals, domain two is 5-9%, etc. Percent subsidized meals is a measure of poverty, which in the United States is connected with educational and health disadvantages (to our shame). Hence, we can assume *a priori* that the average test score is decreasing over these domains, and in fact we see that this is true in the population.

We collect a sample that is stratified by type of school (elementary, middle, high), with sample sizes 80, 80, and 120, respectively. A typical data set is shown in Figure 3.4, where the gray dots represent the sample. The population mean standardized test scores are the black dots, which decrease over these domains. The unconstrained sample means are decreasing overall, but have some deviations from monotonicity. The constrained estimators reflect our *a priori* knowledge, and in addition, imposing the constraints gives smaller confidence intervals and better coverage. As was seen in the simulations, using the mixture covariance matrix further improves the coverage rates.

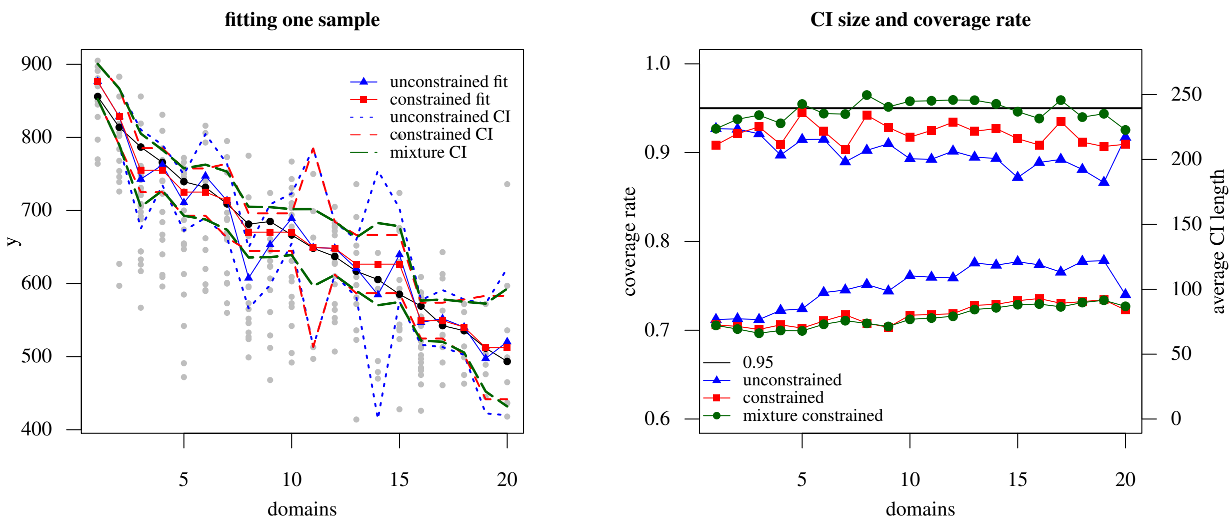

Figure 3.4: Decreasing in poverty levels: On the left is a typical sample, with the population means shown as black dots. On the right are the coverage probabilities and average lengths for 10,000 samples, for a target 95%.

3.5 Discussion

This paper provides an improvement in the theoretical foundation for inference with survey data while utilizing *a priori* information in the form of inequality constraints, such as for ordering or shape. The simpler form of the estimated covariance matrix in (3.7) readily allows computation, with $\tilde{\Sigma}$ obtained from survey software such as the `survey` package in `R`, and a constrained least-squares routine such as `coneproj` used to find $J^{(i)}$ for simulated samples. The paper also provides an improved estimator of the design-based variance that results in increased coverage compared to previous results, exceeding the nominal level in most cases.

Chapter 4

One-sided Testing of Population Domain Means in Surveys

4.1 Introduction

Methods for estimation of population domain means under *a priori* assumptions in the form of linear inequality constraints have been recently established. Suppose interest is in estimating $\bar{\boldsymbol{y}}_U \in \mathbb{R}^D$, a vector of population domain means. Wu et al. (2016) derived an isotonic survey estimator, where it is assumed that $\bar{y}_{U_1} \leq \cdots \leq \bar{y}_{U_D}$. They showed that the constrained estimator is equivalent to a "pooled" estimator, where weighted averages of adjacent sample domain means are used to form an isotonic vector of domain mean estimates. Advantages to the ordered mean estimates are that they "make sense" in terms of satisfying the assumptions, and the confidence intervals for the estimates are typically reduced in length. Oliva-Avilés et al. (2020) proposed a framework for the estimation and inference with more general shape and order constraints in survey contexts. Examples include block orderings, and orderings of domain means arranged in grids. For example, average cholesterol level may be assumed to be increasing in age category and BMI level, but decreasing in exercise category. In another context, suppose average salary is to be estimated by job rank, job type, and location, with average salary assumed to be increasing with rank, and block orderings imposed on job type and location. More recently, Xu et al. (2021) formulated a mixture covariance matrix for constrained estimation that was shown to improve coverage of confidence intervals while retaining the smaller lengths.

The desired linear inequality constraints may be formulated using an $m \times D$ constraint matrix $\boldsymbol{A}$, where the assumption is $\boldsymbol{A}\bar{\boldsymbol{y}}_U \geq \mathbf{0}$. For the isotonic domain means, $m = D - 1$, and the nonzero elements of the constraint matrix are $\boldsymbol{A}_{i,i} = -1$ and $\boldsymbol{A}_{i,i+1} = 1$. For block orderings, where domains are grouped by ordered blocks, each domain in block one, for example, is assumed to have a population mean not larger than each domain in block two, and in block two, each population domain mean does not exceed any of those in block three, etc. Here the number of constraints is $m = \sum_{i=1}^{B-1}\sum_{j=i+1}^{B} k_i k_j$, where B is the number of blocks and k_i is the number of domains in the ith block, $i = 1, \ldots, B$. For a third example, consider domains arranged in a grid; for a context suppose the population units are lakes in a state, and y_i is the level of a certain pollutant in lake i. We are interested in average levels by county and by distance from an industrial plant. If there are 60 counties and 5 categories of distance, there are 300 domains. If we know that the level of pollutant is non-increasing in the distance variable, then there are 60×4=240 constraints, formulated as antitonic within each county.

We propose a test where the null hypothesis is that $\boldsymbol{A}\bar{\boldsymbol{y}}_U = \mathbf{0}$, versus the alternative $\boldsymbol{A}\bar{\boldsymbol{y}}_U \geq \mathbf{0}$, and $\boldsymbol{A}\bar{\boldsymbol{y}}_U$ has at least one positive element. The simplest example is the null hypothesis of constant domain means, versus the alternative of increasing domain means. For the third example above, we can test the null hypothesis that, within each county, the domain means are constant in distance. Using the constraints for a one-sided alternative results in improved power over the equivalent two-sided test.

This test has been widely studied outside of the survey context; see Bartholomew (1959), Bartholomew (1961), Chacko (1963), McDermott and Mudholkar (1993), Robertson et al. (1988), Meyer (2003), Silvapulle and Sen (2005), Sen and Meyer (2017) and others. The null distribution of the likelihood-ratio test statistic for the one-sided test has been derived based on the normal-errors model. In brief, when the model variance is known, the null distribution of the likelihood ratio statistic is shown to be a mixture of chi-square distributions, while for the unknown model variance, the test statistic has the null distribution of a mixture of beta distributions. Similar results for the one sided likelihood ratio test were obtained by Perlman (1969) where the completely unknown model variance was considered. Meyer and Wang (2012) formally proved that the one-sided test will provide higher power than the test using the unconstrained alternative.

In this paper we extend this test to the survey context. In the next section, the test is derived, and in Section 4.3 some large sample theory is given. Simulations in Section 4.4 show that the test performs well compared to the test with the unconstrained alternative, with better power and a test size closer to the target. In Section 4.5 the methods are applied to the National Survey of College Graduates, to test whether salaries are higher for people whose father's education level is higher, controlling for field of study, highest degree attained, and year of degree. The test is available in the `R` package `csurvey`.

4.2 Formulation of the Test Statistic

To establish the notation, let $U = \{1, 2, \cdots, N\}$ be the finite population. A sample $s \subset U$ of size n is to be drawn based on a probability sampling design p, where $p(s)$ is the probability of drawing the sample s. The first order inclusion probability $\pi_i = \Pr(i \in s) = \sum_{i \in s} p(s)$ and the second order inclusion probability $\pi_{ij} = \Pr(i, j \in s) = \sum_{i,j \in s} p(s)$, determined by the sampling design, are both assumed to be positive. In terms of the domains of interest, let $\{U_d : d = 1, \cdots, D\}$ be a partition of the population U and N_d be the population size of domain d, where D is the number of domains. We denote by s_d the intersection of s and U_d, and let n_d be the sample size for s_d.

Let y be the variable of interest and denote by y_i the value for the ith unit in the population. The population domain means are $\bar{\boldsymbol{y}}_U = (\bar{y}_{U_1}, \cdots, \bar{y}_{U_D})^\top$, and $\bar{y}_{U_d}$ is given by:

$$\bar{y}_{U_d} = \frac{\sum_{i \in U_d} y_i}{N_d} \quad d = 1, \cdots, D.$$

Two common design-based estimators of the population means are the Horvitz-Thompson estimator (Horvitz and Thompson (1952)) or the Hájek estimator (Hájek (1971)); because the Hájek estimator $\tilde{y}_{s_d}$ does not require information about the population domain size N_d and has other advantages in practice, we will focus on the Hájek estimator. The results for the Horvitz-Thompson estimator, however, can be derived analogously. The Hájek estimator for domain means is $\tilde{\boldsymbol{y}}_s = (\tilde{y}_{s_1}, \ldots, \tilde{y}_{s_D})$, where

$$\tilde{y}_{s_d} = \frac{\sum_{i \in s_d} y_i / \pi_i}{\hat{N}_d}$$

and $\hat{N}_d = \sum_{i \in s_d} 1/\pi_i$.

We are concerned with testing:

$$H_0 : \bar{\boldsymbol{y}}_U \in V \quad \text{versus} \quad H_1 : \bar{\boldsymbol{y}}_U \in \mathcal{C} \backslash V \tag{4.1}$$

where $V = \{\boldsymbol{y} : \boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}\}$ is the null space of $\boldsymbol{A}$ and the alternative set is the convex cone $\mathcal{C} = \{\boldsymbol{y} : \boldsymbol{A}\boldsymbol{y} \geq \boldsymbol{0}\}$. A set $\mathcal{C}$ is a convex cone if for any $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in $\mathcal{C}$, $\alpha_1\boldsymbol{\theta}_1 + \alpha_2\boldsymbol{\theta}_2$ is in $\mathcal{C}$ for any non-negative α_1 and α_2.

For testing (4.1), we propose the following weighted least squares test statistic:

$$\hat{T} = \frac{\min_{\boldsymbol{\theta}_0 \in V}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0) - \min_{\boldsymbol{\theta}_1 \in \mathcal{C}}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)}$$

where $\tilde{\boldsymbol{\Sigma}}$ is the covariance estimator of $\tilde{\boldsymbol{y}}_s$. We will reject H_0 if $\hat{T}$ is large.

This is similar in structure to the classical test. If $\tilde{\boldsymbol{y}}_s$ were normal with $\mathrm{cov}(\tilde{\boldsymbol{y}}_s) = \tilde{\boldsymbol{\Sigma}}$, then $\hat{T}$ would be distributed as a mixture of beta random variables, under the null hypothesis. In the survey context, we approximate the distribution of $\hat{T}$.

4.3 Asymptotic Distribution of the Test Statistic

The assumptions needed to derive an approximate distribution of T are listed in Appendix C, and are similar to those in Xu et al. (2021). We start with a brief review of the properties of the unconstrained estimator $\tilde{\boldsymbol{y}}_s$. By the Taylor expansion, we can linearize the $\tilde{\boldsymbol{y}}_s$ as follows:

$$\tilde{\boldsymbol{y}}_s = \bar{\boldsymbol{y}}_U + \hat{\boldsymbol{y}}^{center} + O_p(n^{-1})$$

where

$$\hat{\boldsymbol{y}}^{center} = \left(\frac{1}{N_1} \sum_{k \in s_1} \frac{(y_k - \bar{y}_{U_1})}{\pi_k}, \cdots, \frac{1}{N_D} \sum_{k \in s_D} \frac{(y_k - \bar{y}_{U_D})}{\pi_k} \right)^\top .$$

The properties of $\tilde{\boldsymbol{y}}_s - \bar{\boldsymbol{y}}_U$ can be approximated by $\hat{\boldsymbol{y}}^{center}$ and we have that $\mathrm{E}(\hat{\boldsymbol{y}}^{center}) = 0$ and the variance of $\hat{\boldsymbol{y}}^{center}$ is $\boldsymbol{\Sigma}$, the asymptotic variance of $\tilde{\boldsymbol{y}}_s$. The ijth element of $\boldsymbol{\Sigma}$ is:

$$\boldsymbol{\Sigma}_{ij} = \frac{1}{N_i N_j} \sum_{k \in U_i} \sum_{l \in U_j} \Delta_{kl} \frac{(y_k - \bar{y}_{U_i})(y_l - \bar{y}_{U_j})}{\pi_k \pi_l}, \quad i, j = 1, 2, \cdots, D.$$

By the design normal assumption (A5) in the appendix, we have $\boldsymbol{\Sigma}^{-\frac{1}{2}} \hat{\boldsymbol{y}}^{center} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I})$, hence:

$$\boldsymbol{\Sigma}^{-\frac{1}{2}} (\tilde{\boldsymbol{y}}_s - \bar{\boldsymbol{y}}_U) = \boldsymbol{\Sigma}^{-\frac{1}{2}} \hat{\boldsymbol{y}}^{center} + o_p(1) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}).$$

To derive the asymptotic null distribution of $\hat{T}$, we first do the following transformation:

$$\begin{aligned} \hat{T} &= \frac{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0) - \min_{\boldsymbol{\theta}_1 \in \mathcal{C}} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)} \\ &= \frac{\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0)^\top (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0) - \min_{\hat{\boldsymbol{\theta}}_1 \in \hat{\mathcal{C}}} (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_1)^\top (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_1)}{\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0)^\top (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0)} \end{aligned}$$

where $\hat{\boldsymbol{A}} = \boldsymbol{A} \tilde{\boldsymbol{\Sigma}}^{\frac{1}{2}}$, $\hat{\boldsymbol{Z}}_s = \tilde{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \tilde{\boldsymbol{y}}_s$, $\hat{\boldsymbol{\theta}}_0 = \tilde{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}}_1 = \tilde{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \boldsymbol{\theta}_1$, $\hat{V} = \{\hat{\boldsymbol{\theta}}_0 : \hat{\boldsymbol{A}} \hat{\boldsymbol{\theta}}_0 = 0\}$ and $\hat{\mathcal{C}} = \{\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{A}} \hat{\boldsymbol{\theta}}_1 \geq 0\}$. Notice that $\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0)^\top (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0)$ is the squared length of the projection of $\hat{\boldsymbol{Z}}_s$ onto $\hat{V}^\perp$ and the projection of $\hat{\boldsymbol{Z}}_s$ onto $\hat{V}$ has the explicit expression $\hat{\boldsymbol{\theta}}_0^* = (\boldsymbol{I} - \hat{\boldsymbol{A}}^\top (\hat{\boldsymbol{A}} \hat{\boldsymbol{A}}^\top)^- \hat{\boldsymbol{A}}) \hat{\boldsymbol{Z}}_s$, where $(\hat{\boldsymbol{A}} \hat{\boldsymbol{A}}^\top)^-$ is the generalized inverse of $\hat{\boldsymbol{A}} \hat{\boldsymbol{A}}^\top$. Also, from (2.1) in Xu et al. (2021), the projection of $\hat{\boldsymbol{Z}}_s$ onto the cone $\hat{\mathcal{C}}$ can be expressed as:

$$\hat{\boldsymbol{\theta}}_1^* = \sum_J (\boldsymbol{I} - \hat{\boldsymbol{A}}_J^\top (\hat{\boldsymbol{A}}_J \hat{\boldsymbol{A}}_J^\top)^- \hat{\boldsymbol{A}}_J) \hat{\boldsymbol{Z}}_s \mathcal{I}_J(s) \tag{4.2}$$

where $J \subseteq \{1, \ldots, m\}$ such that the rows of $\hat{\boldsymbol{A}}_J$ form a linearly independent set and for each sample s, there is only one subset J for which $\mathcal{I}_J(s) = 1$. In addition, by the consistency of $\tilde{\boldsymbol{\Sigma}}$, we have the following:

$$\begin{aligned}
\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0)^\top (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0) &= (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0^*)^\top (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0^*) \\
&= (\hat{\boldsymbol{A}}^\top (\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^\top)^- \hat{\boldsymbol{A}}\hat{\boldsymbol{Z}}_s)^\top \hat{\boldsymbol{A}}^\top (\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^\top)^- \hat{\boldsymbol{A}}\hat{\boldsymbol{Z}}_s \\
&= \hat{\boldsymbol{Z}}_s^\top \hat{\boldsymbol{A}}^\top (\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^\top)^- \hat{\boldsymbol{A}}\hat{\boldsymbol{Z}}_s \\
&= \tilde{\boldsymbol{y}}_s^\top \boldsymbol{A}^\top (\boldsymbol{A}\tilde{\boldsymbol{\Sigma}}\boldsymbol{A}^\top)^- \boldsymbol{A}\tilde{\boldsymbol{y}}_s \\
&= \tilde{\boldsymbol{y}}_s^\top \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top)^- \boldsymbol{A}\tilde{\boldsymbol{y}}_s + o_p(1) \\
&= \min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0) + o_p(1). \qquad (4.3)
\end{aligned}$$

By (4.3) and Lemma 12 in the Appendix, $\hat{T}$ can be rewritten as:

$$\hat{T} = \frac{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0) - \min_{\boldsymbol{\theta}_1 \in \mathcal{C}} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)} + o_p(1)$$

The denominator in above expression must be bounded away from zero in probability, which is indeed the case because it can be shown that the $\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)$ has, asymptotically, $\chi^2(m)$ distribution under the null and design normal assumption; hence the denominator is bounded away from zero in probability.

Next, let $\tilde{\boldsymbol{Z}}_s = \boldsymbol{\Sigma}^{-\frac{1}{2}} \tilde{\boldsymbol{y}}_s$, $\boldsymbol{Z}_U = \boldsymbol{\Sigma}^{-\frac{1}{2}} \bar{\boldsymbol{y}}_U$, $\tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\theta}_0$, $\tilde{\boldsymbol{\theta}}_1 = \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\theta}_1$ and define $\tilde{V} = \{\tilde{\boldsymbol{\theta}} : \tilde{\boldsymbol{A}}\tilde{\boldsymbol{\theta}} = 0\}$, $\tilde{\mathcal{C}} = \{\tilde{\boldsymbol{\theta}} : \tilde{\boldsymbol{A}}\tilde{\boldsymbol{\theta}} \geq 0\}$, where $\tilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Sigma}^{\frac{1}{2}}$. Then

$$\begin{aligned}
\hat{T} &= \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} ||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_0||^2 - \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{\mathcal{C}}} ||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_1||^2}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} ||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_0||^2} + o_p(1) \\
&= \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} ||\tilde{\boldsymbol{Z}}_s - \boldsymbol{Z}_U + \boldsymbol{Z}_U - \tilde{\boldsymbol{\theta}}_0||^2 - \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{\mathcal{C}}} ||\tilde{\boldsymbol{Z}}_s - \boldsymbol{Z}_U + \boldsymbol{Z}_U - \tilde{\boldsymbol{\theta}}_1||^2}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} ||\tilde{\boldsymbol{Z}}_s - \boldsymbol{Z}_U + \boldsymbol{Z}_U - \tilde{\boldsymbol{\theta}}_0||^2} + o_p(1).
\end{aligned}$$

Let $\boldsymbol{Z}^{center} = \tilde{\boldsymbol{Z}}_s - \boldsymbol{Z}_U$, and recall that under H_0, $\boldsymbol{Z}_U \in \tilde{V}$, so that, in the above expression, minimizing over $\tilde{\boldsymbol{\theta}}_0$ is equivalent to minimizing over $-\boldsymbol{Z}_U + \tilde{\boldsymbol{\theta}}_0$, and similarly for minimizing over

$\tilde{\boldsymbol{\theta}}_1$. Then our test statistic may be expressed as

$$\hat{T} = \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} ||\boldsymbol{Z}^{center} - \tilde{\boldsymbol{\theta}}_0||^2 - \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{\mathcal{C}}} ||\boldsymbol{Z}^{center} - \tilde{\boldsymbol{\theta}}_1||^2}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} ||\boldsymbol{Z}^{center} - \tilde{\boldsymbol{\theta}}_0||^2} + o_p(1).$$

Let $\boldsymbol{Z} \sim$N$(\boldsymbol{0}, \boldsymbol{I})$, and define

$$T = \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} ||\boldsymbol{Z} - \tilde{\boldsymbol{\theta}}_0||^2 - \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{\mathcal{C}}} ||\boldsymbol{Z} - \tilde{\boldsymbol{\theta}}_1||^2}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} ||\boldsymbol{Z} - \tilde{\boldsymbol{\theta}}_0||^2}$$

The random variable T has been shown to be distributed as a mixture of beta random variables under H_0, and the mixing distribution can be found (to within a desired precision) via simulation. Specifically,

$$\text{Pr}(T \leq c) = \sum_{d=0}^{m} \text{Pr}\left\{ Be\left(\frac{m-d}{2}, \frac{d}{2}\right) \leq c \right\} p_d,$$

where $p_0, \ldots, p_D$ are approximated through simulations, and $Be(\alpha, \beta)$ represents a Beta random variable with parameters α and β, respectively. By convention, $Be(0, \beta) = 0$ and $Be(\alpha, 0) = 1$.

If the irreducible constraint matrix $\boldsymbol{A}$ is not full row rank, say $m > D$, then the above result still holds by substituting m_1 for m and $d = 0, 1, \cdots, m_1$, where m_1 is denoted to be the dimension of the space spanned by the rows of the constraint matrix $\boldsymbol{A}$.

Finally, we have $\hat{T} \xrightarrow{\mathcal{D}} T$. This follows from the Lipschitz continuity of the projection of $\boldsymbol{Z}$ onto a convex cone; that is, if $\hat{\boldsymbol{\theta}}$ is the projection of $\boldsymbol{Z}$ onto the cone $\mathcal{C}$, then $\hat{\boldsymbol{\theta}}$ is a continuous function of $\boldsymbol{Z}$; see Proposition 1 and its proof in Meyer and Woodroofe (2000).

The mixture probabilities are approximated as follows:

(1) Generate $\boldsymbol{Z}$ from standard multivariate normal distribution $N(\boldsymbol{0}, \boldsymbol{I})$.

(2) Project the generated $\boldsymbol{Z}$ onto the convex cone $\hat{\mathcal{C}} = \{\boldsymbol{\theta} : \hat{\boldsymbol{A}}\boldsymbol{\theta} \geq 0\}$ to obtain the J set, where $\hat{\boldsymbol{A}} = \boldsymbol{A}\tilde{\boldsymbol{\Sigma}}^{\frac{1}{2}}$. Specifically, let $\hat{\boldsymbol{\theta}}$ be the projection of $\boldsymbol{Z}$ onto the $\hat{\mathcal{C}}$, then $J = \{j : \hat{\boldsymbol{A}}_j\hat{\boldsymbol{\theta}} = 0\}$.

    The R package `coneproj` (Liao and Meyer (2014)) finds $\hat{\boldsymbol{\theta}}$ given the generated $\boldsymbol{Z}$ and $\hat{\boldsymbol{A}}$, and also returns the set of binding constraints J.

(3) Repeat the previous steps B times (say $B = 1000$).

(4) Estimate p_d by the proportion of times that the set J has d elements, $d = 0, 1, \cdots, m$. When the matrix $\boldsymbol{A}$ has more constraints than dimensions, then, the set J has at most m_1 elements. In fact, the cone projection routine in `coneproj` can always find a minimal unique J set.

4.3.1 The Properties of the Test

In this section, we prove consistency and monotonicity of the power function of this test. Under the alternative, if $\hat{T}$ is a good test statistic, then we would expect that the probability of rejecting null increase to one as n increases.

Theorem 10. *Let α be the test size and c_α be the corresponding critical value of the test. Then, the power of the test converges to 1 under the alternative. That is:*

$$P(\hat{T} > c_\alpha | \bar{\boldsymbol{y}}_U \in \mathcal{C} \backslash V) \to 1, \quad \textit{as } n \to \infty.$$

Proof. Since $\hat{T} = 1 - \frac{\min_{\boldsymbol{\theta}_1 \in \mathcal{C}} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)}$, it suffices to show that:

$$\frac{\min_{\boldsymbol{\theta}_1 \in \mathcal{C}} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)} = o_p(1)$$

under the the alternative. For the numerator, we have

$$\min_{\boldsymbol{\theta}_1 \in \mathcal{C}} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1) \leq (\tilde{\boldsymbol{y}}_s - \bar{\boldsymbol{y}}_U)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \bar{\boldsymbol{y}}_U) = O_p\left(\frac{1}{\sqrt{n}}\right) O_p(n) O_p\left(\frac{1}{\sqrt{n}}\right) = O_p(1)$$

where we use the fact that $\tilde{\boldsymbol{y}}_s - \bar{\boldsymbol{y}}_U = O_p(n^{-\frac{1}{2}})$ and $\tilde{\boldsymbol{\Sigma}} = O_p(n^{-1})$ element-wise. For the denominator, we have:

$$\begin{aligned}
&\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0) = \min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} ||\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_0||^2 \\
=&\left[\hat{\boldsymbol{Z}}_s - (\boldsymbol{I} - \hat{\boldsymbol{A}}^\top (\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^\top)^- \hat{\boldsymbol{A}})\hat{\boldsymbol{Z}}_s\right]^\top \left[\hat{\boldsymbol{Z}}_s - (\boldsymbol{I} - \hat{\boldsymbol{A}}^\top (\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^\top)^- \hat{\boldsymbol{A}})\hat{\boldsymbol{Z}}_s\right] \\
=&\hat{\boldsymbol{Z}}_s^\top \hat{\boldsymbol{A}}^\top (\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^\top)^- \hat{\boldsymbol{A}} \hat{\boldsymbol{Z}}_s \\
=&\tilde{\boldsymbol{y}}_s^\top \boldsymbol{A}^\top (\boldsymbol{A}\tilde{\boldsymbol{\Sigma}}\boldsymbol{A}^\top)^- \boldsymbol{A}\tilde{\boldsymbol{y}}_s
\end{aligned}$$

Hence, we have:

$$\begin{aligned}
\frac{\min_{\boldsymbol{\theta}_1 \in \mathcal{C}} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)} &= \frac{\min_{\boldsymbol{\theta}_1 \in \mathcal{C}} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top (n\tilde{\boldsymbol{\Sigma}})^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top (n\tilde{\boldsymbol{\Sigma}})^{-1} (\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)} \\
&= O_p(n^{-1}) \frac{1}{\tilde{\boldsymbol{y}}_s^\top \boldsymbol{A}^\top (\boldsymbol{A} n \tilde{\boldsymbol{\Sigma}} \boldsymbol{A}^\top)^- \boldsymbol{A}\tilde{\boldsymbol{y}}_s} \\
&= O_p(n^{-1}) \frac{1}{\bar{\boldsymbol{y}}_U^\top \boldsymbol{A}^\top (\boldsymbol{A} n \boldsymbol{\Sigma} \boldsymbol{A}^\top)^- \boldsymbol{A}\bar{\boldsymbol{y}}_U + o_p(1)} \\
&= O_p(n^{-1}) O_p(1) = o_p(1)
\end{aligned}$$

because $\tilde{\boldsymbol{y}}_s$ and $\tilde{\boldsymbol{\Sigma}}$ are consistent for $\bar{\boldsymbol{y}}_U$ and $\boldsymbol{\Sigma}$ respectively. Therefore, under the alternative, $\hat{T}$ goes to 1 asymptotically. $\square$

The following is a result on the monotonicity of the power for the test.

Theorem 11. *Suppose that $-\boldsymbol{\Omega} \subseteq \mathcal{C}^o$, where $\boldsymbol{\Omega} = \mathcal{C} \cap V^\perp$, then for any vector $\boldsymbol{\theta} \in \boldsymbol{\Omega}$, the value of the test statistic associated with the data vector $\tilde{\boldsymbol{y}}_s + \boldsymbol{\theta}$ is bigger than the value for $\tilde{\boldsymbol{y}}_s$. Hence, the power for the test with alternative $H_1^* : \boldsymbol{A}\boldsymbol{\theta}_U \geq 0$ is larger than the power of the test with alternative $H_1 : \boldsymbol{A}\bar{\boldsymbol{y}}_U \geq 0$, where $\boldsymbol{\theta}_U = \bar{\boldsymbol{y}}_U + \boldsymbol{\theta}$.*

Proof. Denote:

$$\begin{aligned}
T^{(1)} &= \frac{\min_{\boldsymbol{\theta}_0 \in V}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0) - \min_{\boldsymbol{\theta}_1 \in \mathcal{C}}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)^\top \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{y}}_s - \boldsymbol{\theta}_0)} \\
&= \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}}(\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_0)^\top (\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_0) - \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{\mathcal{C}}}(\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_1)^\top (\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_1)}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}}(\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_0)^\top (\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}_0)} \\
&= 1 - \frac{||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}^{(1)}||^2}{||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{v}}||^2}
\end{aligned}$$

where $\tilde{\boldsymbol{Z}}_s$, $\tilde{\boldsymbol{\theta}}_0$, $\tilde{\boldsymbol{\theta}}_1$, $\tilde{V}$ and $\tilde{\mathcal{C}}$ were defined previously, $\tilde{\boldsymbol{v}}$ is the projection of $\tilde{\boldsymbol{Z}}_s$ onto $\tilde{\boldsymbol{V}}$ and $\tilde{\boldsymbol{\theta}}^{(1)}$ is the projection of $\tilde{\boldsymbol{Z}}_s$ onto $\tilde{\mathcal{C}}$. Similarly, we can define:

$$\begin{aligned}
T^{(2)} &= 1 - \frac{\min_{\boldsymbol{\theta}_1 \in \mathcal{C}}((\tilde{\boldsymbol{y}}_s + \boldsymbol{\theta}) - \boldsymbol{\theta}_1)^\top \boldsymbol{\Sigma}^{-1}((\tilde{\boldsymbol{y}}_s + \boldsymbol{\theta}) - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V}((\tilde{\boldsymbol{y}}_s + \boldsymbol{\theta}) - \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}^{-1}((\tilde{\boldsymbol{y}}_s + \boldsymbol{\theta}) - \boldsymbol{\theta}_0)} \\
&= 1 - \frac{\min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{\mathcal{C}}}((\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}}) - \tilde{\boldsymbol{\theta}}_1)^\top ((\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}}) - \tilde{\boldsymbol{\theta}}_1)}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}}((\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}}) - \tilde{\boldsymbol{\theta}}_0)^\top ((\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}}) - \tilde{\boldsymbol{\theta}}_0)} \\
&= 1 - \frac{||\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^{(2)}||^2}{||\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{v}}||^2}
\end{aligned}$$

where $\tilde{\boldsymbol{\theta}} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}^{(2)}$ is the projection of $\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}}$ onto $\tilde{\mathcal{C}}$. Because $\boldsymbol{\theta}$ is orthogonal to V, so $\tilde{\boldsymbol{\theta}}$ is also orthogonal to $\tilde{V}$. Thus, the projection of $\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}}$ onto $\tilde{V}$ is the same as the projection of $\tilde{\boldsymbol{Z}}_s$ onto $\tilde{V}$, which is $\tilde{\boldsymbol{v}}$. Further, let $\tilde{\boldsymbol{\rho}}^{(1)}$ be the projection of $\tilde{\boldsymbol{Z}}_s$ onto $\tilde{\mathcal{C}}^o$ and $\tilde{\boldsymbol{\rho}}^{(2)}$ be the projection of $\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}}$ onto $\tilde{\mathcal{C}}^o$. Then,

$$T^{(2)} - T^{(1)} = \frac{||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}^{(1)}||^2}{||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{v}}||^2} - \frac{||\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^{(2)}||^2}{||\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{v}}||^2}$$

To show $T^{(2)} - T^{(1)} \geq 0$, it is suffices to show the following is non-negative:

$$
\begin{aligned}
&||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}^{(1)}||^2||\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{v}}||^2 - ||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{v}}||^2||\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^{(2)}||^2 \\
=&||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\theta}}^{(1)}||^2||\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^{(2)} + \tilde{\boldsymbol{\theta}}^{(2)} - \tilde{\boldsymbol{v}}||^2 - ||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{v}}||^2||\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^{(2)}||^2 \\
=&||\tilde{\boldsymbol{\rho}}^{(1)}||^2||\tilde{\boldsymbol{\rho}}^{(2)} + \tilde{\boldsymbol{\theta}}^{(2)} - \tilde{\boldsymbol{v}}||^2 - ||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{v}}||^2||\tilde{\boldsymbol{\rho}}^{(2)}||^2 \\
=&||\tilde{\boldsymbol{\rho}}^{(1)}||^2||\tilde{\boldsymbol{\rho}}^{(2)}||^2 + ||\tilde{\boldsymbol{\rho}}^{(1)}||^2||\tilde{\boldsymbol{\theta}}^{(2)} - \tilde{\boldsymbol{v}}||^2 - ||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{v}}||^2||\tilde{\boldsymbol{\rho}}^{(2)}||^2 \\
=&||\tilde{\boldsymbol{\rho}}^{(1)}||^2\left(||\tilde{\boldsymbol{\theta}}^{(2)}||^2 - ||\tilde{\boldsymbol{v}}||^2\right) - ||\tilde{\boldsymbol{\rho}}^{(2)}||^2\left(||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{v}}||^2 - ||\tilde{\boldsymbol{\rho}}^{(1)}||^2\right) \\
=&||\tilde{\boldsymbol{\rho}}^{(1)}||^2\left(||\tilde{\boldsymbol{\theta}}^{(2)}||^2 - ||\tilde{\boldsymbol{v}}||^2 - \left(||\tilde{\boldsymbol{Z}}_s||^2 - ||\tilde{\boldsymbol{v}}||^2 - ||\tilde{\boldsymbol{\rho}}^{(1)}||^2\right)\right) \\
=&||\tilde{\boldsymbol{\rho}}^{(1)}||^2\left(||\tilde{\boldsymbol{\theta}}^{(2)}||^2 - ||\tilde{\boldsymbol{v}}||^2 - \left(||\tilde{\boldsymbol{\theta}}^{(1)}||^2 - ||\tilde{\boldsymbol{v}}||^2\right)\right) \\
=&||\tilde{\boldsymbol{\rho}}^{(1)}||^2\left(||\tilde{\boldsymbol{\theta}}^{(2)}||^2 - ||\tilde{\boldsymbol{\theta}}^{(1)}||^2\right) \\
=&||\tilde{\boldsymbol{\rho}}^{(1)}||^2\left(||\tilde{\boldsymbol{Z}}_s - (\tilde{\boldsymbol{\rho}}^{(2)} - \tilde{\boldsymbol{\theta}})||^2 - ||\tilde{\boldsymbol{\theta}}^{(1)}||^2\right) \\
\geq&||\tilde{\boldsymbol{\rho}}^{(1)}||^2\left(||\tilde{\boldsymbol{Z}}_s - \tilde{\boldsymbol{\rho}}^{(1)}||^2 - ||\tilde{\boldsymbol{\theta}}^{(1)}||^2\right) \\
=&||\tilde{\boldsymbol{\rho}}^{(1)}||^2\left(||\tilde{\boldsymbol{\theta}}^{(1)}||^2 - ||\tilde{\boldsymbol{\theta}}^{(1)}||^2\right) = 0
\end{aligned}
$$

Here, we use the fact that $\tilde{\boldsymbol{\rho}}^{(1)} = \tilde{\boldsymbol{\rho}}^{(2)}$, because $\tilde{\boldsymbol{\theta}} \in \tilde{\boldsymbol{\Omega}}$ is orthogonal to $\tilde{\mathcal{C}}^o$ and thus the projection of $\tilde{\boldsymbol{Z}}_s + \tilde{\boldsymbol{\theta}}$ onto $\tilde{\mathcal{C}}^o$ is the same as the projection of $\tilde{\boldsymbol{Z}}_s$ onto $\tilde{\mathcal{C}}^o$, where $\tilde{\boldsymbol{\Omega}} = \tilde{\mathcal{C}} \cap \tilde{V}^{\perp}$. Also, when $-\boldsymbol{\theta} \in \mathcal{C}^o$, then $-\tilde{\boldsymbol{\theta}} \in \tilde{\mathcal{C}}^o$ by transformation. Further, the vector $\tilde{\boldsymbol{\rho}}^{(2)} - \tilde{\boldsymbol{\theta}}$ will be in $\tilde{\mathcal{C}}^o$ and must be farther from $\tilde{\boldsymbol{Z}}_s$ than $\tilde{\boldsymbol{\rho}}^{(1)}$.

Let c_α be the critical value for this one-sided test for a specific significance level α. Then we must have:

$$P(T^{(1)} > c_\alpha | H_1) \leq P(T^{(2)} > c_\alpha | H_1^*)$$

The power of the test increases. □

The above result essentially concludes that as the effect size increases, the power of the test increases. In other words, larger effects are easier to detect reliably. Also, the proposed test statistic $\hat{T}$ will inherit above property asymptotically due to the consistency of the covariance estimator.

4.4 Simulation Studies

The simulations involve one or two dimensional grids, with several constraints and population domain means. We present the results in table form from three scenarios: for each, we record the proportions of times the null is rejected in various cases, with different sample sizes, significance levels and the variances for generating the study variables. In each case, we generate a population of size N, then we draw 10,000 samples from the population according to a sampling design. For each sample, we compute the test statistic value and compare it with the critical values under different significance levels, where the critical values are obtained from the asymptotic null distribution of the test statistics. Further, we compare the power of this one sided test with that of ANOVA F test using the unconstrained alternative. That is,

$$H_0 : \boldsymbol{A}\bar{\boldsymbol{y}}_U = \boldsymbol{0} \quad \text{versus} \quad H_2 : \boldsymbol{A}\bar{\boldsymbol{y}}_U \neq \boldsymbol{0}$$

Here, we use `svyglm` function in `survey` package to fit the ANOVA model and compute the P-values of the ANOVA F test by applying the `anova` function in `survey` package.

4.4.1 Monotonicity in One Variable

As in Xu et al. (2021) and Oliva-Avilés et al. (2020), the limiting domain means for generating the study variables are given by the functions as follows:

$$\mu_d^{(0)} \equiv 1, \quad \text{for } d = 1, 2, \cdots, D.$$

$$\mu_d^{(1)} = \frac{\exp(12d/D - 6)}{3.5(1 + \exp(12d/D - 6))}, \quad \text{for } d = 1, 2, \cdots, D.$$

$$\mu_d^{(2)} = \frac{\exp(12d/D - 6)}{2.5(1 + \exp(12d/D - 6))}, \quad \text{for } d = 1, 2, \cdots, D.$$

where $D = 12$ is the number of domains. The study variables $y_1, \ldots, y_N$ are generated by adding independent and identically distributed $N(0, \sigma_i^2)$ $(i = 1, 2)$ errors to the μ_d values from above three functions, respectively, with $\sigma_1 = 1$ and $\sigma_2 = 1.5$. We compare the test size and power for the test of constant versus increasing domain means, with the standard ANOVA test of constant versus non-constant domain means.

We draw the samples from a stratified simple random sampling design without replacement, with $H = 4$ strata that cut across the D domains. The strata are determined using an auxiliary variable z, which is correlated with study variable y. The values of z are created by adding i.i.d. standard normal errors to (d/D). By ranking the values of z, we can create 4 blocks of N/H elements. Then, the stratum membership of the population element is determined by the corresponding ranked z, Finally, the population sizes are set to be $N = 9600$, $N = 19200$, $N = 57600$ and $N = 76800$, respectively, with domain population size $N_d = N/D$. The total sample sizes $n = 200$, $n = 400$, $n = 1200$ and $n = 1600$ are assigned to each strata with sample size $(25, 50, 50, 75)$, $(50, 100, 100, 150)$, $(150, 300, 300, 450)$, $(200, 400, 400, 600)$ in each strata, respectively.

The results in Table 4.1 show that the test size for the proposed one-sided test is closer to the target, while the two-sided test size is somewhat inflated even for the larger sample sizes. For the simulations where the alternative hypothesis is true, the one-sided test has substantially higher power.

4.4.2 Block Monotonic in One Variable

In “block monotonic" ordering case, we assume the population means are ordered among blocks, but there is no ordering imposed within the blocks. Specifically, we organize the limiting domain means in four blocks of three domains as following:

$$\boldsymbol{\mu}^{(0)} = \left(\begin{array}{ccc|ccc|ccc|ccc} .05 & .05 & .05 & .05 & .05 & .05 & .05 & .05 & .05 & .05 & .05 & .05 \end{array} \right)$$

$$\boldsymbol{\mu}^{(1)} = \left(\begin{array}{ccc|ccc|ccc|ccc} -.06 & 0 & .06 & .12 & .06 & .18 & .18 & .24 & .30 & .30 & .36 & .30 \end{array} \right)$$

$$\boldsymbol{\mu}^{(2)} = \left(\begin{array}{ccc|ccc|ccc|ccc} -.08 & 0 & .08 & .16 & .08 & .24 & .24 & .32 & .40 & .40 & .48 & .40 \end{array} \right)$$

Table 4.1: Monotonicity in one variable: the proportions of times null is rejected under various settings and power comparison between the constrained one sided test and the unconstrained test

One sided test										
σ	n	$\alpha_1 = 0.1$			$\alpha_2 = 0.05$			$\alpha_3 = 0.01$		
		$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$
$\sigma = 1$	n=200	0.0996	0.4689	0.6686	0.0533	0.3218	0.5055	0.0134	0.1194	0.2230
	n=400	0.0840	0.6352	0.8529	0.0403	0.4780	0.7268	0.0085	0.2028	0.4054
	n=1200	0.1039	0.9657	0.9986	0.0537	0.9027	0.9941	0.0121	0.6444	0.9133
	n=1600	0.0981	0.9867	0.9999	0.0489	0.9550	0.9988	0.0110	0.7533	0.9654
$\sigma = 1.5$	n=200	0.0994	0.3128	0.4370	0.0528	0.2008	0.2938	0.0133	0.0625	0.1056
	n=400	0.0839	0.4101	0.5946	0.0402	0.2740	0.4338	0.0084	0.0873	0.1770
	n=1200	0.1037	0.7838	0.9461	0.0532	0.6327	0.8679	0.0120	0.3142	0.5773
	n=1600	0.0980	0.8544	0.9751	0.0488	0.7253	0.9334	0.0109	0.3900	0.6928
ANOVA F test										
σ	n	$\alpha_1 = 0.1$			$\alpha_2 = 0.05$			$\alpha_3 = 0.01$		
		$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$
$\sigma = 1$	n=200	0.1412	0.2677	0.4017	0.0746	0.1627	0.2685	0.0147	0.0457	0.0973
	n=400	0.1280	0.3618	0.6034	0.0658	0.2385	0.4627	0.0147	0.0835	0.2259
	n=1200	0.1123	0.8139	0.9854	0.0590	0.7121	0.9694	0.0117	0.4736	0.8943
	n=1600	0.1111	0.9253	0.9986	0.0576	0.8633	0.9964	0.0126	0.6868	0.9814
$\sigma = 1.5$	n=200	0.1412	0.1909	0.2502	0.0746	0.1087	0.1495	0.0147	0.0261	0.0408
	n=400	0.1280	0.2195	0.3278	0.0658	0.1296	0.2094	0.0147	0.0313	0.0661
	n=1200	0.1123	0.4670	0.7538	0.0590	0.3320	0.6361	0.0117	0.1397	0.3902
	n=1600	0.1111	0.5947	0.8795	0.0576	0.4602	0.8014	0.0126	0.2367	0.5932

where the blocks are separated by the vertical lines. Hence, under the alternative, we expect the population mean for each of the domains in block i would be at least as large as those in block $i-1$, for $i = 2, 3, 4$. Also, the effect size of $\bar{\boldsymbol{y}}_U^{(2)}$ generated from $\boldsymbol{\mu}^{(2)}$ would be larger than that of $\bar{\boldsymbol{y}}_U^{(1)}$ from $\boldsymbol{\mu}^{(1)}$. We use exactly the same stratified simple random sampling design as in the previous example.

The results in Table 4.2 show again that one sided test has substantially higher power for simulations where the alternative is true, and for simulations under the null hypothesis, the test size is approximately correct for the one-sided test and the two-sided ANOVA test has inflated test size.

4.4.3 Monotonicity in Two Variables

Here we take into consideration a grid of domains, which represent two variables. In particular, we set the limiting domain means as follows:

Table 4.2: Block monotonicity in one variable: the proportions of times null is rejected under various settings and power comparison between the constrained one sided test and the unconstrained test

One sided test										
σ	n	$\alpha_1 = 0.1$			$\alpha_2 = 0.05$			$\alpha_3 = 0.01$		
		$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$
$\sigma = 1$	n=200	0.1013	0.5114	0.6795	0.0568	0.3590	0.5216	0.0119	0.1397	0.2391
	n=400	0.1036	0.7368	0.8856	0.0534	0.5838	0.7878	0.0109	0.2840	0.4722
	n=1200	0.0964	0.9718	0.9978	0.0487	0.9224	0.9880	0.0089	0.6671	0.8801
	n=1600	0.0976	0.9877	0.9998	0.0492	0.9635	0.9958	0.0098	0.7668	0.9339
$\sigma = 1.5$	n=200	0.1014	0.3421	0.4535	0.0567	0.2191	0.3124	0.0117	0.0731	0.1144
	n=400	0.1031	0.4992	0.6616	0.0534	0.3544	0.5028	0.0109	0.1335	0.2235
	n=1200	0.0965	0.8187	0.9422	0.0485	0.6794	0.8672	0.0091	0.3474	0.5661
	n=1600	0.0974	0.8830	0.9743	0.0497	0.7652	0.9232	0.0099	0.4367	0.6746
ANOVA F test										
σ	n	$\alpha_1 = 0.1$			$\alpha_2 = 0.05$			$\alpha_3 = 0.01$		
		$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$
$\sigma = 1$	n=200	0.1412	0.2941	0.4368	0.0746	0.1847	0.2951	0.0147	0.0551	0.1155
	n=400	0.1280	0.4220	0.6556	0.0658	0.2912	0.5231	0.0147	0.1123	0.2712
	n=1200	0.1123	0.8940	0.9921	0.0590	0.8177	0.9840	0.0117	0.6099	0.9363
	n=1600	0.1111	0.9678	0.9995	0.0576	0.9293	0.9986	0.0126	0.8094	0.9911
$\sigma = 1.5$	n=200	0.1412	0.2052	0.2611	0.0746	0.1173	0.1583	0.0147	0.0281	0.0431
	n=400	0.1280	0.2445	0.3543	0.0658	0.1457	0.2333	0.0147	0.0389	0.0787
	n=1200	0.1123	0.5399	0.8012	0.0590	0.4099	0.6932	0.0117	0.1926	0.4549
	n=1600	0.1111	0.6799	0.9091	0.0576	0.5539	0.8468	0.0126	0.3153	0.6589

$$\boldsymbol{\mu}^{(0)} = \begin{pmatrix} .01 & .01 & .01 & .01 & .01 \\ .02 & .02 & .02 & .02 & .02 \\ .03 & .03 & .03 & .03 & .03 \\ .04 & .04 & .04 & .04 & .04 \end{pmatrix},$$

$$\boldsymbol{\mu}^{(1)} = \begin{pmatrix} 0 & .04 & .16 & .24 & .28 \\ .04 & .08 & .20 & .32 & .40 \\ .04 & .12 & .12 & .20 & .28 \\ .04 & .04 & .12 & .24 & .28 \end{pmatrix},$$

$$\boldsymbol{\mu}^{(2)} = \begin{pmatrix} 0 & .05 & .20 & .30 & .35 \\ .05 & .10 & .25 & .40 & .50 \\ .05 & .15 & .15 & .25 & .35 \\ .05 & .05 & .15 & .30 & .35 \end{pmatrix}.$$

here there are five levels of one variable and four of the other, so there are $D = 20$ domains. By the setting, the μ values are non-decreasing in one variable and there is no constraint on the other variable. It is quite useful in practice. For instance, in National Compensation Survey, the population domain means of the salary are expected to be monotone in job level, but there might be no shape restriction on job type. Further, we expect $\bar{\boldsymbol{y}}_U^{(2)}$ from $\boldsymbol{\mu}^{(2)}$ will have larger effect size than that of $\bar{\boldsymbol{y}}_U^{(1)}$ from $\boldsymbol{\mu}^{(1)}$.

The sampling mechanism and the way we generate the study variable y are the same as that in one dimensional case. However, because there are more number of domains in this case, we set the sample size to be $n = 400$, $n = 800$, $n = 1200$ and $n = 2000$, respectively, corresponding to the population size $N = 8000$, $N = 16000$, $N = 24000$ and $N = 40000$, where the sample sizes are divided among the strata as $(50, 100, 100, 150)$, $(100, 200, 200, 300)$, $(150, 300, 300, 450)$ and $(250, 500, 500, 750)$, respectively.

The simulation results in Table 4.3 demonstrate similar properties as those in the previous scenarios: the tests have higher power as sample size gets larger and the effect size of the population domain means is larger.

4.5 Application to NSCG 2019 Data

To demonstrate the utility of the proposed one sided test procedure in real survey data, we consider the 2019 National Survey of College Graduates (NSCG), which is conducted by the U.S. Census Bureau. The NSCG provides data on the characteristics of the nation's college graduates, with a focus on those in the science and engineering workforce. The data and relevant documentation are available to the public on the NSF website (https://www.nsf.gov/statistics/srvygrads/).

Table 4.3: Monotonicity in two variables: the proportions of times null is rejected under various settings and power comparison between the constrained one sided test and the unconstrained test

One sided test										
σ	n	$\alpha_1 = 0.1$			$\alpha_2 = 0.05$			$\alpha_3 = 0.01$		
		$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$
$\sigma = 1$	n=400	0.1770	0.7738	0.8755	0.1000	0.6415	0.7757	0.0255	0.3460	0.4907
	n=800	0.1203	0.8732	0.9576	0.0590	0.7677	0.8975	0.0129	0.4706	0.6598
	n=1200	0.1097	0.9571	0.9921	0.0562	0.8972	0.9762	0.0102	0.6556	0.8523
	n=2000	0.1093	0.9929	0.9994	0.0558	0.9794	0.9975	0.0103	0.8661	0.9700
$\sigma = 1.5$	n=400	0.1778	0.5837	0.6840	0.1006	0.4301	0.5382	0.0255	0.1844	0.2586
	n=800	0.1210	0.6512	0.7783	0.0594	0.4967	0.6399	0.0133	0.2257	0.3421
	n=1200	0.1098	0.7701	0.8908	0.0565	0.6247	0.7881	0.0100	0.3235	0.4909
	n=2000	0.1089	0.9019	0.9725	0.0560	0.8040	0.9292	0.0103	0.5150	0.7236
ANOVA F test										
σ	n	$\alpha_1 = 0.1$			$\alpha_2 = 0.05$			$\alpha_3 = 0.01$		
		$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$
$\sigma = 1$	n=400	0.1584	0.4337	0.5642	0.0828	0.3005	0.4255	0.0184	0.1075	0.1886
	n=800	0.1338	0.5817	0.7748	0.0703	0.4407	0.6600	0.0154	0.2165	0.4058
	n=1200	0.1273	0.7028	0.8922	0.0662	0.5773	0.8149	0.0140	0.3224	0.6055
	n=2000	0.1289	0.9174	0.9912	0.0697	0.8577	0.9789	0.0149	0.6664	0.9198
$\sigma = 1.5$	n=400	0.1584	0.2899	0.3578	0.0828	0.1759	0.2285	0.0184	0.0510	0.0732
	n=800	0.1338	0.3283	0.4443	0.0703	0.2138	0.3133	0.0154	0.0717	0.1274
	n=1200	0.1273	0.3803	0.5358	0.0662	0.2606	0.4009	0.0140	0.1014	0.1883
	n=2000	0.1289	0.5759	0.7811	0.0697	0.4434	0.6683	0.0149	0.2148	0.4215

We choose the annual salary as the study variable (denoted by SALARY in the dataset). In order to prevent the impact of the extreme annual salary, we only consider those who reported an annual salary between \$30,000 and \$900,000. Also, as the annual salary variable distribution is skewed, a log transformation is implemented. Four predictor variables are considered:

- Field (denoted by NDGMEMG in the dataset): This nominal variable defines the field of study for the highest degree. There are six levels: (1) Computer and mathematical sciences; (2) Biological, agricultural and environmental life sciences; (3) Physical and related sciences; (4) Social and related sciences; (5) Engineering; (6) Other.

- Father's education level (denoted by EDDAD in the dataset): This ordinal variable denotes the highest level of education completed by the respondents' father (or male guardian). The six levels are: (1) Less than high school completed; (2) High school diploma or equivalent; (3) Some college, vocational, or trade school (including 2-year degrees); (4) Bachelors degree (e.g. BS, BA, AB); (5) Masters degree (e.g. MS, MA, MBA); (6) Professional degree (e.g. JD, LLB, MD, DDS, etc.) and Doctorate (e.g. PhD, DSc, EdD, etc.).

- Academic year of award for the highest degree (denoted by HDACYR): This variable gives information about which year each respondent was awarded for their highest degree.
- Highest degree type (denoted by DGRDG): This ordinal variable denotes the highest degree type the respondents earned. The four levels are: (1) Bachelor's; (2) Master's; (3) Doctorate; (4) Professional.

Suppose interest is in the question: for wage-earners whose highest degree is a bachelor's, does the father's education level influence the salary, when controlling for field of study and time since degree? To answer this, we perform separate tests for cohorts in years that the degree was attained, as in Table 4.4. Within each cohort, there are 36 domains, with six levels each of field and father's education level. We test the null hypothesis that the salary is constant over father's education level, within each field, against the alternative that the salary is increasing in father's education level. We compare the p-values for this test with constrained alternative to the ANOVA test with unconstrained alternative. The `svyglm` function in `survey` package is used for the unconstrained alternative, and the F test by applying the `anova` function in `survey` package gives the p-value. The results of the tests for five recent cohorts are in Table 4.4.

Table 4.4: p-values for the null hypothesis that salary is constant in father's education level, controlling for field of study.

year	2007-2009	2010-2011	2012-2013	2014-2015	2016-2017
one-sided mixture beta test	0.0037	0.0119	0.0004	0.0204	0.0018
ANOVA F test	0.0601	0.3601	0.0236	0.2223	0.0655

For each cohort, the p-value for the one-sided test is below $.05$, indicating that salaries increase significantly with father's education level, consistently across years. In contrast, the p-value for the two-sided test is consistently larger, and does not indicate a significant trend for some of the cohorts, and for other years the test results could be considered "borderline." Using the *a priori* knowledge that if father education level affects salary, it must be a positive effect, helps increase the power to see the trend.

4.6 Discussion

In this paper, we developed a testing procedure for testing the linear inequality restrictions of the population domain means within the survey context. Under the design normal assumption of the survey domain means, the proposed test statistic $\hat{T}$ has the asymptotic mixture beta densities, where the mixing probabilities (or the weights) can be easily computed via simulations. The test statistic readily allows computation, with the covariance estimator $\tilde{\boldsymbol{\Sigma}}$ and the unconstrained estimator $\tilde{\boldsymbol{y}}_s$ obtained from the `survey` package in `R` and the constrained least square projection obtained by using the `coneA` function in `coneproj` package. In theory, we showed that the power of the test tends to be one as the sample size increases. Also, larger effect size of the population domain means can boost the power of the test. The simulation studies confirm the properties of the proposed test and find that the test is an exact test in moderate to large-sized samples.

The implementation of the test in the `csurvey` package borrows from the `survey` package. For example, suppose we have a grid of domains in two variables `x1` and `x2` and study variable `y`. The survey design is specified with the `svydesign` command in the `survey` package, and the design object `ds` is used in the implementation of the test. The p-value for the test of constant versus increasing domain means along the `x1` variable, without constraining the domain means in the `x2` variable, is obtained as follows.

```
ansc = csvy(y~incr(x1)*x2, data=data_set_name, design=ds,
            nD=M, test=TRUE)
ansc$pval
```

For more information and examples, see the `csurvey` manual.

Chapter 5

Conclusion and Future Work

In this dissertation, we first proposed a new survey estimator under nonresponse, in which the propensity function is fitted by a constrained least square regression model, with the constraint being a calibration equation. In this manner, we can take advantage of the calibration equation as well as the readily implementation of the least square criterion. Even though we may misspecify the response model to some degree, the estimate will be adjusted towards the true value by the calibration equation, making the estimate more efficiency. We showed that both Horvitz-Thompson type and Hájek type estimators are asymptotic unbiased for the population mean. Also, the asymptotic variance estimators are derived for the proposed estimator and they are proved to be consistent for the corresponding true asymptotic variance. Furthermore, under the design normal assumption, we showed that both estimators are asymptotic normally distributed, regardless of the random response mechanism. In a simulation study, the proposed estimators are shown to have a good performance in terms of unbiasedness, coverage probability and mean square error, compared with other competitive uncalibrated estimators.

Next, we tackled the problem in estimating the variance of inequality constrained domain mean estimators in the finite population context. The proposed mixture variance estimator takes into account the fact that the constrained domain means estimator can be expressed through a projection matrix on a unique linear space derived from the linear constraints. This linear space is sample-dependent and thus so is the covariance of the constrained estimator. This improved variance estimator better reflects the covariance structure of the underlying constrained domain mean estimator by taking into account all possible covariance matrices obtained from the inequality constraints. Also, we formally proved the consistency of the mixture variance estimators.

In the third topic, we proposed a testing procedure for testing the linear inequality constraints of the population domain means within the survey context. The null distribution of the proposed test statistics has been shown to have the asymptotic mixture beta densities. In theory, we proved that the power of the test goes to one as the sample size increases. Also, larger effect size of the population domain means can boost the power of the test. This one-side test is easy to conduct in real practice and the simulations, as well as the real data analysis, confirm the properties of this proposed one-sided test.

Applying the shape constrained methods in survey domain estimation is a new area in survey research. Recently, Wu et al. (2016), Oliva-Avilés et al. (2019), Oliva-Avilés et al. (2020), Xu et al. (2021) proposed a lot of methods on the survey domain estimation and inference under shape restrictions. In the future, to make those work more applicable to survey practitioners, we will work on developing a new `csurvey` package that allows users to implement shape and order constraints on domain mean estimates in surveys. The new package `csurvey` will incorporate the existing methods on constrained domain estimation and inference, with commands to impose a variety of useful constraints in real surveys. Also, we plan to work on the relaxed monotone estimator in surveys. Under some circumstance, imposing strict monotone ordering might not be appropriate, then a relaxed ordering can be used instead. We will try to propose a method that may be used if the domain means can be assumed to be approximately monotone. Specifically, a type of weighted moving average can be assumed to be monotone. we will formulate the moving average with a single tuning parameter, and try to propose a data-driven choice of this parameter.

Bibliography

Alho, J. M. (1990). Adjusting for nonresponse bias using logistic regression. *Biometrika*, 77(3):617–624.

Bartholomew, D. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, 46:36–48.

Bartholomew, D. (1961). A test of homogeneity of means under order restrictions. *Journal of the Royal Statistical Society, Series B*, 23(1):239–272.

Beaumont, J.-F. (2005). Calibrated imputation in surveys under a quasi-model-assisted approach. *Journal of the Royal Statistical Society Series B*, 67:445–458.

Boistard, H., Lopuhaa, H. P., and Ruiz-Gazen, A. (2012). Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electronic Journal of Statistics*, 6:1967–1983.

Breidt, F. J. and Opsomer, J. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28:1026–1053.

Breidt, F. J. and Opsomer, J. D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statistical Science*, 32:190–205.

Cassel, C. M., Sarndal, C.-E., and Wretman, J. H. (1983). Some uses of statistical models in connection with the nonresponse problem. *Incomplete data in sample surveys*, 3:143–160.

Chacko, V. J. (1963). Testing homogeneity against ordered alternatives. *The Annals of Mathematical Statistics*, 34(3):945–956.

Chang, T. and Kott, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika*, 95:555–571.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of American Statistical Association*, 87:376–382.

Ekholm, A. and Laaksonen, S. (1991). Weighing via response modeling in the finnish household budget survey. *Journal of Official Statistics*, 7(3):325–337.

Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. *Proceedings of the American Statistical Association, Social Statistics Section*, pages 197–201.

Folsom, R. E. and Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. *Proceedings of the American Statistical Association, Survey Research Methods Section*, pages 598–603.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Fuller, W. A. (2009). *Sampling Statistics*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Giommi, A. (1984). A simple method for estimating individual response probabilities in sampling from nite populations. *Metron*, 42(4):185–200.

Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A. (2002). *Survey Nonresponse*. Wiley, New York.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35:1491–1523.

Hájek, J. (1971). Comment on a paper by D. Basu. In Godambe, V. P. and Sprott, D. A., editors, *Foundations of Statistical Inference*, page 236. Holt, Rinehart and Winston, Toronto.

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.

Iannacchione, V. G. (2003). Sequential weight adjustments for location and cooperation propensity for the 1995 national survey of family growth. *Journal of Official Statistics*, 19(1):31–43.

Iannacchione, V. G., Milne, J. G., and Folsom, R. E. (1991). Response probability weight adjustments using logistic regression. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 20:637–642.

Kim, J. K. and Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *The Canadian Journal of Statistics*, 35(4):501–514.

Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 32:133–142.

Kott, P. S. and Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of American Statistical Association*, 105(491):1265–1275.

Liao, X. and Meyer, M. C. (2014). coneproj: an R package for the primal or dual cone projections with routines for constrained regression. *Journal of Statistical Software*, 61:1–22.

Little, R. J. A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54(2):139–157.

Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45:503–528.

Lundström, S. and Särndal, C.-E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15:305–327.

McDermott, M. P. and Mudholkar, G. S. (1993). A simple approach to testing homogeneity of order-constrained means. *Journal of the American Statistical Association*, 88:1371–1379.

Meyer, M. C. (1999). An extension of the mixed primal-dual bases algorithm to the case of more constraints than dimensions. *Journal of Statistical Planning and Inference*, 81:13–31.

Meyer, M. C. (2003). A test for linear vs. convex regression function using shape-restricted regression. *Biometrika*, 90(1):223–232.

Meyer, M. C. and Wang, J. (2012). Improved power of one-sided tests. *Statistics and Probability Letters*, 82:1619–1622.

Meyer, M. C. and Woodroofe, M. (2000). On the degrees of freedom in shape-restricted regression. *The Annals of Statistics*, 28(4):1083–1104.

Oliva-Avilés, C., Meyer, M. C., and Opsomer, J. D. (2019). Checking validity of monotone domain mean estimators. *Canadian Journal of Statistics*, 47(2):315–331.

Oliva-Avilés, C., Meyer, M. C., and Opsomer, J. D. (2020). Estimation and inference of domain means subject to qualitative constraints. *Survey Methodology*, 46(2):145–180.

Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *The Annals of Mathematical Statistics*, 40(2):549–567.

Phipps, P. and Toth, D. (2012). Analyzing establishment nonresponse using aninterpretable regression tree model with linked administrative data. *The Annals of Applied Statistics*, 6(2):772–794.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. John Wiley & Sons, New York.

Runarsson, T. P. and Yao, X. (2000). Stochastic ranking for constrained evolutionary optimization. *IEEE Transactions on Evolutionary Computation*, 4(3):284 – 294.

Sarndal, C.-E. and Lundstrom, S. (2006). *Estimation in surveys with nonresponse*. John Wiley & Sons, New York.

Sen, B. and Meyer, M. C. (2017). Testing against a linear regression model using ideas from shape-restricted estimation. *Journal of the Royal Statistical Society Series B*, 79:423–448.

Silva, D. and Opsomer, J. D. (2009). Nonparametric propensity weighting for survey nonresponse through local polynomial regression. *Survey Methodology*, 35(2):165–176.

Silvapulle, M. J. and Sen, P. (2005). *Constrained Statistical Inference*. Wiley, Hoboken, New Jersey.

Wu, J., Meyer, M. C., and Opsomer, J. D. (2016). Survey estimation of domain means that respect natural orderings. *Canadian Journal of Statistics*, 44(4):431–444.

Xu, X., Meyer, M. C., and Opsomer, J. D. (2021). Improved variance estimation for inequality-constrained domain mean estimators using survey data. *Journal of Statistical Planning and Inference*, 215:47–71.

Appendix A

Supplemental Materials for Chapter 2

This appendix contains all the proofs of the theoretical results in Chapter 2.

Proof of Theorem 3. We will complete the proof by showing $n(\hat{V}_i - V_i) = o_p(1), \quad \forall i = 1, 2, 3, 4.$

$$\begin{aligned}
n(\hat{V}_1 - V_1) &= \frac{n}{N^2} \sum_{i \in S} \frac{1}{\pi_i^2} \left(\frac{1}{\hat{p}_i} - 1 \right) y_i^2 \frac{R_i}{\hat{p}_i} - \frac{n}{N^2} \sum_{i \in U} \frac{1}{\pi_i} \left(\frac{1}{p_i} - 1 \right) y_i^2 \\
&= \left\{ \frac{n}{N^2} \sum_{i \in S} \frac{1}{\pi_i^2} \left(\frac{1}{\hat{p}_i} - 1 \right) y_i^2 \frac{R_i}{\hat{p}_i} - \frac{n}{N^2} \sum_{i \in S} \frac{1}{\pi_i^2} \left(\frac{1}{p_i} - 1 \right) y_i^2 \frac{R_i}{p_i} \right\} \\
&\quad + \left\{ \frac{n}{N^2} \sum_{i \in S} \frac{1}{\pi_i^2} \left(\frac{1}{p_i} - 1 \right) y_i^2 \frac{R_i}{p_i} - \frac{n}{N^2} \sum_{i \in S} \frac{1}{\pi_i^2} \left(\frac{1}{p_i} - 1 \right) y_i^2 \right\} \\
&\quad + \left\{ \frac{n}{N^2} \sum_{i \in S} \frac{1}{\pi_i^2} \left(\frac{1}{p_i} - 1 \right) y_i^2 - \frac{n}{N^2} \sum_{i \in U} \frac{1}{\pi_i} \left(\frac{1}{p_i} - 1 \right) y_i^2 \right\} \\
&= a_1 + a_2 + a_3
\end{aligned}$$

We write a_1 as $a_1 = \frac{n}{N^2} \sum_{i \in S} \frac{R_i y_i^2}{\pi_i^2} (\frac{1}{\hat{p}_i^2} - \frac{1}{p_i^2}) - \frac{n}{N^2} \sum_{i \in S} \frac{R_i y_i^2}{\pi_i^2} (\frac{1}{\hat{p}_i} - \frac{1}{p_i})$. Using the similar argument for showing $A = O_p(n^{-\frac{1}{2}})$, it's easy to see that:

$$\frac{n}{N^2} \sum_{i \in S} \frac{R_i y_i^2}{\pi_i^2} (\frac{1}{\hat{p}_i} - \frac{1}{p_i}) = \frac{n}{N} \frac{1}{N} \sum_{i \in S} \frac{R_i y_i^2}{\pi_i^2} (\frac{1}{\hat{p}_i} - \frac{1}{p_i}) = O(1) O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}) = o_p(1)$$

Apply similar argument to the first term of a_1, yielding $\frac{n}{N^2} \sum_{i \in S} \frac{R_i y_i^2}{\pi_i^2} (\frac{1}{\hat{p}_i^2} - \frac{1}{p_i^2}) = o_p(1)$, Hence we have $a_1 = o_p(1) + o_p(1) = o_p(1)$.

We write a_2 as $a_2 = \frac{n}{N^2} \sum_{i \in S} \frac{y_i^2}{\pi_i^2 p_i^2} (R_i - p_i) - \frac{n}{N^2} \sum_{i \in S} \frac{y_i^2}{\pi_i^2 p_i} (R_i - p_i)$. By Kolmogorov's Law of Large Numbers, we have:

$$\frac{1}{n} \sum_{i \in S} \frac{y_i^2}{\pi_i^2 p_i} (R_i - p_i) = o_p(1) \quad \text{and} \quad \frac{1}{n} \sum_{i \in S} \frac{y_i^2}{\pi_i^2 p_i^2} (R_i - p_i) = o_p(1)$$

Hence $a_2 = O(1)o_p(1) + O(1)o_p(1) = o_p(1)$. By the analogous argument for showing (2.9), $a_3 = O(1)O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}) = o_p(1)$. Thus, we have shown that:

$$n(\hat{V}_1 - V_1) = o_p(1) + o_p(1) + o_p(1) = o_p(1)$$

$$\begin{aligned}
n(\hat{V}_2 - V_2) &= \frac{n}{N^2}\sum_{i\in S}\frac{(1-\pi_i)y_i^2}{\pi_i^2}\frac{R_i}{\hat{p}_i} + \frac{n}{N^2}\sum_{i\neq j}\sum_{i,j\in S}\frac{\Delta_{ij}y_iy_j}{\pi_{ij}\pi_i\pi_j}\frac{R_iR_j}{\hat{p}_i\hat{p}_j} - \frac{n}{N^2}\sum_{i,j\in U}\Delta_{ij}\frac{y_iy_j}{\pi_i\pi_j} \\
&= \left\{\frac{n}{N}\left[\frac{1}{N}\sum_{i\in S}\frac{(1-\pi_i)y_i^2}{\pi_i^2}\frac{R_i}{\hat{p}_i} - \frac{1}{N}\sum_{i\in U}\frac{(1-\pi_i)y_i^2}{\pi_i}\right]\right\} \\
&\quad + \left\{\frac{1}{N^2}\left[\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}\frac{R_iR_j}{\hat{p}_i\hat{p}_j} - \sum_{i,j\in U,i\neq j}n\Delta_{ij}\frac{y_iy_j}{\pi_i\pi_j}\right]\right\} \\
&= b_1 + b_2
\end{aligned}$$

Using the same argument for showing $n(\hat{V}_1 - V_1) = o_p(1)$, it's easy to verify that $b_1 = O(1)o_p(1) = o_p(1)$. Now, we can rewrite b_2 as:

$$\begin{aligned}
b_2 &= \left[\frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}\frac{R_iR_j}{\hat{p}_i\hat{p}_j} - \frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}\frac{R_iR_j}{p_ip_j}\right] \\
&= \left[\frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}\frac{R_iR_j}{p_ip_j} - \frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}\right] \\
&= \left[\frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j} - \frac{1}{N^2}\sum_{i,j\in U,i\neq j}n\Delta_{ij}\frac{y_iy_j}{\pi_i\pi_j}\right] \\
&= b_{21} + b_{22} + b_{23}
\end{aligned}$$

$$\begin{aligned}
b_{21} &= \frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}R_iR_j\left(\frac{1}{\hat{p}_i\hat{p}_j} - \frac{1}{p_ip_j}\right) \\
&= \frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}R_iR_j\left[\frac{1}{\hat{p}_j}\left(\frac{1}{\hat{p}_i} - \frac{1}{p_i}\right) + \frac{1}{p_i}\left(\frac{1}{\hat{p}_j} - \frac{1}{p_j}\right)\right] \\
&= \frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}R_iR_j\frac{1}{\hat{p}_j}\left(\frac{1}{\hat{p}_i} - \frac{1}{p_i}\right) + \frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_iy_j}{\pi_i\pi_j}R_iR_j\frac{1}{p_i}\left(\frac{1}{\hat{p}_j} - \frac{1}{p_j}\right)
\end{aligned}$$

By a similar argument for showing $A = o_p(1)$ in Theorem 2, we have:

$$b_{21} = O_p(1)O_p(n^{-\frac{1}{2}}) + O_p(1)O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}) = o_p(1)$$

Also, by Kolmogorov's Law of Large Numbers, we get:

$$b_{22} = \frac{n(n-1)}{N^2}\frac{1}{n(n-1)}\sum_{i,j\in S, i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{y_i y_j}{\pi_i\pi_j p_i p_j}(R_i R_j - p_i p_j) = O(1)o_p(1) = o_p(1)$$

By assumption D.1, D.2 and applying Corollary 5.1.1.1 in Fuller (1996), we can show that $b_{23} = o_p(1)$, hence $b_2 = o_p(1) + o_p(1) + o_p(1) = o_p(1)$ and thus:

$$n(\hat{V}_2 - V_2) = o_p(1) + o_p(1) = o_p(1)$$

Before proving $n(\hat{V}_3 - V_3) = o_p(1)$ and $n(\hat{V}_4 - V_4) = o_p(1)$, we first need the following result:

$$\hat{\boldsymbol{A}} - \boldsymbol{A} = o_p(1) \tag{A.1}$$

$$N(\hat{\boldsymbol{J}}^{-1}(\hat{\boldsymbol{\nu}}) - \boldsymbol{J}^{-1}(\boldsymbol{\nu}^*)) = o_p(1) \tag{A.2}$$

$$\hat{\boldsymbol{D}} - \boldsymbol{D} = o_p(1) \tag{A.3}$$

Using the same procedure for proving $n(\hat{V}_1 - V_1) = o_p(1)$, (A.1) can be easily verified. For (A.2), since $\boldsymbol{J}(.)$ matrix is continuous and invertible, it suffice to show:

$$\frac{1}{N}\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}}) - \frac{1}{N}\boldsymbol{J}(\boldsymbol{\nu}^*) = o_p(1) \tag{A.4}$$

Again, using the similar procedure for proving $n(\hat{V}_1 - V_1) = o_p(1)$, we have that:

$$\frac{1}{N}\sum_{i\in S}\frac{2}{\pi_i}\boldsymbol{B}_i\boldsymbol{B}_i^T \xrightarrow{p} \frac{1}{N}\sum_{i\in U}2\boldsymbol{B}_i\boldsymbol{B}_i^T, \quad \frac{1}{N}\sum_{i\in S}\frac{x_i\boldsymbol{B}_i R_i}{\pi_i\hat{p}_i^2} \xrightarrow{p} \frac{1}{N}\sum_{i\in U}\frac{x_i\boldsymbol{B}_i}{p_i}$$

$$\frac{1}{N}\sum_{i\in S}\frac{x_i}{\pi_i}\left(\frac{1}{\hat{p}_i}-1\right)\frac{R_i}{\hat{p}_i}\boldsymbol{B}_i{}^T \xrightarrow{p} \frac{1}{N}\sum_{i\in U}x_i\left(\frac{1}{p_i}-1\right)\boldsymbol{B}_i{}^T, \quad \frac{1}{N}\sum_{i\in S}\frac{x_i^2R_i}{2\pi_i\hat{p}_i^2}\xrightarrow{p}\frac{1}{N}\sum_{i\in U}\frac{x_i^2}{2p_i}$$

Hence, by Slutsky's Theorem, $\frac{1}{N}\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}}) \xrightarrow{p} \frac{1}{N}\boldsymbol{J}(\boldsymbol{\nu}^*)$, so (A.4) and thus (A.2) are verified.

Similarly, we apply Slutsky's Theorem to the following results, yielding (A.3).

$$\frac{1}{N}\sum_{k\in S}\frac{x_k\boldsymbol{B}_kR_k}{\pi_k\hat{p}_k^2}\xrightarrow{p}\frac{1}{N}\sum_{k\in U}\frac{x_k\boldsymbol{B}_k}{p_k}, \quad \frac{1}{N}\sum_{k\in S}\frac{x_k^2R_k}{\pi_k\hat{p}_k^2}\xrightarrow{p}\frac{1}{N}\sum_{k\in U}\frac{x_k^2}{p_k}$$

Now, based on the expression of $\hat{V}_3$ and V_3, we write $n(\hat{V}_3-V_3)$ as:

$$\begin{aligned}
&n(\hat{V}_3-V_3)\\
=&\frac{n}{N}\left\{\hat{\boldsymbol{A}}^TN[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{D}}\left(\frac{1}{N}\sum_{i\in S}\frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\left[2x_i\left(1-\frac{1}{\hat{p}_i}\right)\right]^2\right)\hat{\boldsymbol{D}}^TN[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{A}}\right.\\
&\left.-\boldsymbol{A}^TN[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{D}\left(\frac{1}{N}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\left[2x_i\left(1-\frac{1}{p_i}\right)\right]^2\right)\boldsymbol{D}^TN[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}\right\}\\
&-\frac{n}{N}\left\{\hat{\boldsymbol{A}}^TN[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{D}}\left(\frac{1}{N}\sum_{i\in S}\frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\left[4x_i\left(1-\frac{1}{\hat{p}_i}\right)\boldsymbol{B}_i{}^T\right]\right)N\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{A}}\right.\\
&\left.-\boldsymbol{A}^TN[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{D}\left(\frac{1}{N}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\left[4x_i\left(1-\frac{1}{p_i}\right)\boldsymbol{B}_i{}^T\right]\right)N[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}\right\}\\
&-\frac{n}{N}\left\{\hat{\boldsymbol{A}}^TN[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\left(\frac{1}{N}\sum_{i\in S}\frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\left[4x_i\left(1-\frac{1}{\hat{p}_i}\right)\boldsymbol{B}_i\right]\right)\hat{\boldsymbol{D}}^TN[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{A}}\right.\\
&\left.-\boldsymbol{A}^TN[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\left(\frac{1}{N}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\left[4x_i\left(1-\frac{1}{p_i}\right)\boldsymbol{B}_i\right]\right)\boldsymbol{D}^TN[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}\right\}\\
&+\frac{n}{N}\left\{\hat{\boldsymbol{A}}^TN[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\left(\frac{1}{N}\sum_{i\in S}\frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}4\boldsymbol{B}_i\boldsymbol{B}_i{}^T\right)N[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{A}}\right.\\
&\left.-\boldsymbol{A}^TN[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\left(\frac{1}{N}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}4\boldsymbol{B}_i\boldsymbol{B}_i{}^T\right)N[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}\right\}\\
=&c_1-c_2-c_3+c_4
\end{aligned}$$

In terms of c_1, we have:

$$
\begin{aligned}
&\frac{1}{N}\sum_{i\in S}\frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\left[2x_i\left(1-\frac{1}{\hat{p}_i}\right)\right]^2-\frac{1}{N}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\left[2x_i\left(1-\frac{1}{p_i}\right)\right]^2\\
=&\frac{1}{N}\sum_{i\in S}\frac{4x_i^2(1-3\hat{p}_i+3\hat{p}_i^2-\hat{p}_i^3)R_i}{\pi_i^2\hat{p}_i^2}-\frac{1}{N}\sum_{i\in U}\frac{4x_i^2(1-3p_i+3p_i^2-p_i^3)}{\pi_i p_i}\\
=&\left(\frac{1}{N}\sum_{i\in S}\frac{4x_i^2R_i}{\pi_i^2\hat{p}_i^2}-\frac{1}{N}\sum_{i\in U}\frac{4x_i^2}{\pi_i p_i}\right)-\left(\frac{1}{N}\sum_{i\in S}\frac{12x_i^2R_i}{\pi_i^2\hat{p}_i}-\frac{1}{N}\sum_{i\in U}\frac{12x_i^2}{\pi_i}\right)\\
&+\left(\frac{1}{N}\sum_{i\in S}\frac{12x_i^2R_i}{\pi_i^2}-\frac{1}{N}\sum_{i\in U}\frac{12x_i^2p_i}{\pi_i}\right)-\left(\frac{1}{N}\sum_{i\in S}\frac{4x_i^2R_i\hat{p}_i}{\pi_i^2}-\frac{1}{N}\sum_{i\in U}\frac{4x_i^2p_i^2}{\pi_i}\right)\\
=&c_{11}-c_{12}+c_{13}-c_{14}
\end{aligned}
$$

Using the same argument for showing $n(\hat{V}_1-V_1)=o_p(1)$, we obtain that $c_{11}=o_p(1)$, $c_{12}=o_p(1)$, $c_{13}=o_p(1)$ and $c_{14}=o_p(1)$. Thus, we have:

$$
\frac{1}{N}\sum_{i\in S}\frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\left[2x_i\left(1-\frac{1}{\hat{p}_i}\right)\right]^2\xrightarrow{p}\frac{1}{N}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\left[2x_i\left(1-\frac{1}{p_i}\right)\right]^2
$$

Together with the result $\hat{\boldsymbol{A}}\xrightarrow{p}\boldsymbol{A}$, $N\hat{\boldsymbol{J}}^{-1}(\hat{\boldsymbol{\nu}})\xrightarrow{p}N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*)$ and $\hat{\boldsymbol{D}}\xrightarrow{p}\boldsymbol{D}$ from (A.1), (A.2) and (A.3), respectively. We have the following result by Slutsky's Theorem:

$$
\begin{aligned}
&\hat{\boldsymbol{A}}^TN[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{D}}\left(\frac{1}{N}\sum_{i\in S}\frac{\hat{p}_i(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\left[2x_i\left(1-\frac{1}{\hat{p}_i}\right)\right]^2\right)\hat{\boldsymbol{D}}^TN[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\hat{\boldsymbol{A}}\xrightarrow{p}\\
&\boldsymbol{A}^TN[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{D}\left(\frac{1}{N}\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}\left[2x_i\left(1-\frac{1}{p_i}\right)\right]^2\right)\boldsymbol{D}^TN[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{A}
\end{aligned}
$$

That is $c_1=o_p(1)$. Using the analogous argument for proving $c_1=o_p(1)$, we have $c_2=o_p(1)$, $c_3=o_p(1)$, $c_4=o_p(1)$ and thus we get $n(\hat{V}_3-V_3)=o_p(1)$.

Since the expressions of $\hat{V}_4$ and V_4 have the similar structure as that for $\hat{V}_3$ and V_3, so using the analogous argument as for showing $n(\hat{V}_3-V_3)=o_p(1)$, we can obtain that $n(\hat{V}_4-V_4)=o_p(1)$.

Therefore, we have:

$$n(\hat{\mathrm{V}}(\bar{y}_{el}) - \mathrm{Var}(\bar{y}_{el})) = \sum_{i=1}^{4} n(\hat{V}_i - V_i) = o_p(1) + o_p(1) + o_p(1) + o_p(1) = o_p(1)$$

□

Proof of Theorem 4. Using the result in (2.13), we compute $\sqrt{n}(\bar{y}_e - \bar{y}_\pi)$ as follows:

$$\begin{aligned}
\sqrt{n}(\bar{y}_e - \bar{y}_\pi) &= \sqrt{n}\left\{\bar{y}_{e^*} + \boldsymbol{A}^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) - \bar{y}_\pi\right\} \\
&= \frac{\sqrt{n}}{N}\sum_{i\in S}\frac{y_i}{\pi_i p_i}(R_i - p_i) + \frac{\sqrt{n}}{N}\sum_{i\in S}\frac{\boldsymbol{A}^T(N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*))\boldsymbol{W}_i}{\pi_i}(R_i - p_i) + o_p(1) \\
&= \frac{\sqrt{n}}{N}\sum_{i\in S}\frac{1}{\pi_i}\left[\frac{y_i}{p_i} + \boldsymbol{A}^T(N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*))\boldsymbol{W}_i\right](R_i - p_i) + o_p(1) \\
&= \frac{\sqrt{n}}{N}\sum_{i\in S} D_i(R_i - p_i) + o_p(1)
\end{aligned} \tag{A.5}$$

where $D_i = \frac{1}{\pi_i}\left[\frac{y_i}{p_i} + \boldsymbol{A}^T(N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*))\boldsymbol{W}_i\right]$ and we assume that, for all i, $0 < |D_i| \le M$ for some constant M. Now, we want to show the following result:

$$\sqrt{n}(\bar{y}_e - \bar{y}_U) \xrightarrow{d} N(0, V_\pi + V_R) \tag{A.6}$$

where $V_\pi = \lim_{N\to\infty} \frac{n}{N^2}\sum_{i,j\in U}\frac{\Delta_{ij}y_i y_j}{\pi_i\pi_j}$ and $V_R = \lim_{N\to\infty}\frac{n}{N^2}\sum_{i\in U}\pi_i D_i^2 p_i(1-p_i)$.

By assumption D.3, we have $\sqrt{n}(\bar{y}_\pi - \bar{y}_U) \xrightarrow{d} N(0, V_\pi)$, if we can show:

$$\sqrt{n}(\bar{y}_e - \bar{y}_\pi)|S \xrightarrow{d} N(0, V_R) \quad a.s \tag{A.7}$$

then by Theorem 1.3.6 in Fuller (2009), (A.6) is proved.

Given sample $S = \{1, 2, \cdots, n\}$, denote $T_n = \sum_{i\in S} D_i(R_i - p_i)$ and $S_n^2 = \mathrm{Var}(\sum_{i\in S} D_i(R_i - p_i)) = \sum_{i\in S} D_i^2 p_i(1-p_i)$, we will prove:

$$\frac{T_n}{S_n} \xrightarrow{d} N(0,1)$$

by showing that the Lyapunov condition holds.

Set $\delta = 1$, then:

$$\begin{aligned}
\mathrm{E}(|D_i(R_i - p_i)|^{2+\delta}) &= \mathrm{E}(|D_i(R_i - p_i)|^3) \\
&= |D_i|^3[p_i(1-p_i)^3 + (1-p_i)p_i^3] \\
&\leq |D_i|^3 p_i(1-p_i) \\
&\leq M D_i^2 p_i(1-p_i)
\end{aligned}$$

Thus, $\sum_{i \in S} \mathrm{E}(|D_i(R_i - p_i)|^3) \leq M \sum_{i \in S} D_i^2 p_i(1-p_i) = M S_n^2$. So we have that:

$$\frac{\sum_{i \in S} \mathrm{E}(|D_i(R_i - p_i)|^3)}{S_n^3} \leq \frac{M S_n^2}{S_n^3} = \frac{M}{S_n} \longrightarrow 0, \quad as \quad n \to \infty$$

Hence, Lyapunov condition is satisfied. By Lyapunov central limit theorem, we have:

$$\frac{T_n}{S_n} = \frac{\sum_{i \in S} D_i(R_i - p_i)}{\sqrt{\sum_{i \in S} D_i^2 p_i(1-p_i)}} = \frac{\frac{\sqrt{n}}{N} \sum_{i \in S} D_i(R_i - p_i)}{\sqrt{\frac{n}{N^2} \sum_{i \in S} D_i^2 p_i(1-p_i)}} \xrightarrow{d} N(0,1)$$

Now, we are ready to show $\left. \frac{\sum_{i \in S} D_i(R_i - p_i)}{\sqrt{\sum_{i \in S} D_i^2 p_i(1-p_i)}} \right| S \xrightarrow{d} N(0,1) \quad a.s.$

By Berry Esseen Theorem, for any value t, there exists a absolute constant C_0, such that:

$$\begin{aligned}
&\left| P\left\{ \frac{\sum_{i \in S} D_i(R_i - p_i)}{\sqrt{\sum_{i \in S} D_i^2 p_i(1-p_i)}} < t \middle| S \right\} - \Phi(t) \right| \\
\leq& C_0 \frac{\sum_{i \in S} \mathrm{E}(|D_i(R_i - p_i)|^3)}{S_n^3} \\
\leq& \frac{C_0 M}{\sqrt{\sum_{i \in S} D_i^2 p_i(1-p_i)}}
\end{aligned}$$

where $\Phi(\cdot)$ denote the cumulative distribution function of the standard normal distribution. Hence, by assumption D.2 and above inequality, we have that:

$$
\begin{aligned}
&\mathrm{E}\left|P\left\{\frac{\sum_{i\in S} D_i(R_i-p_i)}{\sqrt{\sum_{i\in S} D_i^2 p_i(1-p_i)}}<t\,\middle|\,S\right\}-\Phi(t)\right|^4\\
\leq&\mathrm{E}\left[\frac{C_0 M}{\sqrt{\sum_{i\in S} D_i^2 p_i(1-p_i)}}\right]^4\\
=&\frac{C_0^4 M^4}{N^2}\mathrm{E}\left[\left(\frac{1}{N}\sum_{i\in U} I_i D_i^2 p_i(1-p_i)\right)^{-2}\right]\\
<&\frac{C_0^4 M^4}{n^2[\min_i D_i^2 p_i(1-p_i)]^2}\mathrm{E}\left[\left(\frac{1}{N}\sum_{i\in U} I_i\right)^{-2}\right]\\
\leq&\frac{C_0^4 M^4}{[\min_i D_i^2 p_i(1-p_i)]^2(\pi^*)^2}\frac{1}{n^2}
\end{aligned}
$$

which implies:

$$
\sum_{n=1}^{\infty}\mathrm{E}\left|P\left\{\frac{\sum_{i=1}^n D_i(R_i-p_i)}{\sqrt{\sum_{i=1}^n D_i^2 p_i(1-p_i)}}<t\,\middle|\,S\right\}-\Phi(t)\right|^4 \leq \frac{C_0^4 M^4}{[\inf_i D_i^2 p_i(1-p_i)]^2(\pi^*)^2}\sum_{n=1}^{\infty}\frac{1}{n^2}<\infty \tag{A.8}
$$

Given any $\varepsilon>0$, by Markov's inequality and (A.8), we have:

$$
\begin{aligned}
&\sum_{n=1}^{\infty}P\left\{\left|P\left[\frac{\sum_{i\in S} D_i(R_i-p_i)}{\sqrt{\sum_{i\in S} D_i^2 p_i(1-p_i)}}<t\,\middle|\,S\right]-\Phi(t)\right|>\varepsilon\right\}\\
=&\sum_{n=1}^{\infty}P\left\{\left|P\left[\frac{\sum_{i\in S} D_i(R_i-p_i)}{\sqrt{\sum_{i\in S} D_i^2 p_i(1-p_i)}}<t\,\middle|\,S\right]-\Phi(t)\right|^4>\varepsilon^4\right\}\\
\leq&\sum_{n=1}^{\infty}\mathrm{E}\left|P\left[\frac{\sum_{i=1}^n D_i(R_i-p_i)}{\sqrt{\sum_{i=1}^n D_i^2 p_i(1-p_i)}}<t\,\middle|\,S\right]-\Phi(t)\right|^4\Big/\varepsilon^4\\
<&\infty
\end{aligned}
$$

By Borel-Cantelli lemma, we have:

$$P\left\{\left|P\left[\frac{\sum_{i\in S} D_i(R_i-p_i)}{\sqrt{\sum_{i\in S} D_i^2 p_i(1-p_i)}}<t\,\middle|\, S\right]-\Phi(t)\right|>\varepsilon,\ i.o.\right\}=0$$

which is equivalent to:

$$\left.\frac{\sum_{i\in S} D_i(R_i-p_i)}{\sqrt{\sum_{i\in S} D_i^2 p_i(1-p_i)}}\right| S \xrightarrow{d} N(0,1) \quad a.s \tag{A.9}$$

By similar proof for (2.9), we have:

$$\frac{1}{N}\sum_{i\in S} D_i^2 p_i(1-p_i) \xrightarrow{p} \frac{1}{N}\sum_{i\in U}\pi_i D_i^2 p_i(1-p_i) \tag{A.10}$$

Using (A.10), (A.9), (A.5) and applying Slutsky's Theorem, we can get (A.7) and thus (A.6) is proved.

Notice that $V_\pi + V_R = \lim_{N\to\infty} n\left\{\frac{1}{N^2}\sum_{i,j\in U}\frac{\Delta_{ij}y_iy_j}{\pi_i\pi_j}+\frac{1}{N^2}\sum_{i\in U}\pi_i D_i^2 p_i(1-p_i)\right\}$, where:

$$\begin{aligned}
&\frac{1}{N^2}\sum_{i,j\in U}\frac{\Delta_{ij}y_iy_j}{\pi_i\pi_j}+\frac{1}{N^2}\sum_{i\in U}\pi_i D_i^2 p_i(1-p_i)\\
&=\mathrm{Var}(\bar{y}_\pi)+\frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left[\frac{y_i}{p_i}+\boldsymbol{A}^T(N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*))\boldsymbol{W}_i\right]^2 p_i(1-p_i)\\
&=\mathrm{Var}(\bar{y}_\pi)+\frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left[\frac{y_i^2}{p_i^2}+(\boldsymbol{A}^T(N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*))\boldsymbol{W}_i)^2+\frac{2y_i}{p_i}\boldsymbol{A}^T(N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*))\boldsymbol{W}_i\right]p_i(1-p_i)\\
&=\mathrm{Var}(\bar{y}_\pi)+\frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left[\frac{y_i^2}{p_i^2}+(C_iN)^2+\frac{2y_i}{p_i}C_iN\right]p_i(1-p_i)\\
&=\mathrm{Var}(\bar{y}_\pi)+\frac{1}{N^2}\sum_{i\in U}\frac{y_i^2}{\pi_i}\frac{1-p_i}{p_i}+\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}C_i^2+\frac{2}{N}\sum_{i\in U}\frac{y_iC_i(1-p_i)}{\pi_i}\\
&=\mathrm{Var}(\bar{y}_{el})
\end{aligned}$$

So by the result: $\frac{\sqrt{n}(\bar{y}_e-\bar{y}_U)}{\sqrt{V_\pi+V_R}} \xrightarrow{d} N(0,1)$ from (A.6), we have that:

$$\frac{\sqrt{n}(\bar{y}_e-\bar{y}_U)}{\sqrt{n\mathrm{Var}(\bar{y}_{el})}} \xrightarrow{d} N(0,1)$$

Apply Theorem 3 and Slutsky's Theorem to above result, yielding:

$$\frac{\sqrt{n}(\bar{y}_e - \bar{y}_U)}{\sqrt{n\hat{\mathrm{V}}(\bar{y}_{el})}} = \frac{\bar{y}_e - \bar{y}_U}{\sqrt{\hat{V}(\bar{y}_{el})}} \xrightarrow{d} N(0,1)$$

$\square$

Proof of Theorem 6. By Lemma 4, $\mathrm{Var}(\bar{y}_{HL})$ is expressed as:

$$\begin{aligned}
\mathrm{Var}(\bar{y}_{HL}) &= \frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)(y_i-\bar{y}_U)^2 + \frac{1}{N^2}\sum_{i,j\in U}\Delta_{ij}\frac{(y_i-\bar{y}_U)(y_j-\bar{y}_U)}{\pi_i\pi_j} \\
&\quad + \sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}C_{Hi}^2 + \frac{2}{N}\sum_{i\in U}\frac{C_{Hi}(1-p_i)(y_i-\bar{y}_U)}{\pi_i} \\
&= V_{H1} + V_{H2} + V_{H3} + V_{H4}
\end{aligned}$$

We will show $n(\hat{V}_{Hi} - V_{Hi}) = o_p(1), \forall i = 1,2,3,4.$

$$\begin{aligned}
&n(\hat{V}_{H1} - V_{H1}) \\
=&\frac{n}{\hat{N}^2}\sum_{i\in S}\frac{1}{\pi_i^2}\left(\frac{1}{\hat{p}_i}-1\right)(y_i^2 - 2y_i\bar{y}_H + \bar{y}_H^2)\frac{R_i}{\hat{p}_i} - \frac{n}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)(y_i^2 - 2y_i\bar{y}_U + \bar{y}_U^2) \\
=&\left\{\frac{N^2}{\hat{N}^2}\frac{n}{N^2}\sum_{i\in S}\frac{1}{\pi_i^2}\left(\frac{1}{\hat{p}_i}-1\right)y_i^2\frac{R_i}{\hat{p}_i} - \frac{n}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)y_i^2\right\} \\
&-\left\{\frac{2\bar{y}_H N^2}{\hat{N}^2}\frac{n}{N^2}\sum_{i\in S}\frac{1}{\pi_i^2}\left(\frac{1}{\hat{p}_i}-1\right)y_i\frac{R_i}{\hat{p}_i} - \frac{2\bar{y}_U n}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)y_i\right\} \\
&+\left\{\frac{\bar{y}_H^2 N^2}{\hat{N}^2}\frac{n}{N^2}\sum_{i\in S}\frac{1}{\pi_i^2}\left(\frac{1}{\hat{p}_i}-1\right)\frac{R_i}{\hat{p}_i} - \frac{\bar{y}_U^2 n}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)\right\} \\
=&d_1 + d_2 + d_3
\end{aligned}$$

Using a similar proof for Theorem 2, it's easy to show that $\frac{\hat{N}}{N} \xrightarrow{p} 1$, and thus $\frac{\hat{N}^2}{N^2} \xrightarrow{p} 1$ by Slutsky's theorem. By Theorem 5, $\bar{y}_H - \bar{y}_U = o_p(1)$, so we also have $\bar{y}_H^2 \xrightarrow{p} \bar{y}_U^2$ by Slutsky's theorem.

From Theorem 3, we have $n(\hat{V}_1 - V_1) = o_p(1)$, together with the result $\frac{\hat{N}^2}{N^2} \xrightarrow{p} 1$, we have: $d_1 = o_p(1)$. By a similar argument for showing $n(\hat{V}_1 - V_1) = o_p(1)$, we get the following results:

$$\frac{n}{N^2}\sum_{i\in S}\frac{1}{\pi_i^2}\left(\frac{1}{\hat{p}_i}-1\right)y_i\frac{R_i}{\hat{p}_i} \xrightarrow{p} \frac{n}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right)y_i \tag{A.11}$$

$$\frac{n}{N^2}\sum_{i\in S}\frac{1}{\pi_i^2}\left(\frac{1}{\hat{p}_i}-1\right)\frac{R_i}{\hat{p}_i} \xrightarrow{p} \frac{n}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left(\frac{1}{p_i}-1\right) \tag{A.12}$$

Apply Slutsky's theorem to (A.11), $\frac{\hat{N}^2}{N^2} \xrightarrow{p} 1$ and $\bar{y}_H - \bar{y}_U = o_p(1)$, we have $d_2 = o_p(1)$. Also, apply Slutsky's theorem to (A.12), $\frac{\hat{N}^2}{N^2} \xrightarrow{p} 1$ and $\bar{y}_H^2 \xrightarrow{p} \bar{y}_U^2$, we have $d_3 = o_p(1)$. Therefore, we have:

$$n(\hat{V}_{H1} - V_{H1}) = o_p(1) + o_p(1) + o_p(1) = o_p(1)$$

Now, we rewrite $n(\hat{V}_{H2} - V_{H2})$ as:

$$\begin{aligned}
&n(\hat{V}_{H2} - V_{H2}) \\
=&\left\{\frac{n}{\hat{N}^2}\sum_{i\in S}\frac{(1-\pi_i)(y_i-\bar{y}_H)^2}{\pi_i^2}\frac{R_i}{\hat{p}_i} - \frac{n}{N^2}\sum_{i\in U}\frac{(1-\pi_i)(y_i-\bar{y}_U)^2}{\pi_i}\right\} \\
&+\left\{\frac{n}{\hat{N}^2}\sum_{i,j\in S, i\neq j}\frac{\Delta_{ij}}{\pi_{ij}}\frac{(y_i-\bar{y}_H)(y_j-\bar{y}_H)}{\pi_i\pi_j}\frac{R_iR_j}{\hat{p}_i\hat{p}_j} - \frac{n}{N^2}\sum_{i,j\in U, i\neq j}\Delta_{ij}\frac{(y_i-\bar{y}_U)(y_j-\bar{y}_U)}{\pi_i\pi_j}\right\} \\
=&e_1 + e_2
\end{aligned}$$

Using the same argument for proving $n(\hat{V}_{H1}-V_{H1})=o_p(1)$, it's easy to show that $e_1=o_p(1)$. We now write e_2 as:

$$
\begin{aligned}
e_2 &= \frac{N^2}{\hat{N}^2}\frac{n}{N^2}\sum_{i,j\in S,i\neq j}\frac{\Delta_{ij}}{\pi_{ij}}\frac{R_iR_j}{\pi_i\pi_j\hat{p}_i\hat{p}_j}(y_iy_j-y_i\bar{y}_H-y_j\bar{y}_H+\bar{y}_H^2)\\
&\quad -\frac{n}{N^2}\sum_{i,j\in U,i\neq j}\Delta_{ij}\frac{1}{\pi_i\pi_j}(y_iy_j-y_i\bar{y}_U-y_j\bar{y}_U+\bar{y}_U^2)\\
&=\left[\frac{N^2}{\hat{N}^2}\frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{R_iR_j}{\pi_i\pi_j\hat{p}_i\hat{p}_j}y_iy_j-\frac{1}{N^2}\sum_{i,j\in U,i\neq j}n\Delta_{ij}\frac{1}{\pi_i\pi_j}y_iy_j\right]\\
&\quad -\left[\frac{N^2}{\hat{N}^2}\frac{\bar{y}_H}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{R_iR_j}{\pi_i\pi_j\hat{p}_i\hat{p}_j}(y_i+y_j)-\frac{\bar{y}_U}{N^2}\sum_{i,j\in U,i\neq j}n\Delta_{ij}\frac{1}{\pi_i\pi_j}(y_i+y_j)\right]\\
&\quad +\left[\frac{N^2}{\hat{N}^2}\frac{\bar{y}_H^2}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{R_iR_j}{\pi_i\pi_j\hat{p}_i\hat{p}_j}-\frac{\bar{y}_U^2}{N^2}\sum_{i,j\in U,i\neq j}n\Delta_{ij}\frac{1}{\pi_i\pi_j}\right]\\
&= e_{21}+e_{22}+e_{23}
\end{aligned}
$$

By the result in Theorem 3, we have $b_2=\frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{R_iR_jy_iy_j}{\pi_i\pi_j\hat{p}_i\hat{p}_j}-\frac{1}{N^2}\sum_{i,j\in U,i\neq j}n\Delta_{ij}\frac{y_iy_j}{\pi_i\pi_j}=o_p(1)$, together with $\frac{\hat{N}^2}{N^2}\xrightarrow{p}1$, we can get $e_{21}=o_p(1)$. By the same argument for proving $b_2=o_p(1)$, we have:

$$
\frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{R_iR_j}{\pi_i\pi_j\hat{p}_i\hat{p}_j}(y_i+y_j)\xrightarrow{p}\frac{1}{N^2}\sum_{i,j\in U,i\neq j}n\Delta_{ij}\frac{1}{\pi_i\pi_j}(y_i+y_j) \tag{A.13}
$$

$$
\frac{1}{N^2}\sum_{i,j\in S,i\neq j}\frac{n\Delta_{ij}}{\pi_{ij}}\frac{R_iR_j}{\pi_i\pi_j\hat{p}_i\hat{p}_j}\xrightarrow{p}\frac{1}{N^2}\sum_{i,j\in U,i\neq j}n\Delta_{ij}\frac{1}{\pi_i\pi_j} \tag{A.14}
$$

Apply Slutsky's theorem to (A.13), $\frac{\hat{N}^2}{N^2}\xrightarrow{p}1$ and $\bar{y}_H-\bar{y}_U=o_p(1)$, we have $e_{22}=o_p(1)$. Also, apply Slutsky's theorem to (A.14), $\frac{\hat{N}^2}{N^2}\xrightarrow{p}1$ and $\bar{y}_H^2\xrightarrow{p}\bar{y}_U^2$, we have $e_{23}=o_p(1)$. Therefore, we have: $e_2=o_p(1)+o_p(1)+o_p(1)=o_p(1)$. Thus,

$$
n(\hat{V}_{H2}-V_{H2})=o_p(1)+o_p(1)=o_p(1)
$$

Before showing $n(\hat{V}_{H3} - V_{H3}) = o_p(1)$ and $n(\hat{V}_{H4} - V_{H4}) = o_p(1)$, we first need to show $\hat{\boldsymbol{A}}_H - \boldsymbol{A}_H = o_p(1)$.

$$
\begin{aligned}
\hat{\boldsymbol{A}}_H - \boldsymbol{A}_H &= -\left(\frac{N}{\hat{N}}\frac{1}{N}\sum_{i\in S}\frac{\boldsymbol{B}_i R_i y_i}{\pi_i \hat{p}_i^2} - \frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i y_i}{p_i}\right) + \left(\frac{N}{\hat{N}}\frac{\bar{y}_H}{N}\sum_{i\in S}\frac{\boldsymbol{B}_i R_i}{\pi_i \hat{p}_i^2} - \frac{\bar{y}_U}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i}{p_i}\right)\\
&= f_1 + f_2
\end{aligned}
$$

Since we have $\hat{\boldsymbol{A}} - \boldsymbol{A} = o_p(1)$ and $\frac{\hat{N}}{N} \xrightarrow{p} 1$, so $f_1 = o_p(1)$. Apply Slutsky's theorem to $\frac{\hat{N}}{N} \xrightarrow{p} 1$, $\bar{y}_H - \bar{y}_U = o_p(1)$ and $\frac{1}{N}\sum_{i\in S}\frac{\mathbf{B_i} R_i}{\pi_i \hat{p}_i^2} \xrightarrow{p} \frac{1}{N}\sum_{i\in U}\frac{\boldsymbol{B}_i}{p_i}$, we get $f_2 = o_p(1)$, therefore, we have:

$$
\hat{\boldsymbol{A}}_H - \boldsymbol{A}_H = o_p(1)
$$

Using exactly the same procedure for showing $n(\hat{V}_3 - V_3) = o_p(1)$ in Theorem 3 (just replace $\boldsymbol{A}$ with $\boldsymbol{A}_H$ and replace $\hat{\boldsymbol{A}}$ with $\hat{\boldsymbol{A}}_H$), we have that:

$$
n(\hat{V}_{H3} - V_{H3}) = o_p(1)
$$

Now, we write $n(\hat{V}_{H4} - V_{H4})$ as:

$$
\begin{aligned}
& n(\hat{V}_{H4} - V_{H4})\\
=& 2\frac{N}{\hat{N}}\frac{n}{N}\hat{\boldsymbol{A}}_H^T[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\sum_{i\in S}\frac{(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\hat{\boldsymbol{W}}_i(y_i - \bar{y}_H)\\
& - 2\frac{n}{N}\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{(1-p_i)}{\pi_i}\boldsymbol{W}_i(y_i - \bar{y}_U)\\
=& \left(2\frac{N}{\hat{N}}\frac{n}{N}\hat{\boldsymbol{A}}_H^T[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\sum_{i\in S}\frac{(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\hat{\boldsymbol{W}}_i y_i - 2\frac{n}{N}\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{(1-p_i)}{\pi_i}\boldsymbol{W}_i y_i\right)\\
& - \left(2\bar{y}_H\frac{N}{\hat{N}}\frac{n}{N}\hat{\boldsymbol{A}}_H^T[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\sum_{i\in S}\frac{(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\hat{\boldsymbol{W}}_i - 2\bar{y}_U\frac{n}{N}\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{(1-p_i)}{\pi_i}\boldsymbol{W}_i\right)\\
=& g_1 + g_2
\end{aligned}
$$

Using the similar proof for showing $n(\hat{V}_3 - V_3) = o_p(1)$, we have:

$$\frac{n}{N}\hat{\boldsymbol{A}}_H^T[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\sum_{i\in S}\frac{(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\hat{\boldsymbol{W}}_i y_i \xrightarrow{p} \frac{n}{N}\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{(1-p_i)}{\pi_i}\boldsymbol{W}_i y_i$$

$$\frac{n}{N}\hat{\boldsymbol{A}}_H^T[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\nu}})]^{-1}\sum_{i\in S}\frac{(1-\hat{p}_i)}{\pi_i^2}\frac{R_i}{\hat{p}_i}\hat{\boldsymbol{W}}_i \xrightarrow{p} \frac{n}{N}\boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\sum_{i\in U}\frac{(1-p_i)}{\pi_i}\boldsymbol{W}_i$$

together with the result $\frac{\hat{N}}{N} \xrightarrow{p} 1$ and $\bar{y}_H - \bar{y}_U = o_p(1)$, we have $g_1 = o_p(1)$ and $g_2 = o_p(1)$. Hence, we get:

$$n(\hat{V}_{H4} - V_{H4}) = o_p(1)$$

Overall, $n(\hat{V}(\bar{y}_{HL}) - \mathrm{Var}(\bar{y}_{HL})) = o_p(1) + o_p(1) + o_p(1) + o_p(1) = o_p(1)$ □

Proof of Theorem 7. From (2.14), we have:

$$\begin{aligned}
&\bar{y}_H - \bar{y}_U \\
=&\bar{y}_{HL} + o_p(n^{-\frac{1}{2}}) \\
=&\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)R_i}{\pi_i p_i} + \boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + o_p(n^{-\frac{1}{2}}) \\
=&\frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)(R_i-p_i)}{\pi_i p_i} + \boldsymbol{A}_H^T[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{S}_L(\boldsymbol{\nu}^*) + \frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)}{\pi_i} + o_p(n^{-\frac{1}{2}}) \\
=&\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left[\frac{(y_i-\bar{y}_U)}{p_i} + \boldsymbol{A}_H^T N[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{W}_i\right](R_i-p_i) + \frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)}{\pi_i} + o_p(n^{-\frac{1}{2}}) \\
=&\frac{1}{N}\sum_{i\in S}D_{Hi}(R_i-p_i) + \frac{1}{N}\sum_{i\in S}\frac{(y_i-\bar{y}_U)}{\pi_i} + o_p(n^{-\frac{1}{2}}) \qquad \text{(A.15)}
\end{aligned}$$

where $D_{Hi} = \frac{1}{\pi_i}\left[\frac{(y_i-\bar{y}_U)}{p_i} + \boldsymbol{A}_H^T N[\boldsymbol{J}(\boldsymbol{\nu}^*)]^{-1}\boldsymbol{W}_i\right]$ and we assume that, for all i, $|D_{Hi}| \le M_H$ for some constant M_H.

Now, we want to prove the following result:

$$\sqrt{n}\bar{y}_{HL} \xrightarrow{d} N(0, V_{H\pi} + V_{HR}) \qquad \text{(A.16)}$$

where $V_{H\pi} = \lim_{N\to\infty} \frac{n}{N^2} \sum_{i,j\in U} \frac{\Delta_{ij}(y_i-\bar{y}_U)(y_j-\bar{y}_U)}{\pi_i\pi_j}$ and $V_{HR} = \lim_{N\to\infty} \frac{n}{N^2} \sum_{i\in U} \pi_i D_{Hi}^2 p_i(1-p_i)$. Denote $\bar{y}_{H\pi} = \frac{1}{N}\sum_{i\in S} \frac{(y_i-\bar{y}_U)}{\pi_i}$, From assumption D.3, we have $\sqrt{n}\bar{y}_{H\pi} \xrightarrow{d} N(0, V_{H\pi})$, if we can show:

$$\sqrt{n}(\bar{y}_{HL} - \bar{y}_{H\pi})|S \xrightarrow{d} N(0, V_{HR}), \quad a.s \tag{A.17}$$

then by Theorem 1.3.6 in Fuller (2009), (A.16) is proved. Given sample $S = \{1, 2, \cdots, n\}$, denote $T_{Hn} = \sum_{i\in S} D_{Hi}(R_i - p_i)$ and $S_{Hn}^2 = \mathrm{Var}(\sum_{i\in S} D_{Hi}(R_i - p_i)) = \sum_{i\in S} D_{Hi}^2 p_i(1-p_i)$, we will prove:

$$\frac{T_{Hn}}{S_{Hn}} \xrightarrow{d} N(0,1)$$

by showing that the Lyapunov condition holds. Let $\delta = 1$, then:

$$\begin{aligned}
\mathrm{E}(|D_{Hi}(R_i - p_i)|^{2+\delta}) &= \mathrm{E}(|D_{Hi}(R_i - p_i)|^3) \\
&= |D_{Hi}|^3[p_i(1-p_i)^3 + (1-p_i){p_i}^3] \\
&\le |D_{Hi}|^3 p_i(1-p_i) \\
&\le M_H D_{Hi}^2 p_i(1-p_i)
\end{aligned}$$

Thus, $\sum_{i\in S} \mathrm{E}(|D_{Hi}(R_i - p_i)|^3) \le M_H \sum_{i\in S} D_{Hi}^2 p_i(1-p_i) = M_H S_{Hn}^2$. So we have that:

$$\frac{\sum_{i\in S} \mathrm{E}(|D_{Hi}(R_i - p_i)|^3)}{S_{Hn}^3} \le \frac{M_H S_{Hn}^2}{S_{Hn}^3} = \frac{M_H}{S_{Hn}} \longrightarrow 0, \quad as \quad n\to\infty$$

Hence, by Lyapunov central limit theorem, we have:

$$\frac{T_{Hn}}{S_{Hn}} = \frac{\sum_{i\in S} D_{Hi}(R_i - p_i)}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}} = \frac{\frac{\sqrt{n}}{N}\sum_{i\in S} D_{Hi}(R_i - p_i)}{\sqrt{\frac{n}{N^2}\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}} \xrightarrow{d} N(0,1)$$

Now, it's ready to show $\left.\frac{\sum_{i\in S} D_{Hi}(R_i-p_i)}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}}\right| S \xrightarrow{d} N(0,1) \quad a.s.$ By Berry-Esseen Theorem, for any value t, there exists a absolute constant C_0, such that:

$$\begin{aligned}
&\left|P\left\{\left.\frac{\sum_{i\in S} D_{Hi}(R_i-p_i)}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}} < t \right| S\right\} - \Phi(t)\right| \\
\leq & C_0 \frac{\sum_{i\in S} \mathrm{E}(|D_{Hi}(R_i-p_i)|^3)}{S_{Hn}^3} \\
\leq & \frac{C_0 M_H}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}}
\end{aligned}$$

where $\Phi(\cdot)$ denote the cumulative distribution function of the standard normal distribution. Hence, by assumption D.2 and above result, we have that:

$$\begin{aligned}
& \mathrm{E}\left|P\left\{\left.\frac{\sum_{i\in S} D_{Hi}(R_i-p_i)}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}} < t \right| S\right\} - \Phi(t)\right|^4 \\
\leq & \mathrm{E}\left[\frac{C_0 M_H}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}}\right]^4 \\
= & \frac{C_0^4 M_H^4}{N^2} \mathrm{E}\left[\left(\frac{1}{N}\sum_{i\in U} I_i D_{Hi}^2 p_i(1-p_i)\right)^{-2}\right] \\
< & \frac{C_0^4 M_H^4}{n^2[\min_i D_{Hi}^2 p_i(1-p_i)]^2} \mathrm{E}\left[\left(\frac{1}{N}\sum_{i\in U} I_i\right)^{-2}\right] \\
\leq & \frac{C_0^4 M_H^4}{[\min_i D_{Hi}^2 p_i(1-p_i)]^2(\pi^*)^2} \frac{1}{n^2}
\end{aligned}$$

which implies:

$$\sum_{n=1}^{\infty} \mathrm{E}\left|P\left\{\left.\frac{\sum_{i=1}^n D_{Hi}(R_i-p_i)}{\sqrt{\sum_{i=1}^n D_{Hi}^2 p_i(1-p_i)}} < t \right| S\right\} - \Phi(t)\right|^4 \leq \frac{C_0^4 M_H^4}{[\inf_i D_{Hi}^2 p_i(1-p_i)]^2(\pi^*)^2} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty \tag{A.18}$$

Therefore, for any $\varepsilon > 0$, by Markov's inequality and (A.18):

$$\begin{aligned}
&\sum_{n=1}^{\infty} P\left\{\left|P\left[\frac{\sum_{i\in S} D_{Hi}(R_i - p_i)}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}} < t \middle| S\right] - \Phi(t)\right| > \varepsilon\right\} \\
=&\sum_{n=1}^{\infty} P\left\{\left|P\left[\frac{\sum_{i\in S} D_{Hi}(R_i - p_i)}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}} < t \middle| S\right] - \Phi(t)\right|^4 > \varepsilon^4\right\} \\
\leq&\sum_{n=1}^{\infty} \mathrm{E}\left|P\left[\frac{\sum_{i=1}^{n} D_{Hi}(R_i - p_i)}{\sqrt{\sum_{i=1}^{n} D_{Hi}^2 p_i(1-p_i)}} < t \middle| S\right] - \Phi(t)\right|^4 \Bigg/ \varepsilon^4 \\
<&\infty
\end{aligned}$$

By Borel-Cantelli lemma, we have:

$$P\left\{\left|P\left[\frac{\sum_{i\in S} D_{Hi}(R_i - p_i)}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}} < t \middle| S\right] - \Phi(t)\right| > \varepsilon,\ i.o.\right\} = 0$$

which is equivalent to:

$$\left.\frac{\sum_{i\in S} D_{Hi}(R_i - p_i)}{\sqrt{\sum_{i\in S} D_{Hi}^2 p_i(1-p_i)}}\right| S \xrightarrow{d} N(0,1) \quad a.s \tag{A.19}$$

By an analogous proof for (2.9), we have:

$$\frac{1}{N}\sum_{i\in S} D_{Hi}^2 p_i(1-p_i) \xrightarrow{p} \frac{1}{N}\sum_{i\in U} \pi_i D_{Hi}^2 p_i(1-p_i) \tag{A.20}$$

By (A.19), (A.20), (A.15) and apply Slutsky's theorem, we can obtain (A.17) and thus (A.16) is proved. Notice that:

$$V_{H\pi} + V_{HR} = \lim_{N\to\infty} n\left\{\frac{1}{N^2}\sum_{i,j\in U} \frac{\Delta_{ij}(y_i - \bar{y}_U)(y_j - \bar{y}_U)}{\pi_i\pi_j} + \frac{1}{N^2}\sum_{i\in U} \pi_i D_{Hi}^2 p_i(1-p_i)\right\}$$

where:

$$
\begin{aligned}
&\frac{1}{N^2}\sum_{i,j\in U}\frac{\Delta_{ij}(y_i-\bar{y}_U)(y_j-\bar{y}_U)}{\pi_i\pi_j}+\frac{1}{N^2}\sum_{i\in U}\pi_i D_{Hi}^2 p_i(1-p_i)\\
&=\mathrm{Var}(\bar{y}_{H\pi})+\frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left[\frac{(y_i-\bar{y}_U)^2}{p_i^2}+(\boldsymbol{A}_H^T(N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*))\boldsymbol{W}_i)^2\right.\\
&\quad\left.+\frac{2(y_i-\bar{y}_U)}{p_i}\boldsymbol{A}_H^T(N\boldsymbol{J}^{-1}(\boldsymbol{\nu}^*))\boldsymbol{W}_i\right]p_i(1-p_i)\\
&=\mathrm{Var}(\bar{y}_{H\pi})+\frac{1}{N^2}\sum_{i\in U}\frac{1}{\pi_i}\left[\frac{(y_i-\bar{y}_U)^2}{p_i^2}+(C_{Hi}N)^2+\frac{2(y_i-\bar{y}_U)}{p_i}C_{Hi}N\right]p_i(1-p_i)\\
&=\mathrm{Var}(\bar{y}_{H\pi})+\frac{1}{N^2}\sum_{i\in U}\frac{(y_i-\bar{y}_U)^2}{\pi_i}\frac{1-p_i}{p_i}+\sum_{i\in U}\frac{p_i(1-p_i)}{\pi_i}C_{Hi}^2+\frac{2}{N}\sum_{i\in U}\frac{(y_i-\bar{y}_U)C_{Hi}(1-p_i)}{\pi_i}\\
&=\mathrm{Var}(\bar{y}_{HL})
\end{aligned}
$$

So by the result $\frac{\sqrt{n}\bar{y}_{HL}}{\sqrt{V_{H\pi}+V_{HR}}}\xrightarrow{d}N(0,1)$, we have that:

$$\frac{\sqrt{n}\bar{y}_{HL}}{\sqrt{n\mathrm{Var}(\bar{y}_{HL})}}\xrightarrow{d}N(0,1)$$

By Theorem 6 and Slutsky's theorem, we have:

$$\frac{\sqrt{n}\bar{y}_{HL}}{\sqrt{n\hat{\mathrm{V}}(\bar{y}_{HL})}}\xrightarrow{d}N(0,1)$$

Finally, from (2.14), we get the desired result:

$$\frac{\sqrt{n}(\bar{y}_H-\bar{y}_U)}{\sqrt{n\hat{\mathrm{V}}(\bar{y}_{HL})}}=\frac{\bar{y}_H-\bar{y}_U}{\sqrt{\hat{\mathrm{V}}(\bar{y}_{HL})}}\xrightarrow{d}N(0,1)$$

□

Appendix B

Supplemental Materials for Chapter 3

B.1 Proof of Proposition 1

B.1.1 Derivation of the Expression in (3.2)

From the unweighted constrained least square problem in Section 3.2.1, the solution $\tilde{\boldsymbol{\phi}}$ is the projection of $\tilde{\boldsymbol{z}}_s$ onto the set

$$\boldsymbol{\Omega}_s = \{\boldsymbol{\phi} \in \mathbb{R}^D : \boldsymbol{A}_s\boldsymbol{\phi} \geq \boldsymbol{0}\},$$

which defines a convex cone in $\mathbb{R}^D$. The necessary and sufficient conditions for a vector $\tilde{\boldsymbol{\phi}}$ to be the projection of $\tilde{\boldsymbol{z}}_s$ onto $\boldsymbol{\Omega}_s$ can be found in Robertson et al. (1988), Chapter 1, or Silvapulle and Sen (2005), Chapter 3.

When the constraint matrix $\boldsymbol{A}_s$ is not full row rank, the cone projection is more efficiently solved by computing the projection onto the polar cone $\boldsymbol{\Omega}_s^0$, defined by

$$\boldsymbol{\Omega}_s^0 = \{\boldsymbol{\rho} \in \mathrm{R}^D : \boldsymbol{\rho} = \sum_{j=1}^{m} a_j\boldsymbol{\gamma}_j, a_j \geq 0, j = 1, \cdots, m\},$$

where $\boldsymbol{\gamma}_j$, $j = 1, \cdots, m$, are the rows of $-\boldsymbol{A}_s$ and it can be shown that $\tilde{\boldsymbol{\phi}} = \tilde{\boldsymbol{z}}_s - \tilde{\boldsymbol{\rho}}$, where $\tilde{\boldsymbol{\rho}}$ is the projection of $\tilde{\boldsymbol{z}}_s$ onto the polar cone $\boldsymbol{\Omega}_s^0$. Further, we have the fact that the projection of $\tilde{\boldsymbol{z}}_s$ onto the polar cone is exactly the projection of $\tilde{\boldsymbol{z}}_s$ onto the linear space generated by the edges $\boldsymbol{\gamma}_j$ such that $\langle \tilde{\boldsymbol{z}}_s - \tilde{\boldsymbol{\rho}}, \boldsymbol{\gamma}_j \rangle = 0$. That is, for a given sample s, there is a set $J_s \subseteq \{1, \ldots, m\}$ such that the projection of $\tilde{\boldsymbol{z}}_s$ onto $\boldsymbol{\Omega}_s^0$ coincides with the projection of $\tilde{\boldsymbol{z}}_s$ onto the linear space spanned by $\boldsymbol{\gamma}_j$, for $j \in J_s$. Let $\mathcal{L}_{J_s}$ be the linear space generated by $\boldsymbol{\gamma}_j$, $j \in J_s$ and $\boldsymbol{M}_{J_s}$ be the projection matrix corresponding to this linear space, where $\boldsymbol{M}_{\emptyset}$ is the matrix of zeros by convention. Then

$\tilde{\boldsymbol{\rho}} = \boldsymbol{M}_{J_s}\tilde{\boldsymbol{z}}_s$ and we have:

$$
\begin{aligned}
\tilde{\boldsymbol{\theta}} &= \tilde{\boldsymbol{y}}_s - \boldsymbol{W}_s^{-1/2}\boldsymbol{M}_{J_s}\boldsymbol{W}_s^{1/2}\tilde{\boldsymbol{y}}_s \\
&= \tilde{\boldsymbol{y}}_s - \boldsymbol{W}_s^{-1/2}\boldsymbol{A}'_{s,J_s}(\boldsymbol{A}_{s,J_s}\boldsymbol{A}'_{s,J_s})^{-1}\boldsymbol{A}_{s,J_s}\boldsymbol{W}_s^{1/2}\tilde{\boldsymbol{y}}_s \\
&= \left(\boldsymbol{I}_{D\times D} - \boldsymbol{W}_s^{-1}\boldsymbol{A}'_{J_s}(\boldsymbol{A}_{J_s}\boldsymbol{W}_s^{-1}\boldsymbol{A}'_{J_s})^{-1}\boldsymbol{A}_{J_s}\right)\tilde{\boldsymbol{y}}_s
\end{aligned}
$$

where $\boldsymbol{A}_{s,J_s}$, $\boldsymbol{A}_{J_s}$ denotes the matrix formed by the rows of $\boldsymbol{A}_s$ and $\boldsymbol{A}$ in positions J_s, respectively.

Define $\mathcal{C}_{J_s}$ to be the set of points $\tilde{\boldsymbol{z}}_s \in \mathbb{R}^D$ where the projection of $\tilde{\boldsymbol{z}}_s$ onto the polar cone coincides with the projection onto $\mathcal{L}_{J_s}$, and let $\mathcal{I}_J(s) = 1$ if $\tilde{\boldsymbol{z}}_s \in \mathcal{C}_{J_s}$ and $\mathcal{I}_J(s) = 0$ otherwise. If $\boldsymbol{A}_s$ is not full row rank, then the set J_s may not be unique. Theorem 3.1 from Oliva-Avilés et al. (2020) guarantees, however, that the projection $\tilde{\boldsymbol{\rho}}$ is the same for all such J_s, and that it is always possible to find a minimal J_s^* that is a subset of all J_s such that $\tilde{\boldsymbol{z}}_s \in \mathcal{C}_{J_s}$, and the vectors $\boldsymbol{\gamma}_j$, $j \in J_s^*$ form a linearly independent set. Assuming J_s is this unique set and taking into account that different sample s might correspond to a different J_s, the general expression of the constrained estimator in (3.2) is obtained.

B.1.2 Review of the Covariance Estimator in Oliva-Avilés, et al. (2020)

We derive an expression for the asymptotic variance of the constrained domain mean estimator, similarly to the derivation in Oliva-Avilés et al. (2020). Based on the expression of $\tilde{\boldsymbol{\theta}}$ in (3.2), we have that for each domain d:

$$
\tilde{\theta}_d = \sum_J \left[\tilde{y}_{s_d} - \frac{\hat{N}}{\hat{N}_d}\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_s^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\tilde{\boldsymbol{y}}_s\right\}_d\right]\mathcal{I}_J(s)
$$

where $\tilde{\theta}_d = \tilde{y}_{s_d}$ if $J = \emptyset$. By Taylor expansion, for each domain d, we can linearize the $\tilde{\theta}_d$ as follows:

$$
\tilde{\theta}_d = \sum_J \left[\theta_{d,J} + \sum_{j=1}^{D}\alpha_{dj,J}(\hat{t}_j - t_j) + \sum_{j=1}^{D}\beta_{dj,J}(\hat{N}_j - N_j) + O_p(n^{-1})\right]\mathcal{I}_J(s)
$$

where

$$\theta_{d,J} = \bar{y}_{U_d} - \frac{N}{N_d}\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_d,$$

and $\boldsymbol{W}_U$ is the diagonal matrix with elements $N_1/N, N_2/N, \cdots, N_D/N$. In addition,

$$\hat{t}_j \text{ is the HT estimator of } t_j = \sum_{k\in U_j} y_k,$$

$$\alpha_{dj,J} = \frac{\partial\tilde{\theta}_d}{\partial\hat{t}_j}\Big|_{(\hat{t}_1,\cdots,\hat{t}_D,\hat{N}_1,\cdots,\hat{N}_D)=(t_1,\cdots,t_D,N_1,\cdots,N_D)},$$

and

$$\beta_{dj,J} = \frac{\partial\tilde{\theta}_d}{\partial\hat{N}_j}\Big|_{(\hat{t}_1,\cdots,\hat{t}_D,\hat{N}_1,\cdots,\hat{N}_D)=(t_1,\cdots,t_D,N_1,\cdots,N_D)}.$$

Direct computation through matrix differentiation yields:

$$\alpha_{dd,J} = \frac{1}{N_d} - \frac{N}{N_d^2}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{dd},$$

$$\alpha_{dj,J} = -\frac{N}{N_dN_j}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{dj},$$

$$\begin{aligned}
\beta_{dd,J} &= -\frac{\bar{y}_{U_d}}{N_d} + \frac{N\bar{y}_{U_d}}{N_d^2}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{dd} + \frac{N}{N_d^2}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{d\cdot}\bar{\boldsymbol{y}}_U \\
&\quad - \frac{N^2}{N_d^3}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{dd}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{d\cdot}\bar{\boldsymbol{y}}_U \\
&= -\frac{1}{N_d}\theta_{d,J} + \frac{N}{N_d^2}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{dd}\theta_{d,J},
\end{aligned}$$

and

$$\begin{aligned}
\beta_{dj,J} &= \frac{N\bar{y}_{U_j}}{N_dN_j}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{dj} \\
&\quad - \frac{N^2}{N_dN_j^2}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{dj}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{j\cdot}\bar{\boldsymbol{y}}_U \\
&= \frac{N}{N_dN_j}\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\}_{dj}\theta_{j,J}.
\end{aligned}$$

The variance of the constrained estimator for domain d can be approximated by the variance of $(\sum_{j=1}^{D} \alpha_{dj,J}(\hat{t}_j - t_j) + \sum_{j=1}^{D} \beta_{dj,J}(\hat{N}_j - N_j))$ for the observed set J. The dd'th element of the asymptotic covariance matrix of $\tilde{\boldsymbol{\theta}}$ is given by:

$$\begin{aligned}
\{\mathrm{AV}(\tilde{\boldsymbol{\theta}})\}_{dd'} =& \mathrm{Acov}(\tilde{\theta}_d, \tilde{\theta}_{d'}) \\
=& \mathrm{cov}\left(\sum_{j=1}^{D}\sum_{k\in s_j} \frac{(\alpha_{dj,J}y_k + \beta_{dj,J})}{\pi_k}, \sum_{j=1}^{D}\sum_{l\in s_j} \frac{(\alpha_{d'j,J}y_l + \beta_{d'j,J})}{\pi_l}\right) \\
=& \sum_{j=1}^{D}\left[\sum_{k,l\in U_j} \Delta_{kl}\frac{(\alpha_{dj,J}y_k + \beta_{dj,J})(\alpha_{d'j,J}y_l + \beta_{d'j,J})}{\pi_k\pi_l}\right] \\
&+ \sum_{j\neq i}\sum_{k\in U_j}\sum_{l\in U_i} \Delta_{kl}\frac{(\alpha_{dj,J}y_k + \beta_{dj,J})(\alpha_{d'i,J}y_l + \beta_{d'i,J})}{\pi_k\pi_l}. \qquad \text{(B.1)}
\end{aligned}$$

B.1.3 Some Lemmas

Lemma 5. *Assume $\boldsymbol{A\mu} \geq 0$ and let $J_{\boldsymbol{\mu}} = \{j : \boldsymbol{A}_j\boldsymbol{\mu} = 0\}$, then for $J \neq \emptyset$ and $J \not\subseteq J_{\boldsymbol{\mu}}$, we have that:*

$$P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J) = O\left(n^{-1}\right).$$

In other words, the probability for $J \neq \emptyset$ and $J \not\subseteq J_{\boldsymbol{\mu}}$ has measure 0 asymptotically.

Proof of Lemma 5. If $J \not\subseteq J_{\boldsymbol{\mu}}$, then there must exist $j \in J$, but $j \notin J_{\boldsymbol{\mu}}$. So we have $\boldsymbol{A}_j\tilde{\boldsymbol{y}}_s < 0$ and $\boldsymbol{A}_j\boldsymbol{\mu} > 0$. Using Markov's inequality, we have the following:

$$\begin{aligned}
P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J) &\leq P(-\boldsymbol{A}_j\tilde{\boldsymbol{y}}_s > 0) \\
&= P(-\boldsymbol{A}_j\tilde{\boldsymbol{y}}_s + \boldsymbol{A}_j\boldsymbol{\mu} > \boldsymbol{A}_j\boldsymbol{\mu}) \\
&\leq \frac{\mathrm{E}(\boldsymbol{A}_j\tilde{\boldsymbol{y}}_s - \boldsymbol{A}_j\boldsymbol{\mu})^2}{(\boldsymbol{A}_j\boldsymbol{\mu})^2} \\
&= \frac{\mathrm{E}(\sum_{d=1}^{D} a_{jd}^2(\tilde{y}_{s_d} - \mu_d)^2 + \sum_{d\neq d'} a_{jd}a_{jd'}(\tilde{y}_{s_d} - \mu_d)(\tilde{y}_{s_{d'}} - \mu_{d'}))}{(\boldsymbol{A}_j\boldsymbol{\mu})^2} \\
&= \frac{\sum_{d=1}^{D} a_{jd}^2\mathrm{E}(\tilde{y}_{s_d} - \mu_d)^2 + \sum_{d\neq d'} a_{jd}a_{jd'}\mathrm{E}(\tilde{y}_{s_d} - \mu_d)(\tilde{y}_{s_{d'}} - \mu_{d'})}{(\boldsymbol{A}_j\boldsymbol{\mu})^2}
\end{aligned}$$

where a_{jd} is the jdth element of $\boldsymbol{A}$. Further, we have the following result in survey context.

$$\begin{aligned}
\mathrm{E}(\tilde{y}_{s_d} - \mu_d)^2 &= \mathrm{E}(\tilde{y}_{s_d} - \bar{y}_{U_d} + \bar{y}_{U_d} - \mu_d)^2 \\
&= \mathrm{E}(\tilde{y}_{s_d} - \bar{y}_{U_d})^2 + \mathrm{E}(\bar{y}_{U_d} - \mu_d)^2 + 2\mathrm{E}[(\tilde{y}_{s_d} - \bar{y}_{U_d})(\bar{y}_{U_d} - \mu_d)] \\
&\leq O\left(\frac{1}{n}\right) + O\left(\frac{1}{N}\right) + 2\sqrt{O\left(\frac{1}{n}\right) O\left(\frac{1}{N}\right)} \\
&= O\left(\frac{1}{n}\right)
\end{aligned}$$

And Cauchy Schwarz inequality yields $\mathrm{E}(\tilde{y}_{s_d} - \mu_d)(\tilde{y}_{s_{d'}} - \mu_{d'}) \leq O\left(n^{-1}\right)$, so we finally have:

$$P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J) = O\left(\frac{1}{n}\right).$$

□

By using the same technique, we can prove the following result.

Lemma 6. *Assume $\boldsymbol{A\mu} > 0$ strictly, then for $J \neq \emptyset$, we have that:*

$$P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J) = O\left(n^{-1}\right).$$

In other words, the probability for $J \neq \emptyset$ has measure 0 asymptotically.

Lemma 7. *Under assumptions (A1)-(A5), the constrained estimator $\tilde{\boldsymbol{\theta}}$ is consistent for $\bar{\boldsymbol{y}}_U$ with respect to the sampling mechanism. That is,*

$$\tilde{\boldsymbol{\theta}} = \bar{\boldsymbol{y}}_U + O_p\left(\frac{1}{\sqrt{n}}\right) \tag{B.2}$$

Proof of Lemma 7. For each domain d, we can write $\tilde{\theta}_d - \bar{y}_{U_d}$ as:

$$\begin{aligned}\tilde{\theta}_d - \bar{y}_{U_d} =& (\tilde{y}_{s_d} - \bar{y}_{U_d})\mathcal{I}_{(J=\emptyset)} + \sum_{J\neq\emptyset, J\subseteq J_{\boldsymbol{\mu}}} (\tilde{\theta}_{d,J} - \bar{y}_{U_d})\mathcal{I}_{(J\neq\emptyset, J\subseteq J_{\boldsymbol{\mu}})} \\ &+ \sum_{J\neq\emptyset, J\nsubseteq J_{\boldsymbol{\mu}}} (\tilde{\theta}_{d,J} - \bar{y}_{U_d})\mathcal{I}_{(J\neq\emptyset, J\nsubseteq J_{\boldsymbol{\mu}})}\end{aligned}$$

By Lemma 5, $P(J \neq \emptyset, J \nsubseteq J_{\boldsymbol{\mu}}) = O(n^{-1})$, that is, $\mathcal{I}_{(J\neq\emptyset, J\nsubseteq J_{\boldsymbol{\mu}})} = O_p(n^{-1})$. Also, we know that $\tilde{y}_{s_d} - \bar{y}_{U_d} = O_p(n^{-\frac{1}{2}})$. So we only need to look at the second term of $\tilde{\theta}_d - \bar{y}_{U_d}$. When $J \neq \emptyset$ and $J \subseteq J_{\boldsymbol{\mu}}$,

$$\begin{aligned}\tilde{\theta}_{d,J} - \bar{y}_{U_d} &= (\tilde{y}_{s_d} - \bar{y}_{U_d}) - \{\boldsymbol{W}_s^{-1}\boldsymbol{A}_J'(\boldsymbol{A}_J\boldsymbol{W}_s^{-1}\boldsymbol{A}_J')^{-1}\boldsymbol{A}_J\tilde{\boldsymbol{y}}_s\}_d \\ &= O_p(n^{-\frac{1}{2}}) - \{\boldsymbol{W}_s^{-1}\boldsymbol{A}_J'(\boldsymbol{A}_J\boldsymbol{W}_s^{-1}\boldsymbol{A}_J')^{-1}\boldsymbol{A}_J(\boldsymbol{\mu} + O_p(n^{-\frac{1}{2}}))\}_d \\ &= O_p(n^{-\frac{1}{2}}) + O_p(n^{-\frac{1}{2}}) \\ &= O_p(n^{-\frac{1}{2}})\end{aligned}$$

So overall, $\tilde{\theta}_d - \bar{y}_{U_d} = O_p(n^{-\frac{1}{2}})$ and (B.2) is verified. □

By using the result in Lemma 7, together with the assumption in (A5) that $\bar{y}_{U_d} - \mu_d = O(N^{-\frac{1}{2}})$, we have the following result.

Lemma 8. *Under assumptions (A1)-(A5), the constrained estimator $\tilde{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\mu}$ with respect to the sampling mechanism. That is,*

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\mu} + O_p\left(\frac{1}{\sqrt{n}}\right).$$

B.1.4 Proof of Propostion 1

Proof. For the observed set J, denote $\boldsymbol{B}_J = \boldsymbol{A}'_J[\boldsymbol{A}_J \boldsymbol{W}_U^{-1} \boldsymbol{A}'_J]^{-1} \boldsymbol{A}_J$. Then, we have:

$$\alpha_{dd,J} = \frac{1}{N_d} - \frac{1}{N_d}\left(\frac{N}{N_d}\{\boldsymbol{B}_J\}_{dd}\right), \alpha_{dj,J} = -\frac{N}{N_d N_j}\{\boldsymbol{B}_J\}_{dj} \quad \text{for } d \neq j$$

$$\beta_{dd,J} = -\frac{1}{N_d}\theta_{d,J} + \frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}\theta_{d,J}, \beta_{dj,J} = \frac{N}{N_d N_j}\{\boldsymbol{B}_J\}_{dj}\theta_{j,J} \quad \text{for } d \neq j$$

First, we show that the asymptotic $\text{cov}(\tilde{\theta}_d, \tilde{\theta}_d)$ is the ddth element of $(\boldsymbol{I} - \boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I} - \boldsymbol{P}_J)'$.

$$\begin{aligned}
\text{Acov}(\tilde{\theta}_d, \tilde{\theta}_d) &= \text{cov}\left(\sum_{j=1}^{D} \alpha_{dj,J}\hat{t}_j + \sum_{j=1}^{D} \beta_{dj,J}\hat{N}_j, \sum_{j=1}^{D} \alpha_{dj,J}\hat{t}_j + \sum_{j=1}^{D} \beta_{dj,J}\hat{N}_j\right) \\
&= \text{cov}\left(\sum_{j=1}^{D}\sum_{k \in s_j} \frac{(\alpha_{dj,J}y_k + \beta_{dj,J})}{\pi_k}, \sum_{j=1}^{D}\sum_{l \in s_j} \frac{(\alpha_{dj,J}y_l + \beta_{dj,J})}{\pi_l}\right) \\
&= \text{cov}\left(\sum_{k \in s_d} \frac{(\alpha_{dd,J}y_k + \beta_{dd,J})}{\pi_k}, \sum_{l \in s_d} \frac{(\alpha_{dd,J}y_l + \beta_{dd,J})}{\pi_l}\right) \\
&+ 2\text{cov}\left(\sum_{k \in s_d} \frac{(\alpha_{dd,J}y_k + \beta_{dd,J})}{\pi_k}, \sum_{j \neq d}\sum_{l \in s_j} \frac{(\alpha_{dj,J}y_l + \beta_{dj,J})}{\pi_l}\right) \\
&+ \text{cov}\left(\sum_{j \neq d}\sum_{k \in s_j} \frac{(\alpha_{dj,J}y_k + \beta_{dj,J})}{\pi_k}, \sum_{j \neq d}\sum_{l \in s_j} \frac{(\alpha_{dj,J}y_l + \beta_{dj,J})}{\pi_l}\right)
\end{aligned}$$

where

$$
\begin{aligned}
&\operatorname{cov}\left(\sum_{k\in s_d}\frac{(\alpha_{dd,J}y_k+\beta_{dd,J})}{\pi_k},\sum_{l\in s_d}\frac{(\alpha_{dd,J}y_l+\beta_{dd,J})}{\pi_l}\right)\\
=&\sum_{k,l\in U_d}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[\left(\frac{1}{N_d}-\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}\right)y_k-\frac{1}{N_d}\theta_{d,J}+\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}\theta_{d,J}\right]\\
&\left[\left(\frac{1}{N_d}-\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}\right)y_l-\frac{1}{N_d}\theta_{d,J}+\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}\theta_{d,J}\right]\\
=&\sum_{k,l\in U_d}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[\frac{1}{N_d}(y_k-\theta_{d,J})-\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}(y_k-\theta_{d,J})\right]\left[\frac{1}{N_d}(y_l-\theta_{d,J})-\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}(y_l-\theta_{d,J})\right]\\
=&\{\boldsymbol{\Sigma}_J\}_{dd}-\frac{2N}{N_d}\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{\Sigma}_J\}_{dd}+\left(\frac{N}{N_d}\right)^2\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{\Sigma}_J\}_{dd}\{\boldsymbol{B}_J\}_{dd}\\
=&A_1+A_2+A_3.
\end{aligned}
$$

and

$$
\begin{aligned}
&2\operatorname{cov}\left(\sum_{k\in s_d}\frac{(\alpha_{dd,J}y_k+\beta_{dd,J})}{\pi_k},\sum_{j\neq d}\sum_{l\in s_j}\frac{(\alpha_{dj,J}y_l+\beta_{dj,J})}{\pi_l}\right)\\
=&2\sum_{j\neq d}\sum_{k\in U_d,l\in U_j}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[\frac{1}{N_d}(y_k-\theta_{d,J})-\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}(y_k-\theta_{d,J})\right]\left[-\frac{N}{N_dN_j}\{\boldsymbol{B}_J\}_{dj}(y_l-\theta_{j,J})\right]\\
=&-2\sum_{j\neq d}\frac{N}{N_d}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{dj}+2\sum_{j\neq d}\left(\frac{N}{N_d}\right)^2\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{dj}\\
=&B_1+B_2
\end{aligned}
$$

and the third term is:

$$
\begin{aligned}
&\operatorname{cov}\left(\sum_{j\neq d}\sum_{k\in s_j}\frac{(\alpha_{dj,J}y_k+\beta_{dj,J})}{\pi_k},\sum_{j\neq d}\sum_{l\in s_j}\frac{(\alpha_{dj,J}y_l+\beta_{dj,J})}{\pi_l}\right)\\
=&\sum_{j\neq d}\sum_{k,l\in U_j}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_j}\{\boldsymbol{B}_J\}_{dj}(y_k-\theta_{j,J})\right]\left[-\frac{N}{N_dN_j}\{\boldsymbol{B}_J\}_{dj}(y_l-\theta_{j,J})\right]\\
+&\sum_{j\neq i\neq d}\sum_{k\in U_j,l\in U_i}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_j}\{\boldsymbol{B}_J\}_{dj}(y_k-\theta_{j,J})\right]\left[-\frac{N}{N_dN_i}\{\boldsymbol{B}_J\}_{di}(y_l-\theta_{i,J})\right]\\
=&\sum_{j\neq d}\left(\frac{N}{N_d}\right)^2\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{jj}+\sum_{j\neq i\neq d}\left(\frac{N}{N_d}\right)^2\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{B}_J\}_{di}\{\boldsymbol{\Sigma}_J\}_{ji}\\
=&C_1+C_2
\end{aligned}
$$

Now, we can write the $\{(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\}_{dd}$ as follows:

$$
\begin{aligned}
&\{(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\}_{dd}\\
=&\left(\boldsymbol{e}_d'-\frac{N}{N_d}\{\boldsymbol{B}_J\}_{d\cdot}\right)\boldsymbol{\Sigma}_J\left(\boldsymbol{e}_d-\frac{N}{N_d}\{\boldsymbol{B}_J\}_{\cdot d}\right)\\
=&\{\boldsymbol{\Sigma}_J\}_{dd}-\frac{N}{N_d}\{\boldsymbol{B}_J\}_{d\cdot}\{\boldsymbol{\Sigma}_J\}_{\cdot d}-\frac{N}{N_d}\{\boldsymbol{\Sigma}_J\}_{d\cdot}\{\boldsymbol{B}_J\}_{\cdot d}+\left(\frac{N}{N_d}\right)^2\{\boldsymbol{B}_J\}_{d\cdot}\boldsymbol{\Sigma}_J\{\boldsymbol{B}_J\}_{\cdot d}\\
=&\{\boldsymbol{\Sigma}_J\}_{dd}-2\frac{N}{N_d}\{\boldsymbol{\Sigma}_J\}_{d\cdot}\{\boldsymbol{B}_J\}_{\cdot d}+\left(\frac{N}{N_d}\right)^2\{\boldsymbol{B}_J\}_{d\cdot}\boldsymbol{\Sigma}_J\{\boldsymbol{B}_J\}_{\cdot d}\\
=&I_1^{dd}+I_2^{dd}+I_3^{dd}
\end{aligned}
$$

Obviously, $I_1^{dd}=A_1$.

$$
\begin{aligned}
I_2^{dd}=&-2\frac{N}{N_d}\{\boldsymbol{\Sigma}_J\}_{d\cdot}\{\boldsymbol{B}_J\}_{\cdot d}=-2\frac{N}{N_d}\sum_{j=1}^{D}\{\boldsymbol{\Sigma}_J\}_{dj}\{\boldsymbol{B}_J\}_{jd}\\
=&-2\frac{N}{N_d}\{\boldsymbol{\Sigma}_J\}_{dd}\{\boldsymbol{B}_J\}_{dd}-2\frac{N}{N_d}\sum_{j\neq d}\{\boldsymbol{\Sigma}_J\}_{dj}\{\boldsymbol{B}_J\}_{jd}\\
=&A_2+B_1
\end{aligned}
$$

$$
\begin{aligned}
I_3^{dd} =& \left(\frac{N}{N_d}\right)^2 \{\boldsymbol{B}_J\}_{d\cdot}\boldsymbol{\Sigma}_J\{\boldsymbol{B}_J\}_{\cdot d} = \left(\frac{N}{N_d}\right)^2 \sum_{i=1}^{D}\sum_{j=1}^{D}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{ji}\{\boldsymbol{B}_J\}_{id} \\
=& \left(\frac{N}{N_d}\right)^2 \left[\sum_{i\neq d}\sum_{j=1}^{D}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{ji}\{\boldsymbol{B}_J\}_{id} + \sum_{j=1}^{D}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{jd}\{\boldsymbol{B}_J\}_{dd}\right] \\
=& \left(\frac{N}{N_d}\right)^2 \left[\sum_{i\neq d}\sum_{j\neq d}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{ji}\{\boldsymbol{B}_J\}_{id} + \sum_{i\neq d}\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{\Sigma}_J\}_{di}\{\boldsymbol{B}_J\}_{id}\right. \\
&\left. +\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{\Sigma}_J\}_{dd}\{\boldsymbol{B}_J\}_{dd} + \sum_{j\neq d}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{dj}\{\boldsymbol{B}_J\}_{dd}\right] \\
=& \left(\frac{N}{N_d}\right)^2 \left[\sum_{j\neq d}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{jj}\{\boldsymbol{B}_J\}_{jd} + \sum_{i\neq j\neq d}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{ji}\{\boldsymbol{B}_J\}_{id}\right. \\
&\left. +\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{\Sigma}_J\}_{dd}\{\boldsymbol{B}_J\}_{dd} + 2\sum_{j\neq d}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{dj}\{\boldsymbol{B}_J\}_{dd}\right] \\
=& C_1 + C_2 + A_3 + B_2
\end{aligned}
$$

Hence, we verified that $\{(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\}_{dd} = \text{AV}(\tilde{\theta}_d)$. Now, we will show that the dd'th element of $\{(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\}$ is exactly $\text{Acov}(\tilde{\theta}_d, \tilde{\theta}_{d'})$ for $d \neq d'$.

$$
\begin{aligned}
&\text{Acov}(\tilde{\theta}_d, \tilde{\theta}_{d'}) \\
=&\text{cov}\left(\sum_{j=1}^{D}\sum_{k\in s_j}\frac{(\alpha_{dj,J}y_k + \beta_{dj,J})}{\pi_k}, \sum_{j=1}^{D}\sum_{l\in s_j}\frac{(\alpha_{d'j,J}y_l + \beta_{d'j,J})}{\pi_l}\right) \\
=&\text{cov}\left(\sum_{k\in s_d}\frac{(\alpha_{dd,J}y_k + \beta_{dd,J})}{\pi_k} + \sum_{k\in s_{d'}}\frac{(\alpha_{dd',J}y_k + \beta_{dd',J})}{\pi_k} + \sum_{j\neq d,d'}\sum_{k\in s_j}\frac{(\alpha_{dj,J}y_k + \beta_{dj,J})}{\pi_k},\right. \\
&\left.\sum_{k\in s_d}\frac{(\alpha_{d'd,J}y_k + \beta_{d'd,J})}{\pi_k} + \sum_{k\in s_{d'}}\frac{(\alpha_{d'd',J}y_k + \beta_{d'd',J})}{\pi_k} + \sum_{j\neq d,d'}\sum_{k\in s_j}\frac{(\alpha_{d'j,J}y_k + \beta_{d'j,J})}{\pi_k}\right)
\end{aligned}
$$

We compute above 9 terms one by one as follows:

$$\begin{aligned}
&\text{cov}\left(\sum_{k\in s_d}\frac{(\alpha_{dd,J}y_k+\beta_{dd,J})}{\pi_k},\sum_{l\in s_d}\frac{(\alpha_{d'd,J}y_l+\beta_{d'd,J})}{\pi_l}\right)\\
=&\sum_{k,l\in U_d}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[\frac{1}{N_d}(y_k-\theta_{d,J})-\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}(y_k-\theta_{d,J})\right]\left[-\frac{N}{N_{d'}N_d}\{\boldsymbol{B}_J\}_{d'd}(y_l-\theta_{d,J})\right]\\
=&-\frac{N}{N_{d'}}\{\boldsymbol{B}_J\}_{d'd}\{\boldsymbol{\Sigma}_J\}_{dd}+\frac{N^2}{N_{d'}N_d}\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{B}_J\}_{d'd}\{\boldsymbol{\Sigma}_J\}_{dd}\\
=&D_1+D_2
\end{aligned}$$

$$\begin{aligned}
&\text{cov}\left(\sum_{k\in s_d}\frac{(\alpha_{dd,J}y_k+\beta_{dd,J})}{\pi_k},\sum_{l\in s_{d'}}\frac{(\alpha_{d'd',J}y_l+\beta_{d'd',J})}{\pi_l}\right)\\
=&\sum_{k\in U_d,l\in U_{d'}}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[\frac{1}{N_d}(y_k-\theta_{d,J})-\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}(y_k-\theta_{d,J})\right]\\
&\times\left[\frac{1}{N_{d'}}(y_l-\theta_{d',J})-\frac{N}{N_{d'}^2}\{\boldsymbol{B}_J\}_{d'd'}(y_l-\theta_{d',J})\right]\\
=&\{\boldsymbol{\Sigma}_J\}_{dd'}-\frac{N}{N_{d'}}\{\boldsymbol{B}_J\}_{d'd'}\{\boldsymbol{\Sigma}_J\}_{dd'}-\frac{N}{N_d}\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{\Sigma}_J\}_{dd'}+\frac{N^2}{N_{d'}N_d}\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{B}_J\}_{d'd'}\{\boldsymbol{\Sigma}_J\}_{dd'}\\
=&D_3+D_4+D_5+D_6
\end{aligned}$$

$$\begin{aligned}
&\text{cov}\left(\sum_{k\in s_d}\frac{(\alpha_{dd,J}y_k+\beta_{dd,J})}{\pi_k},\sum_{j\neq d,d'}\sum_{l\in s_j}\frac{(\alpha_{d'j,J}y_l+\beta_{d'j,J})}{\pi_l}\right)\\
=&\sum_{j\neq d,d'}\text{cov}\left(\sum_{k\in s_d}\frac{(\alpha_{dd,J}y_k+\beta_{dd,J})}{\pi_k},\sum_{l\in s_j}\frac{(\alpha_{d'j,J}y_l+\beta_{d'j,J})}{\pi_l}\right)\\
=&\sum_{j\neq d,d'}\sum_{k\in U_d,l\in U_j}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[\frac{1}{N_d}(y_k-\theta_{d,J})-\frac{N}{N_d^2}\{\boldsymbol{B}_J\}_{dd}(y_k-\theta_{d,J})\right]\left[-\frac{N}{N_{d'}N_j}\{\boldsymbol{B}_J\}_{d'j}(y_l-\theta_{j,J})\right]\\
=&\sum_{j\neq d,d'}\left[-\frac{N}{N_{d'}}\{\boldsymbol{B}_J\}_{d'j}\{\boldsymbol{\Sigma}_J\}_{dj}+\frac{N^2}{N_{d'}N_d}\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{B}_J\}_{d'j}\{\boldsymbol{\Sigma}_J\}_{dj}\right]\\
=&D_7+D_8
\end{aligned}$$

$$
\begin{aligned}
&\operatorname{cov}\left(\sum_{k\in s_{d'}}\frac{(\alpha_{dd',J}y_k+\beta_{dd',J})}{\pi_k},\sum_{l\in s_d}\frac{(\alpha_{d'd,J}y_l+\beta_{d'd,J})}{\pi_l}\right)\\
=&\sum_{k\in U_{d'},l\in U_d}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{dd'}(y_k-\theta_{d',J})\right]\left[-\frac{N}{N_{d'}N_d}\{\boldsymbol{B}_J\}_{d'd}(y_l-\theta_{d,J})\right]\\
=&\frac{N^2}{N_{d'}N_d}\{\boldsymbol{B}_J\}_{dd'}\{\boldsymbol{B}_J\}_{d'd}\{\boldsymbol{\Sigma}_J\}_{d'd}\\
=&E_1
\end{aligned}
$$

$$
\begin{aligned}
&\operatorname{cov}\left(\sum_{k\in s_{d'}}\frac{(\alpha_{dd',J}y_k+\beta_{dd',J})}{\pi_k},\sum_{l\in s_{d'}}\frac{(\alpha_{d'd',J}y_l+\beta_{d'd',J})}{\pi_l}\right)\\
=&\sum_{k,l\in U_{d'}}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{dd'}(y_k-\theta_{d',J})\right]\left[\frac{1}{N_{d'}}(y_l-\theta_{d',J})-\frac{N}{N_{d'}^2}\{\boldsymbol{B}_J\}_{d'd'}(y_l-\theta_{d',J})\right]\\
=&-\frac{N}{N_d}\{\boldsymbol{B}_J\}_{dd'}\{\boldsymbol{\Sigma}_J\}_{d'd'}+\frac{N^2}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{dd'}\{\boldsymbol{B}_J\}_{d'd'}\{\boldsymbol{\Sigma}_J\}_{d'd'}\\
=&E_2+E_3
\end{aligned}
$$

$$
\begin{aligned}
&\operatorname{cov}\left(\sum_{k\in s_{d'}}\frac{(\alpha_{dd',J}y_k+\beta_{dd',J})}{\pi_k},\sum_{j\neq d,d'}\sum_{l\in s_j}\frac{(\alpha_{d'j,J}y_l+\beta_{d'j,J})}{\pi_l}\right)\\
=&\sum_{j\neq d,d'}\sum_{k\in U_{d'},l\in U_j}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{dd'}(y_k-\theta_{d',J})\right]\left[-\frac{N}{N_{d'}N_j}\{\boldsymbol{B}_J\}_{d'j}(y_l-\theta_{j,J})\right]\\
=&\sum_{j\neq d,d'}\frac{N^2}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{dd'}\{\boldsymbol{B}_J\}_{d'j}\{\boldsymbol{\Sigma}_J\}_{d'j}\\
=&E_4
\end{aligned}
$$

$$
\begin{aligned}
&\text{cov}\left(\sum_{j\neq d,d'}\sum_{k\in s_j}\frac{(\alpha_{dj,J}y_k+\beta_{dj,J})}{\pi_k},\sum_{l\in s_d}\frac{(\alpha_{d'd,J}y_l+\beta_{d'd,J})}{\pi_l}\right)\\
=&\sum_{j\neq d,d'}\sum_{k\in U_j,l\in U_d}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_j}\{\boldsymbol{B}_J\}_{dj}(y_k-\theta_{j,J})\right]\left[-\frac{N}{N_{d'}N_d}\{\boldsymbol{B}_J\}_{d'd}(y_l-\theta_{d,J})\right]\\
=&\sum_{j\neq d,d'}\frac{N^2}{N_{d'}N_d}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{B}_J\}_{d'd}\{\boldsymbol{\Sigma}_J\}_{jd}\\
=&F_1
\end{aligned}
$$

$$
\begin{aligned}
&\text{cov}\left(\sum_{j\neq d,d'}\sum_{k\in s_j}\frac{(\alpha_{dj,J}y_k+\beta_{dj,J})}{\pi_k},\sum_{l\in s_{d'}}\frac{(\alpha_{d'd',J}y_l+\beta_{d'd',J})}{\pi_l}\right)\\
=&\sum_{j\neq d,d'}\sum_{k\in U_j,l\in U_{d'}}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_j}\{\boldsymbol{B}_J\}_{dj}(y_k-\theta_{j,J})\right]\left[\frac{1}{N_{d'}}(y_l-\theta_{d',J})-\frac{N}{N_{d'}^2}\{\boldsymbol{B}_J\}_{d'd'}(y_l-\theta_{d',J})\right]\\
=&\sum_{j\neq d,d'}\left[-\frac{N}{N_d}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{jd'}+\frac{N^2}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{B}_J\}_{d'd'}\{\boldsymbol{\Sigma}_J\}_{jd'}\right]\\
=&F_2+F_3
\end{aligned}
$$

$$
\begin{aligned}
&\operatorname{cov}\left(\sum_{j\neq d,d'}\sum_{k\in s_j}\frac{(\alpha_{dj,J}y_k+\beta_{dj,J})}{\pi_k},\sum_{j\neq d,d'}\sum_{l\in s_j}\frac{(\alpha_{d'j,J}y_l+\beta_{d'j,J})}{\pi_l}\right)\\
=&\sum_{j\neq d,d'}\operatorname{cov}\left(\sum_{k\in s_j}\frac{(\alpha_{dj,J}y_k+\beta_{dj,J})}{\pi_k},\sum_{l\in s_j}\frac{(\alpha_{d'j,J}y_l+\beta_{d'j,J})}{\pi_l}\right)+\\
&\sum_{j\neq i\neq d,d'}\operatorname{cov}\left(\sum_{k\in s_j}\frac{(\alpha_{dj,J}y_k+\beta_{dj,J})}{\pi_k},\sum_{l\in s_i}\frac{(\alpha_{d'i,J}y_l+\beta_{d'i,J})}{\pi_l}\right)\\
=&\sum_{j\neq d,d'}\sum_{k,l\in U_j}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_j}\{\boldsymbol{B}_J\}_{dj}(y_k-\theta_{j,J})\right]\left[-\frac{N}{N_{d'}N_j}\{\boldsymbol{B}_J\}_{d'j}(y_l-\theta_{j,J})\right]\\
&+\sum_{j\neq i\neq d,d'}\sum_{k\in U_j,l\in U_i}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[-\frac{N}{N_dN_j}\{\boldsymbol{B}_J\}_{dj}(y_k-\theta_{j,J})\right]\left[-\frac{N}{N_{d'}N_i}\{\boldsymbol{B}_J\}_{d'i}(y_l-\theta_{i,J})\right]\\
=&\sum_{j\neq d,d'}\frac{N^2}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{B}_J\}_{d'j}\{\boldsymbol{\Sigma}_J\}_{jj}+\sum_{j\neq i\neq d,d'}\frac{N^2}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{B}_J\}_{d'i}\{\boldsymbol{\Sigma}_J\}_{ji}\\
=&F_4+F_5
\end{aligned}
$$

So, for $d\neq d'$, we have that:

$$
\begin{aligned}
\text{Acov}(\tilde{\theta}_d,\tilde{\theta}_{d'})=&D_1+D_2+D_3+D_4+D_5+D_6+D_7+D_8\\
&+E_1+E_2+E_3+E_4\\
&+F_1+F_2+F_3+F_4+F_5
\end{aligned}
$$

Now, we compute the dd'th element of $(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'$ as follows:

$$
\begin{aligned}
&\{(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\}_{dd'}\\
=&\left(\boldsymbol{e}_d'-\frac{N}{N_d}\{\boldsymbol{B}_J\}_{d\cdot}\right)\boldsymbol{\Sigma}_J\left(\boldsymbol{e}_{d'}-\frac{N}{N_{d'}}\{\boldsymbol{B}_J\}_{\cdot d'}\right)\\
=&\{\boldsymbol{\Sigma}_J\}_{dd'}-\frac{N}{N_d}\{\boldsymbol{B}_J\}_{d\cdot}\{\boldsymbol{\Sigma}_J\}_{\cdot d'}-\frac{N}{N_{d'}}\{\boldsymbol{\Sigma}_J\}_{d\cdot}\{\boldsymbol{B}_J\}_{\cdot d'}+\frac{N^2}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{d\cdot}\boldsymbol{\Sigma}_J\{\boldsymbol{B}_J\}_{\cdot d'}\\
=&I_1^{dd'}+I_2^{dd'}+I_3^{dd'}+I_4^{dd'}
\end{aligned}
$$

It is clear that we have $I_1^{dd'} = D_3$ directly.

$$
\begin{aligned}
I_2^{dd'} =& -\frac{N}{N_d}\{\boldsymbol{B}_J\}_{d\cdot}\{\boldsymbol{\Sigma}_J\}_{\cdot d'} = -\frac{N}{N_d}\sum_{j=1}^{D}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{jd'} \\
=& -\frac{N}{N_d}\left[\{\boldsymbol{B}_J\}_{dd}\{\boldsymbol{\Sigma}_J\}_{dd'} + \{\boldsymbol{B}_J\}_{dd'}\{\boldsymbol{\Sigma}_J\}_{d'd'} + \sum_{j\neq d,d'}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{jd'}\right] \\
=& D_5 + E_2 + F_2
\end{aligned}
$$

$$
\begin{aligned}
I_3^{dd'} =& -\frac{N}{N_{d'}}\{\boldsymbol{\Sigma}_J\}_{d\cdot}\{\boldsymbol{B}_J\}_{\cdot d'} = -\frac{N}{N_{d'}}\sum_{j=1}^{D}\{\boldsymbol{\Sigma}_J\}_{dj}\{\boldsymbol{B}_J\}_{jd'} \\
=& -\frac{N}{N_{d'}}\left[\{\boldsymbol{\Sigma}_J\}_{dd}\{\boldsymbol{B}_J\}_{dd'} + \{\boldsymbol{\Sigma}_J\}_{dd'}\{\boldsymbol{B}_J\}_{d'd'} + \sum_{j\neq d,d'}\{\boldsymbol{\Sigma}_J\}_{dj}\{\boldsymbol{B}_J\}_{jd'}\right] \\
=& -\frac{N}{N_{d'}}\left[\{\boldsymbol{B}_J\}_{d'd}\{\boldsymbol{\Sigma}_J\}_{dd} + \{\boldsymbol{B}_J\}_{d'd'}\{\boldsymbol{\Sigma}_J\}_{dd'} + \sum_{j\neq d,d'}\{\boldsymbol{B}_J\}_{d'j}\{\boldsymbol{\Sigma}_J\}_{dj}\right] \\
=& D_1 + D_4 + D_7
\end{aligned}
$$

$$
\begin{aligned}
I_4^{dd'} =& \frac{N^2}{N_dN_{d'}}\{\boldsymbol{B}_J\}_{d\cdot}\boldsymbol{\Sigma}_J\{\boldsymbol{B}_J\}_{\cdot d'} = \frac{N^2}{N_dN_{d'}}\sum_{i=1}^{D}\sum_{j=1}^{D}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{ji}\{\boldsymbol{B}_J\}_{id'} \\
=& \frac{N^2}{N_dN_{d'}}\left[\sum_{j=1}^{D}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{jd}\{\boldsymbol{B}_J\}_{dd'} + \sum_{j=1}^{D}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{jd'}\{\boldsymbol{B}_J\}_{d'd'}\right. \\
&\left. + \sum_{i\neq d,d'}\sum_{j=1}^{D}\{\boldsymbol{B}_J\}_{dj}\{\boldsymbol{\Sigma}_J\}_{ji}\{\boldsymbol{B}_J\}_{id'}\right] \\
=& I_{41}^{dd'} + I_{42}^{dd'} + I_{43}^{dd'}
\end{aligned}
$$

where:

$$
\begin{aligned}
I_{41}^{dd'} =& \frac{N^2}{N_d N_{d'}} \sum_{j=1}^{D} \{\boldsymbol{B}_J\}_{dj} \{\boldsymbol{\Sigma}_J\}_{jd} \{\boldsymbol{B}_J\}_{dd'} \\
=& \frac{N^2}{N_d N_{d'}} \left[\{\boldsymbol{B}_J\}_{dd} \{\boldsymbol{\Sigma}_J\}_{dd} \{\boldsymbol{B}_J\}_{dd'} + \{\boldsymbol{B}_J\}_{dd'} \{\boldsymbol{\Sigma}_J\}_{d'd} \{\boldsymbol{B}_J\}_{dd'} \right. \\
& \left. + \sum_{j \neq d, d'} \{\boldsymbol{B}_J\}_{dj} \{\boldsymbol{\Sigma}_J\}_{jd} \{\boldsymbol{B}_J\}_{dd'} \right] \\
=& D_2 + E_1 + F_1
\end{aligned}
$$

Similarly, we have:

$$
\begin{aligned}
I_{42}^{dd'} =& \frac{N^2}{N_d N_{d'}} \sum_{j=1}^{D} \{\boldsymbol{B}_J\}_{dj} \{\boldsymbol{\Sigma}_J\}_{jd'} \{\boldsymbol{B}_J\}_{d'd'} \\
=& \frac{N^2}{N_d N_{d'}} \left[\{\boldsymbol{B}_J\}_{dd} \{\boldsymbol{\Sigma}_J\}_{dd'} \{\boldsymbol{B}_J\}_{d'd'} + \{\boldsymbol{B}_J\}_{dd'} \{\boldsymbol{\Sigma}_J\}_{d'd'} \{\boldsymbol{B}_J\}_{d'd'} \right. \\
& \left. + \sum_{j \neq d, d'} \{\boldsymbol{B}_J\}_{dj} \{\boldsymbol{\Sigma}_J\}_{jd'} \{\boldsymbol{B}_J\}_{d'd'} \right] \\
=& D_6 + E_3 + F_3
\end{aligned}
$$

$$
\begin{aligned}
I_{43}^{dd'} =& \frac{N^2}{N_d N_{d'}} \sum_{i \neq d, d'} \sum_{j=1}^{D} \{\boldsymbol{B}_J\}_{dj} \{\boldsymbol{\Sigma}_J\}_{ji} \{\boldsymbol{B}_J\}_{id'} \\
=& \frac{N^2}{N_d N_{d'}} \left[\sum_{i \neq d, d'} \{\boldsymbol{B}_J\}_{dd} \{\boldsymbol{\Sigma}_J\}_{di} \{\boldsymbol{B}_J\}_{id'} + \sum_{i \neq d, d'} \{\boldsymbol{B}_J\}_{dd'} \{\boldsymbol{\Sigma}_J\}_{d'i} \{\boldsymbol{B}_J\}_{id'} \right. \\
& \left. + \sum_{i \neq d, d'} \sum_{j \neq d, d'} \{\boldsymbol{B}_J\}_{dj} \{\boldsymbol{\Sigma}_J\}_{ji} \{\boldsymbol{B}_J\}_{id'} \right] \\
=& \frac{N^2}{N_d N_{d'}} \left[\sum_{j \neq d, d'} \{\boldsymbol{B}_J\}_{dd} \{\boldsymbol{\Sigma}_J\}_{dj} \{\boldsymbol{B}_J\}_{jd'} + \sum_{j \neq d, d'} \{\boldsymbol{B}_J\}_{dd'} \{\boldsymbol{\Sigma}_J\}_{d'j} \{\boldsymbol{B}_J\}_{jd'} \right. \\
& \left. + \sum_{j \neq d, d'} \{\boldsymbol{B}_J\}_{dj} \{\boldsymbol{\Sigma}_J\}_{jj} \{\boldsymbol{B}_J\}_{jd'} + \sum_{i \neq j \neq d, d'} \{\boldsymbol{B}_J\}_{dj} \{\boldsymbol{\Sigma}_J\}_{ji} \{\boldsymbol{B}_J\}_{id'} \right] \\
=& D_8 + E_4 + F_4 + F_5
\end{aligned}
$$

Hence, we verified that $\{(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\}_{dd'} = \mathrm{Acov}(\tilde{\theta}_d, \tilde{\theta}_{d'})$, and thus overall, we have:

$$AV(\tilde{\boldsymbol{\theta}}) = (\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'$$

This completes the verification of (3.3). □

B.2 Proof of Theorem 8

Proof. Define $J_{\boldsymbol{\mu}} = \{j : \boldsymbol{A}_j\boldsymbol{\mu} = 0\}$. Then, we can write $n(\hat{V}(\tilde{\boldsymbol{\theta}}) - AV(\tilde{\boldsymbol{\theta}}))$ as follows:

$$\begin{aligned}
&n(\hat{V}(\tilde{\boldsymbol{\theta}}) - AV(\tilde{\boldsymbol{\theta}})) \\
=&n\left[\sum_J (\boldsymbol{I}-\hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)'\mathcal{I}_J(s) - \sum_J (\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\mathcal{I}_J(s)\right] \\
=&\sum_J \left[(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)n\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)' - (\boldsymbol{I}-\boldsymbol{P}_J)n\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\right]\mathcal{I}_J(s) \\
=&n(\tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\mathcal{I}_\emptyset(s) + \sum_{J\neq\emptyset, J\subseteq J_{\boldsymbol{\mu}}} \left[(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)n\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)' - (\boldsymbol{I}-\boldsymbol{P}_J)n\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\right]\mathcal{I}_J(s) \\
&+ \sum_{J\neq\emptyset, J\not\subseteq J_{\boldsymbol{\mu}}} \left[(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)n\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)' - (\boldsymbol{I}-\boldsymbol{P}_J)n\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\right]\mathcal{I}_J(s) \\
=&I_1 + I_2 + I_3
\end{aligned}$$

For I_1, we will prove the convergence of diagonal and non-diagonal elements of $\tilde{\boldsymbol{\Sigma}}$ separately.

Let us consider the diagonal element first. Denote $\hat{\boldsymbol{\Sigma}}_{dd} = \frac{1}{N_d^2}\sum_{k,l\in s_d} \frac{\Delta_{kl}}{\pi_{kl}} \frac{(y_k - \tilde{y}_{s_d})(y_l - \tilde{y}_{s_d})}{\pi_k \pi_l}$. We will prove $n(\hat{\boldsymbol{\Sigma}}_{dd} - \boldsymbol{\Sigma}_{dd}) = o_p(1)$ by showing:

$$n\mathrm{E}|\hat{\boldsymbol{\Sigma}}_{dd} - \boldsymbol{\Sigma}_{dd}| \to 0 \quad \text{as} \quad N_d \to \infty \tag{B.3}$$

where

$$
\begin{aligned}
& n\mathrm{E}|\hat{\mathbf{\Sigma}}_{dd} - \mathbf{\Sigma}_{dd}| \\
=& n\mathrm{E}\left|\frac{1}{N_d^2}\sum_{k,l\in U_d}\frac{\Delta_{kl}}{\pi_{kl}}\frac{(y_k - \bar{y}_{U_d} + \bar{y}_{U_d} - \tilde{y}_{s_d})(y_l - \bar{y}_{U_d} + \bar{y}_{U_d} - \tilde{y}_{s_d})}{\pi_k\pi_l} I_k I_l \right. \\
& \left. - \frac{1}{N_d^2}\sum_{k,l\in U_d}\Delta_{kl}\frac{(y_k - \bar{y}_{U_d})(y_l - \bar{y}_{U_d})}{\pi_k\pi_l}\right| \\
\leq& n\mathrm{E}\left|\frac{1}{N_d^2}\sum_{k,l\in U_d}\Delta_{kl}\frac{(y_k - \bar{y}_{U_d})(y_l - \bar{y}_{U_d})}{\pi_k\pi_l}\left(\frac{I_k I_l}{\pi_{kl}} - 1\right)\right| \\
& + n\mathrm{E}\left|\frac{1}{N_d^2}\sum_{k,l\in U_d}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[2(y_k - \bar{y}_{U_d})(\bar{y}_{U_d} - \tilde{y}_{s_d}) + (\bar{y}_{U_d} - \tilde{y}_{s_d})^2\right]\frac{I_k I_l}{\pi_{kl}}\right| \\
=& A^{dd} + B^{dd}
\end{aligned}
$$

Now,

$$
\begin{aligned}
& n^2\mathrm{E}\left(\frac{1}{N_d^2}\sum_{k,l\in U_d}\Delta_{kl}\frac{(y_k - \bar{y}_{U_d})(y_l - \bar{y}_{U_d})}{\pi_k\pi_l}\left(\frac{I_k I_l}{\pi_{kl}} - 1\right)\right)^2 \\
=& n^2\sum_{k,i\in U_d}\frac{1-\pi_k}{\pi_k}\frac{1-\pi_i}{\pi_i}\frac{(y_k - \bar{y}_{U_d})^2(y_i - \bar{y}_{U_d})^2}{N_d^4}\frac{\pi_{ki} - \pi_k\pi_i}{\pi_k\pi_i} \\
& + 2n^2\sum_{k\in U_d}\sum_{i,j\in U_d, i\neq j}\frac{1-\pi_k}{\pi_k}\frac{\Delta_{ij}}{\pi_i\pi_j}\frac{(y_k - \bar{y}_{U_d})^2(y_i - \bar{y}_{U_d})(y_j - \bar{y}_{U_d})}{N_d^4}\mathrm{E}\left(\frac{I_k - \pi_k}{\pi_k}\frac{I_i I_j - \pi_{ij}}{\pi_{ij}}\right) \\
& + n^2\sum_{k,l\in U_d, k\neq l}\sum_{i,j\in U_d, i\neq j}\frac{\Delta_{kl}}{\pi_k\pi_l}\frac{\Delta_{ij}}{\pi_i\pi_j}\frac{(y_k - \bar{y}_{U_d})(y_l - \bar{y}_{U_d})(y_i - \bar{y}_{U_d})(y_j - \bar{y}_{U_d})}{N_d^4} \\
& \times \mathrm{E}\left(\frac{I_k I_l - \pi_{kl}}{\pi_{kl}}\frac{I_i I_j - \pi_{ij}}{\pi_{ij}}\right) \\
=& A_1^{dd} + A_2^{dd} + A_3^{dd}
\end{aligned}
$$

But

$$
\begin{aligned}
A_1^{dd} &= \frac{n^2}{N_d^2}\frac{1}{N_d^2}\sum_{k\in U_d}\left(\frac{1-\pi_k}{\pi_k}\right)^3 (y_k-\bar{y}_{U_d})^4 \\
&+ \frac{n^2}{N_d^2}\frac{1}{N_d^2}\sum_{k,i\in U_d,k\neq i}\frac{1-\pi_k}{\pi_k}\frac{1-\pi_i}{\pi_i}\left(y_k-\bar{y}_{U_d}\right)^2\left(y_i-\bar{y}_{U_d}\right)^2\frac{\Delta_{ki}}{\pi_k\pi_i} \\
&\leq \left[\frac{1}{N_d\lambda_1^3}\left(\frac{n^2}{N_d^2}\right)\right]\frac{1}{N_d}\sum_{k\in U_d}(y_k-\bar{y}_{U_d})^4 \\
&+ \left[\frac{1}{N_d\lambda_1^4}\left(\frac{n}{N_d}\right)n\max_{k,i\in U_d,k\neq i}|\Delta_{ki}|\right]\frac{\sum_{k\in U_d}(y_k-\bar{y}_{U_d})^4}{N_d}
\end{aligned}
$$

which goes to 0 as $N_d\to\infty$, and:

$$
\begin{aligned}
A_3^{dd} \leq& \frac{(n\max_{k,l\in U_d,k\neq l}|\Delta_{kl}|)^2}{\lambda_1^4}\sum_{k,l\in U_d,k\neq l}\sum_{i,j\in U_d,i\neq j}\frac{|(y_k-\bar{y}_{U_d})(y_l-\bar{y}_{U_d})(y_i-\bar{y}_{U_d})(y_j-\bar{y}_{U_d})|}{N_d^4} \\
&\times\left|\mathrm{E}\left(\frac{I_kI_l-\pi_{kl}}{\pi_{kl}}\frac{I_iI_j-\pi_{ij}}{\pi_{ij}}\right)\right| \\
\leq& O(N_d^{-1}) + \frac{(n\max_{k,l\in U_d,k\neq l}|\Delta_{kl}|)^2}{\lambda_1^4\lambda_2^2}\max_{(k,l,i,j)\in D_4,N_d}\left|\mathrm{E}\left[(I_kI_l-\pi_{kl})(I_iI_j-\pi_{ij})\right]\right| \\
&\times\frac{1}{N_d}\sum_{k\in N_d}(y_k-\bar{y}_{U_d})^4
\end{aligned}
$$

which converges to 0 by assumption (A6) as $N_d\to\infty$. Also, the Cauchy Schwarz inequality implies that $A_2^{dd}\to 0$ as $N_d\to\infty$, and thus it follows that $A^{dd}\to 0$ as $N_d\to\infty$.

Next,

$$
\begin{aligned}
B^{dd} =& n\mathrm{E}\left|\frac{1}{N_d^2}\sum_{k,l\in U_d}\frac{\Delta_{kl}}{\pi_k\pi_l}\left[2(y_k-\bar{y}_{U_d})(\bar{y}_{U_d}-\tilde{y}_{s_d})+(\bar{y}_{U_d}-\tilde{y}_{s_d})^2\right]\frac{I_kI_l}{\pi_{kl}}\right| \\
&\leq\left(\frac{2n}{N_d\lambda_1^2}+\frac{2n\max_{k,l\in U_d,k\neq l}|\Delta_{kl}|}{\lambda_1^2\lambda_2}\right)\left[\frac{\sum_{k\in U_d}(y_k-\bar{y}_{U_d})^2}{N_d}\mathrm{E}(\tilde{y}_{s_d}-\bar{y}_{U_d})^2\right]^{\frac{1}{2}} \\
&+\left(\frac{n}{N_d\lambda_1^2}+\frac{n\max_{k,l\in U_d,k\neq l}|\Delta_{kl}|}{\lambda_1^2\lambda_2}\right)\mathrm{E}(\tilde{y}_{s_d}-\bar{y}_{U_d})^2 \\
&\to 0
\end{aligned}
$$

using the fact that $\mathrm{E}(\tilde{y}_{s_d} - \bar{y}_{U_d})^2 \to 0$ as $N_d \to \infty$. So, (B.3) is verified. Hence, we have:

$$\begin{aligned} n(\tilde{\mathbf{\Sigma}}_{dd} - \mathbf{\Sigma}_{dd}) &= n(\tilde{\mathbf{\Sigma}}_{dd} - \hat{\mathbf{\Sigma}}_{dd}) + n(\hat{\mathbf{\Sigma}}_{dd} - \mathbf{\Sigma}_{dd}) \\ &= \left(\frac{N_d^2}{\hat{N}_d^2} - 1\right)\left(\frac{n}{N_d^2}\sum_{k,l\in s_d}\frac{\Delta_{kl}}{\pi_{kl}}\frac{(y_k - \tilde{y}_{s_d})(y_l - \tilde{y}_{s_d})}{\pi_k\pi_l}\right) + o_p(1) \\ &= o_p(1)O_p(1) + o_p(1) = o_p(1) \end{aligned}$$

Now, when $d \neq d', d, d' = 1, \cdots, D$, define

$$\hat{\mathbf{\Sigma}}_{dd'} = \frac{1}{N_d N_{d'}}\sum_{k\in s_d}\sum_{l\in s_{d'}}\frac{\Delta_{kl}}{\pi_{kl}}\frac{(y_k - \tilde{y}_{s_d})(y_l - \tilde{y}_{s_{d'}})}{\pi_k\pi_l}$$

We will prove $n(\hat{\mathbf{\Sigma}}_{dd'} - \mathbf{\Sigma}_{dd'}) = o_p(1)$ by showing:

$$n\mathrm{E}|\hat{\mathbf{\Sigma}}_{dd'} - \mathbf{\Sigma}_{dd'}| \to 0 \quad \text{as} \quad N_d, N_{d'} \to \infty \tag{B.4}$$

and

$$\begin{aligned} &n\mathrm{E}|\hat{\mathbf{\Sigma}}_{dd'} - \mathbf{\Sigma}_{dd'}| \\ =&n\mathrm{E}\left|\frac{1}{N_d N_{d'}}\sum_{k\in s_d}\sum_{l\in s_{d'}}\frac{\Delta_{kl}}{\pi_{kl}}\frac{(y_k - \tilde{y}_{s_d})(y_l - \tilde{y}_{s_{d'}})}{\pi_k\pi_l} - \frac{1}{N_d N_{d'}}\sum_{k\in U_d}\sum_{l\in U_{d'}}\Delta_{kl}\frac{(y_k - \bar{y}_{U_d})(y_l - \bar{y}_{U_{d'}})}{\pi_k\pi_l}\right| \\ \leq&n\mathrm{E}\left|\frac{1}{N_d N_{d'}}\sum_{k\in U_d}\sum_{l\in U_{d'}}\frac{\Delta_{kl}}{\pi_k\pi_l}(y_k - \bar{y}_{U_d})(y_l - \bar{y}_{U_{d'}})\left(\frac{I_k I_l}{\pi_{kl}} - 1\right)\right| \\ &+ n\mathrm{E}\left|\frac{1}{N_d N_{d'}}\sum_{k\in U_d}\sum_{l\in U_{d'}}\frac{\Delta_{kl}}{\pi_k\pi_l}[(y_k - \bar{y}_{U_d})(\bar{y}_{U_{d'}} - \tilde{y}_{s_{d'}})\right. \\ &\left.+(y_l - \bar{y}_{U_{d'}})(\bar{y}_{U_d} - \tilde{y}_{s_d}) + (\bar{y}_{U_d} - \tilde{y}_{s_d})(\bar{y}_{U_{d'}} - \tilde{y}_{s_{d'}})]\frac{I_k I_l}{\pi_{kl}}\right| \\ =&A^{dd'} + B^{dd'} \end{aligned}$$

Now,

$$
\begin{aligned}
&n^2\mathrm{E}\left(\frac{1}{N_dN_{d'}}\sum_{k\in U_d}\sum_{l\in U_{d'}}\frac{\Delta_{kl}}{\pi_k\pi_l}(y_k-\bar{y}_{U_d})(y_l-\bar{y}_{U_{d'}})\left(\frac{I_kI_l}{\pi_{kl}}-1\right)\right)^2\\
=&\frac{n^2}{N_d^2N_{d'}^2}\mathrm{E}\left[\sum_{k\in U_d}\sum_{l\in U_{d'}}\left(\frac{\Delta_{kl}}{\pi_k\pi_l}\right)^2(y_k-\bar{y}_{U_d})^2(y_l-\bar{y}_{U_{d'}})^2\left(\frac{I_kI_l}{\pi_{kl}}-1\right)^2\right]\\
&+\frac{n^2}{N_d^2N_{d'}^2}\mathrm{E}\left[\sum_{k\in U_d}\sum_{l,j\in U_{d'},l\neq j}\left(\frac{\Delta_{kl}\Delta_{kj}}{\pi_k^2\pi_l\pi_j}\right)(y_k-\bar{y}_{U_d})^2(y_l-\bar{y}_{U_{d'}})\right.\\
&\left.(y_j-\bar{y}_{U_{d'}})\left(\frac{I_kI_l}{\pi_{kl}}-1\right)\left(\frac{I_kI_j}{\pi_{kj}}-1\right)\right]\\
&+\frac{n^2}{N_d^2N_{d'}^2}\mathrm{E}\left[\sum_{k,i\in U_d,k\neq i}\sum_{l\in U_{d'}}\left(\frac{\Delta_{kl}\Delta_{il}}{\pi_k\pi_i\pi_l^2}\right)(y_k-\bar{y}_{U_d})(y_i-\bar{y}_{U_d})\right.\\
&\left.(y_l-\bar{y}_{U_{d'}})^2\left(\frac{I_kI_l}{\pi_{kl}}-1\right)\left(\frac{I_iI_l}{\pi_{il}}-1\right)\right]\\
&+\frac{n^2}{N_d^2N_{d'}^2}\mathrm{E}\left[\sum_{k,i\in U_d,k\neq i}\sum_{l,j\in U_{d'},l\neq j}\left(\frac{\Delta_{kl}\Delta_{ij}}{\pi_k\pi_l\pi_i\pi_j}\right)(y_k-\bar{y}_{U_d})(y_l-\bar{y}_{U_{d'}})\right.\\
&\left.(y_i-\bar{y}_{U_d})(y_j-\bar{y}_{U_{d'}})\left(\frac{I_kI_l}{\pi_{kl}}-1\right)\left(\frac{I_iI_j}{\pi_{ij}}-1\right)\right]\\
&=A_1^{dd'}+A_2^{dd'}+A_3^{dd'}+A_4^{dd'}
\end{aligned}
$$

where:

$$
\begin{aligned}
A_1^{dd'}&\leq\frac{(n\max_{k,l\in U_d,k\neq l}|\Delta_{kl}|)^2}{N_dN_{d'}\lambda_1^4}\frac{1}{N_dN_{d'}}\sum_{k\in U_d}\sum_{l\in U_{d'}}(y_k-\bar{y}_{U_d})^2(y_l-\bar{y}_{U_{d'}})^2\mathrm{E}\left(\frac{I_kI_l}{\pi_{kl}}-1\right)^2\\
&\leq\frac{(n\max_{k,l\in U_d,k\neq l}|\Delta_{kl}|)^2}{N_dN_{d'}\lambda_1^4\lambda_2^2}\left(\frac{1}{N_d}\sum_{k\in U_d}(y_k-\bar{y}_{U_d})^2\right)\left(\frac{1}{N_{d'}}\sum_{l\in U_{d'}}(y_l-\bar{y}_{U_{d'}})^2\right)\\
&\to 0
\end{aligned}
$$

as $N_d, N_{d'}\to\infty$. $A_2^{dd'}$ can be bounded as follows:

$$
\begin{aligned}
A_2^{dd'} \leq & \frac{n^2 \max_{k,l\in U, k\neq l} |\Delta_{kl}| \max_{k,j\in U, k\neq j} |\Delta_{kj}|}{N_d \lambda_1^4} \frac{1}{N_d N_{d'}^2} \sum_{k\in U_d} \sum_{l,j\in U_{d'}, l\neq j} (y_k - \bar{y}_{U_d})^2 \\
& |(y_l - \bar{y}_{U_{d'}})(y_j - \bar{y}_{U_{d'}})| \left| \mathrm{E}\left[\left(\frac{I_k I_l}{\pi_{kl}} - 1\right)\left(\frac{I_k I_j}{\pi_{kj}} - 1\right)\right]\right| \\
& \leq \frac{n^2 \max_{k,l\in U, k\neq l} |\Delta_{kl}| \max_{k,j\in U, k\neq j} |\Delta_{kj}|}{N_d \lambda_1^4 \lambda_2^2} \left(\frac{1}{N_d}\sum_{k\in U_d} (y_k - \bar{y}_{U_d})^2\right) \\
& \times \left[\frac{1}{N_{d'}^2}\left(\sum_{l\in U_{d'}} |y_l - \bar{y}_{U_{d'}}|\right)^2\right] \\
& \leq \frac{n^2 \max_{k,l\in U, k\neq l} |\Delta_{kl}| \max_{k,j\in U, k\neq j} |\Delta_{kj}|}{N_d \lambda_1^4 \lambda_2^2} \left(\frac{1}{N_d}\sum_{k\in U_d} (y_k - \bar{y}_{U_d})^2\right)\left(\frac{1}{N_{d'}}\sum_{k\in U_{d'}} (y_k - \bar{y}_{U_{d'}})^2\right)
\end{aligned}
$$

which goes to 0 as $N_d \to \infty$. Similarly, we have $A_3^{dd'} \to 0$ as $N_{d'} \to \infty$. Now, we bound $A_4^{dd'}$ as follows:

$$
\begin{aligned}
A_4^{dd'} \leq & \frac{n^2 \max_{k,l\in U, k\neq l} |\Delta_{kl}| \max_{i,j\in U, i\neq j} |\Delta_{ij}|}{\lambda_1^4} \frac{1}{N_d^2 N_{d'}^2} \sum_{k,i\in U_d, k\neq i} \sum_{l,j\in U_{d'}, l\neq j} |(y_k - \bar{y}_{U_d})(y_l - \bar{y}_{U_{d'}}) \\
& (y_i - \bar{y}_{U_d})(y_j - \bar{y}_{U_{d'}})| \left| \mathrm{E}\left[\left(\frac{I_k I_l}{\pi_{kl}} - 1\right)\left(\frac{I_i I_j}{\pi_{ij}} - 1\right)\right]\right| \\
& \leq \frac{n^2 \max_{k,l\in U, k\neq l} |\Delta_{kl}| \max_{i,j\in U, i\neq j} |\Delta_{ij}|}{\lambda_1^4 \lambda_2^2} \max_{(k,l,i,j)\in D_{4,N}} |\mathrm{E}\left[(I_k I_l - 1)(I_i I_j - 1)\right]| \\
& \left(\frac{1}{N_d}\sum_{k\in U_d} (y_k - \bar{y}_{U_d})^2\right)\left(\frac{1}{N_{d'}}\sum_{k\in U_{d'}} (y_k - \bar{y}_{U_{d'}})^2\right)
\end{aligned}
$$

which goes to 0 by assumption (A6). Hence, we have $A^{dd'} \to 0$ asymptotically. Now,

$$\begin{aligned}
B^{dd'} =& n\mathrm{E}\left|\frac{1}{N_d N_{d'}}\sum_{k\in U_d}\sum_{l\in U_{d'}}\frac{\Delta_{kl}}{\pi_k\pi_l}[(y_k-\bar{y}_{U_d})(\bar{y}_{U_{d'}}-\tilde{y}_{s_{d'}})\right.\\
&\left.+(y_l-\bar{y}_{U_{d'}})(\bar{y}_{U_d}-\tilde{y}_{s_d})+(\bar{y}_{U_d}-\tilde{y}_{s_d})(\bar{y}_{U_{d'}}-\tilde{y}_{s_{d'}})]\frac{I_kI_l}{\pi_{kl}}\right|\\
\leq&\frac{n\max_{k,l\in U,k\neq l}|\Delta_{kl}|}{\lambda_1^2\lambda_2}\frac{1}{N_dN_{d'}}\sum_{k\in U_d}\sum_{l\in U_{d'}}\mathrm{E}\left[|(y_k-\bar{y}_{U_d})(\bar{y}_{U_{d'}}-\tilde{y}_{s_{d'}})|\right.\\
&\left.+|(y_l-\bar{y}_{U_{d'}})(\bar{y}_{U_d}-\tilde{y}_{s_d})|+|(\bar{y}_{U_d}-\tilde{y}_{s_d})(\bar{y}_{U_{d'}}-\tilde{y}_{s_{d'}})|\right]\\
=&\frac{n\max_{k,l\in U,k\neq l}|\Delta_{kl}|}{\lambda_1^2\lambda_2}\left[\frac{1}{N_d}\sum_{k\in U_d}|y_k-\bar{y}_{U_d}|\mathrm{E}\left|\bar{y}_{U_{d'}}-\tilde{y}_{s_{d'}}\right|\right.\\
&\left.+\frac{1}{N_{d'}}\sum_{l\in U_{d'}}\left|y_l-\bar{y}_{U_{d'}}\right|\mathrm{E}\left|\bar{y}_{U_d}-\tilde{y}_{s_d}\right|+\mathrm{E}\left|(\bar{y}_{U_d}-\tilde{y}_{s_d})(\bar{y}_{U_{d'}}-\tilde{y}_{s_{d'}})\right|\right]
\end{aligned}$$

We have that, for $d=1,\cdots,D$, $\mathrm{E}\left|\bar{y}_{U_d}-\tilde{y}_{s_d}\right|\to 0$ asymptotically since $\mathrm{E}\left(\bar{y}_{U_d}-\tilde{y}_{s_d}\right)^2\to 0$, and by Cauchy Schwarz inequality, $\mathrm{E}\left|(\bar{y}_{U_d}-\tilde{y}_{s_d})(\bar{y}_{U_{d'}}-\tilde{y}_{s_{d'}})\right|\to 0$. Hence, we have $B^{dd'}\to 0$ as $N_d, N_{d'}\to\infty$. (B.4) is verified. Thus $n(\tilde{\mathbf{\Sigma}}_{dd'}-\mathbf{\Sigma}_{dd'})=n(\tilde{\mathbf{\Sigma}}_{dd'}-\hat{\mathbf{\Sigma}}_{dd'})+n(\hat{\mathbf{\Sigma}}_{dd'}-\mathbf{\Sigma}_{dd'})=o_p(1)O_p(1)+o_p(1)=o_p(1)$. So, by (B.3) and (B.4), we have:

$$n(\tilde{\mathbf{\Sigma}}-\mathbf{\Sigma})=o_p(1) \tag{B.5}$$

and thus:

$$I_1=n(\tilde{\mathbf{\Sigma}}-\mathbf{\Sigma})\mathcal{I}_{\emptyset}(s)=o_p(1)O_p(1)=o_p(1)$$

Now, if the observed $J\neq\emptyset$ and $J\subseteq J_{\boldsymbol{\mu}}$, we have the following result:

$$|n(\mathbf{\Sigma}-\mathbf{\Sigma}_J)|\to 0,\quad \text{as } N\to\infty \tag{B.6}$$

because the difference of the ijth element of above two matrices can be expressed as:

$$
\begin{aligned}
&|n(\{\boldsymbol{\Sigma}\}_{ij} - \{\boldsymbol{\Sigma}_J\}_{ij})| \\
=&\left|\frac{1}{N_iN_j}\sum_{k\in U_i}\sum_{l\in U_j} n\Delta_{kl}\frac{(y_k-\bar{y}_{U_i})(y_l-\bar{y}_{U_j})}{\pi_k\pi_l} - \frac{1}{N_iN_j}\sum_{k\in U_i}\sum_{l\in U_j} n\Delta_{kl}\frac{(y_k-\theta_{i,J})(y_l-\theta_{j,J})}{\pi_k\pi_l}\right| \\
=&\left|\frac{1}{N_iN_j}\sum_{k\in U_i}\sum_{l\in U_j}\frac{n\Delta_{kl}}{\pi_k\pi_l}\left[(y_k-\bar{y}_{U_i})\frac{N}{N_j}\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_j\right.\right. \\
&+\frac{N}{N_i}\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_i(y_l-\bar{y}_{U_j}) \\
&\left.\left.+\frac{N^2}{N_iN_j}\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_i\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_j\right]\right| \\
\leq&\left|\frac{N}{N_j}\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_j\frac{1}{N_iN_j}\sum_{k\in U_i}\sum_{l\in U_j}\frac{n\Delta_{kl}}{\pi_k\pi_l}(y_k-\bar{y}_{U_i})\right| \\
&+\left|\frac{N}{N_i}\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_i\frac{1}{N_iN_j}\sum_{k\in U_i}\sum_{l\in U_j}\frac{n\Delta_{kl}}{\pi_k\pi_l}(y_l-\bar{y}_{U_j})\right| \\
&+\left|\frac{N^2}{N_iN_j}\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_i\left\{\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U\right\}_j\frac{1}{N_iN_j}\sum_{k\in U_i}\sum_{l\in U_j}\frac{n\Delta_{kl}}{\pi_k\pi_l}\right| \\
=&O\left(\frac{1}{\sqrt{N}}\right)O(1)+O\left(\frac{1}{\sqrt{N}}\right)O(1)+O\left(\frac{1}{\sqrt{N}}\right)O\left(\frac{1}{\sqrt{N}}\right)O(1) \\
=&O\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

where we use the fact that:

$$
\begin{aligned}
\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J\bar{\boldsymbol{y}}_U =&\boldsymbol{A}'_J(\boldsymbol{A}_J\boldsymbol{W}_U^{-1}\boldsymbol{A}'_J)^{-1}\boldsymbol{A}_J(\boldsymbol{\mu}+O(N^{-\frac{1}{2}})) \\
=&0+O(N^{-\frac{1}{2}})=O(N^{-\frac{1}{2}})
\end{aligned}
$$

So, for the dd'th element of $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}_J$, $d, d' = 1, \cdots, D$, we have:

$$\begin{aligned} n\mathrm{E}|\hat{\boldsymbol{\Sigma}}_{dd'} - \boldsymbol{\Sigma}_{Jdd'}| =& n\mathrm{E}|\hat{\boldsymbol{\Sigma}}_{dd'} - \boldsymbol{\Sigma}_{dd'} + \boldsymbol{\Sigma}_{dd'} - \boldsymbol{\Sigma}_{Jdd'}| \\ \leq& n\mathrm{E}|\hat{\boldsymbol{\Sigma}}_{dd'} - \boldsymbol{\Sigma}_{dd'}| + n|\boldsymbol{\Sigma}_{dd'} - \boldsymbol{\Sigma}_{Jdd'}| \\ &\to 0 \end{aligned}$$

by using (B.3), (B.4) and (B.6). Hence, we have $n(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_J) = o_p(1)$ and thus

$$n(\tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_J) = n(\tilde{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}}) + n(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_J) = o_p(1)O_p(1) + o_p(1) = o_p(1)$$

Next, by Taylor expansion, we have $\boldsymbol{W}_s^{-1} \xrightarrow{p} \boldsymbol{W}_U^{-1}$ and thus $\hat{\boldsymbol{P}}_J \xrightarrow{p} \boldsymbol{P}_J$. Hence:

$$\begin{aligned} &(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)n\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)' - (\boldsymbol{I} - \boldsymbol{P}_J)n\boldsymbol{\Sigma}_J(\boldsymbol{I} - \boldsymbol{P}_J)' \\ =&(\boldsymbol{I} - \boldsymbol{P}_J)n\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \boldsymbol{P}_J)' + o_p(1) - (\boldsymbol{I} - \boldsymbol{P}_J)n\boldsymbol{\Sigma}_J(\boldsymbol{I} - \boldsymbol{P}_J)' \\ =&(\boldsymbol{I} - \boldsymbol{P}_J)n(\tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_J)(\boldsymbol{I} - \boldsymbol{P}_J)' + o_p(1) \\ =&o_p(1) + o_p(1) = o_p(1). \end{aligned}$$

and thus

$$I_2 = o_p(1)O_p(1) = o_p(1)$$

By Lemma 5, we have $P(J \neq \emptyset, J \not\subseteq J_{\boldsymbol{\mu}}) = O(n^{-1})$. So, if observed $J \neq \emptyset$ and $J \not\subseteq J_{\boldsymbol{\mu}}$, then $\mathcal{I}_J(s) = O_p(n^{-1}) = o_p(1)$. Thus:

$$I_3 = (O_p(1) + O(1))o_p(1) = o_p(1)$$

Overall, (3.5) is verified. $\square$

B.3 Proof of Theorem 9

Lemma 9. *Assume $\boldsymbol{A}\boldsymbol{\mu} > 0$ strictly, then for $J \neq \emptyset$, we have that:*

$$P(\boldsymbol{y}^{(1)} \in \mathcal{C}_J|s) = O_p\left(n^{-1}\right).$$

In other words, for the observed J set that corresponds to a simulated $\boldsymbol{y}^{(i)}$, the probability of $J \neq \emptyset$ has measure 0 asymptotically.

Proof. Suppose $J \neq \emptyset$, then for $j \in J$, we must have $\boldsymbol{A}_j\boldsymbol{y}^{(1)} < 0$ and $\boldsymbol{A}_j\boldsymbol{\mu} > 0$. Under a given sample s, using Markov's inequality, we have the following:

$$\begin{aligned}
P(\boldsymbol{y}^{(1)} \in \mathcal{C}_J|s) &\leq P(-\boldsymbol{A}_j\boldsymbol{y}^{(1)} + \boldsymbol{A}_j\boldsymbol{\mu} > \boldsymbol{A}_j\boldsymbol{\mu}|s) \\
&\leq \frac{\mathrm{E}[(\boldsymbol{A}_j\boldsymbol{y}^{(1)} - \boldsymbol{A}_j\boldsymbol{\mu})^2|s]}{(\boldsymbol{A}_j\boldsymbol{\mu})^2} \\
&= \frac{\mathrm{E}[(\boldsymbol{A}_j(\tilde{\boldsymbol{\theta}} - \boldsymbol{\mu}) + \boldsymbol{A}_j\boldsymbol{\epsilon})^2|s]}{(\boldsymbol{A}_j\boldsymbol{\mu})^2}
\end{aligned}$$

Notice that by the way we generate $\boldsymbol{y}^{(i)}$'s, we can express $\boldsymbol{y}^{(1)}$ as:

$$\boldsymbol{y}^{(1)} = \tilde{\boldsymbol{\theta}} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathrm{MVN}(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}})$$

and the numerator in above fraction can be expressed as:

$$\begin{aligned}
&\mathrm{E}[(\boldsymbol{A}_j(\tilde{\boldsymbol{\theta}} - \boldsymbol{\mu}) + \boldsymbol{A}_j\boldsymbol{\epsilon})^2|s] \\
=&\mathrm{E}[(\boldsymbol{A}_j(\tilde{\boldsymbol{\theta}} - \boldsymbol{\mu}))^2|s] + \mathrm{E}[(\boldsymbol{A}_j\boldsymbol{\epsilon})^2|s] + 2\mathrm{E}[(\boldsymbol{A}_j(\tilde{\boldsymbol{\theta}} - \boldsymbol{\mu})\boldsymbol{A}_j\boldsymbol{\epsilon})|s] \\
=&\mathrm{E}\left[\left(\sum_{d=1}^{D} a_{jd}^2(\tilde{\theta}_d - \mu_d)^2 + \sum_{d\neq d'} a_{jd}a_{jd'}(\tilde{\theta}_d - \mu_d)(\tilde{\theta}_{d'} - \mu_{d'})\right)\middle| s\right] + \mathrm{var}(\boldsymbol{A}_j\boldsymbol{\epsilon}|s) + 0 \\
=&\sum_{d=1}^{D} a_{jd}^2\mathrm{E}[(\tilde{\theta}_d - \mu_d)^2|s] + \sum_{d\neq d'} a_{jd}a_{jd'}\mathrm{E}[(\tilde{\theta}_d - \mu_d)(\tilde{\theta}_{d'} - \mu_{d'})|s] + \boldsymbol{A}_j\tilde{\boldsymbol{\Sigma}}\boldsymbol{A}_j' \\
=&O_p(n^{-1}) + O_p(n^{-1}) = O_p(n^{-1})
\end{aligned}$$

where, conditioning on the s, $(\tilde{\theta}_d - \mu_d)^2 = O_p(n^{-1})$ by Lemma 8 and $\tilde{\boldsymbol{\Sigma}}$ has order n^{-1}. Also, $\mathrm{E}[(\tilde{\theta}_d - \mu_d)(\tilde{\theta}_{d'} - \mu_{d'})|s] \leq O_p\left(n^{-1}\right)$ by the Cauchy Schwarz inequality. so we finally have:

$$P(\boldsymbol{y}^{(1)} \in \mathcal{C}_J|s) = O_p\left(n^{-1}\right).$$

□

Proof of Theorem 9

Proof. We can express $n(\hat{V}^m(\tilde{\boldsymbol{\theta}}) - AV^m(\tilde{\boldsymbol{\theta}}))$ as:

$$\begin{aligned}
&n(\hat{V}^m(\tilde{\boldsymbol{\theta}}) - AV^m(\tilde{\boldsymbol{\theta}})) \\
=&n\left[\frac{1}{B}\sum_{i=1}^{B}(\boldsymbol{I} - \hat{\boldsymbol{P}}_{J^{(i)}})\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_{J^{(i)}})' - \sum_{J}(\boldsymbol{I} - \boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I} - \boldsymbol{P}_J)'P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J)\right] \\
=&n\left[\frac{1}{B}\sum_{i=1}^{B}\left(\sum_{J}(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)'I(\boldsymbol{y}^{(i)} \in \mathcal{C}_J|s)\right)\right. \\
&\left. - \sum_{J}(\boldsymbol{I} - \boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I} - \boldsymbol{P}_J)'P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J)\right] \\
=&n\sum_{J}\left[(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)'\frac{1}{B}\sum_{i=1}^{B}I(\boldsymbol{y}^{(i)} \in \mathcal{C}_J|s) - (\boldsymbol{I} - \boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I} - \boldsymbol{P}_J)'P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J)\right] \\
=&n\left[\tilde{\boldsymbol{\Sigma}}\frac{1}{B}\sum_{i=1}^{B}I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_\emptyset|s\right) - \boldsymbol{\Sigma}P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_\emptyset)\right] \\
&+ n\sum_{J\neq\emptyset}\left[(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{P}}_J)'\frac{1}{B}\sum_{i=1}^{B}I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_J|s\right) - (\boldsymbol{I} - \boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I} - \boldsymbol{P}_J)'P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J)\right] \\
=&L_1 + L_2
\end{aligned}$$

By (B.5), we can express L_1 as:

$$
\begin{aligned}
L_1 =& n\left[\tilde{\boldsymbol{\Sigma}}\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_{\emptyset}|s\right) - \boldsymbol{\Sigma}P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_{\emptyset})\right] \\
=& n\boldsymbol{\Sigma}\left[\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_{\emptyset}|s\right) - P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_{\emptyset})\right] + o_p(1)
\end{aligned}
$$

and we have the following result:

$$
\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_{\emptyset}|s\right) - P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_{\emptyset}) = o_p(1) \tag{B.7}
$$

because

$$
\begin{aligned}
&\mathrm{E}\left|\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_{\emptyset}|s\right) - P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_{\emptyset})\right| \\
\leq& \frac{1}{B}\sum_{i=1}^{B} \mathrm{E}\left|I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_{\emptyset}|s\right) - P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_{\emptyset})\right| \\
=& \frac{1}{B}\sum_{i=1}^{B} \left[P\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_{\emptyset}|s\right)\left|1 - P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_{\emptyset})\right| + \left(1 - P\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_{\emptyset}|s\right)\right)P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_{\emptyset})\right] \to 0,
\end{aligned}
$$

where we use the fact that $1 - P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_{\emptyset}) = O(n^{-1})$ and $1 - P\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_{\emptyset}|s\right) = O(n^{-1})$ by Lemma 6 and Lemma 9. Thus, it follows that $L_1 = o_p(1)$.

Now, if the observed $J \neq \emptyset$, we have $\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_J|s\right) = o_p(1)$, since by Lemma 9:

$$
\mathrm{E}\left|\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_J|s\right)\right| \leq \frac{1}{B}\sum_{i=1}^{B} \mathrm{E}\left|I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_J|s\right)\right| = P\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_J|s\right) \to 0
$$

Also, using the same argument for showing (B.7), we have:

$$
\left[\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)} \in \mathcal{C}_J|s\right) - P(\tilde{\boldsymbol{y}}_s \in \mathcal{C}_J)\right] = o_p(1)
$$

Hence, in L_2

$$
\begin{aligned}
&n\left[(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)'\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)}\in\mathcal{C}_J|s\right)-(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'P(\tilde{\boldsymbol{y}}_s\in\mathcal{C}_J)\right]\\
=&n\left[\left((\boldsymbol{I}-\hat{\boldsymbol{P}}_J)\tilde{\boldsymbol{\Sigma}}(\boldsymbol{I}-\hat{\boldsymbol{P}}_J)'-(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\right)\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)}\in\mathcal{C}_J|s\right)\right]\\
&+n(\boldsymbol{I}-\boldsymbol{P}_J)\boldsymbol{\Sigma}_J(\boldsymbol{I}-\boldsymbol{P}_J)'\left[\frac{1}{B}\sum_{i=1}^{B} I\left(\boldsymbol{y}^{(i)}\in\mathcal{C}_J|s\right)-P(\tilde{\boldsymbol{y}}_s\in\mathcal{C}_J)\right]\\
=&(O_p(1)-O(1))o_p(1)+O(1)o_p(1)=o_p(1)
\end{aligned}
$$

It follows that $L_2=o_p(1)$ and the proof is complete. $\square$

Appendix C

Supplemental Materials for Chapter 4

C.1 Assumptions

(A1) The number of domains D is a known fixed integer and $\liminf_{N\to\infty} \frac{N_d}{N} > 0, \limsup_{N\to\infty} \frac{N_d}{N} < 1$ for $d = 1, 2, \cdots, D$.

(A2) The boundedness property of the finite population fourth moment holds. That is, we have:

$$\limsup_{N\to\infty} N^{-1} \sum_{i\in U} y_i^4 < \infty.$$

(A3) The sample size n is non-random and there exists a $\pi \in (0, 1)$ such that $\min_d \frac{n_d}{N_d} \geq \pi$ for $d = 1, \cdots, D$.

(A4) For all N, $\min_{i\in U} \pi_i \geq \lambda_1 > 0$ and $\min_{i,j\in U} \pi_{ij} \geq \lambda_2 > 0$, and

$$\limsup_{N\to\infty} n \max_{i,j\in U, i\neq j} |\Delta_{ij}| < \infty$$

where $\Delta_{ij} = \mathrm{cov}(I_i, I_j) = \pi_{ij} - \pi_i\pi_j$.

(A5) For any vector $\boldsymbol{x} \in \mathbb{R}^D$ with finite fourth population moment, we have:

$$\mathrm{var}(\hat{\boldsymbol{x}}_s)^{-\frac{1}{2}}(\hat{\boldsymbol{x}}_s - \bar{\boldsymbol{x}}_U) \xrightarrow{d} N(0, \boldsymbol{I}_{D\times D})$$

where $\hat{\boldsymbol{x}}_s$ is the HT domain mean estimator of $\bar{\boldsymbol{x}}_U = (N_1^{-1}\sum_{k\in U_1} x_k, \cdots, N_D^{-1}\sum_{k\in U_D} x_k)^\top$, $\boldsymbol{I}_{D\times D}$ is the identity matrix of dimension D, the design covariance matrix $\mathrm{var}(\hat{\boldsymbol{x}}_s)$ is positive definite.

The assumption (A1) states that the number of domains remains constant as the population size N changes and ensures that there is no asymptotically vanishing domains. Assumption (A2) is a condition needed for showing the variance consistency of the Horvitz-Thompson estimator and this condition generally can be satisfied for most survey data. In (A3) we guarantee that no matter how N and n changes, there is no vanishing sampling fraction for each domain d, which is a mild condition in the design-based context.

Assumption (A4) illustrates that the design is both a probability sampling design and a measurable design. The assumption on the Δ_{ij} states that the covariance between sample membership indicators is sufficiently small, which goes to zero at rate of n^{-1}. These conditions hold in many classical sampling designs, including simple random sampling with and without replacement, and many other unequal probability sampling designs.

The asymptotic normal assumption in (A5) is usually assumed explicitly and it is satisfied for many specific sampling designs, including simple random sampling with or without replacement. Also, it holds for Poisson sampling and unequal probability sampling with replacement. The design asymptotic normal assumption, taken together with the variance consistency of the Horvitz-Thompson estimator, can be used to derive the asymptotic distribution of the constrained domain mean estimator. More importantly, it is this normal assumption that makes it possible for us to take advantage of the available techniques in the one sided test literatures and obtain the null distribution of the test statistics approximately. Otherwise, we have to resort to the bootstrap method to get the empirical distribution of the test statistics when the properties of the design estimator are completely unknown.

C.2 Supplemental Materials for Section 4.3

Lemma 10. *Denote $\boldsymbol{\mu}$ to be the super-population domain means. Let J be the set that is associated with $\hat{\boldsymbol{\theta}}_1^*$ in (4.2) and $J_{\boldsymbol{\mu}}^0$ be the corresponding set for the solution $\boldsymbol{\theta}_{\boldsymbol{\mu}}^*$ that minimizes $(\boldsymbol{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_1)^\top(\boldsymbol{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_1)$ subject to $\boldsymbol{\theta}_1 \in \mathcal{C}_{\boldsymbol{\mu}} = \{\boldsymbol{\theta} : \boldsymbol{A}_{\boldsymbol{\mu}}\boldsymbol{\theta} \geq 0\}$, where $\boldsymbol{Z}_{\boldsymbol{\mu}} = \boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-\frac{1}{2}}\boldsymbol{\mu}$, $\mathcal{C}_{\boldsymbol{\mu}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ are super-population versions of $\hat{\boldsymbol{Z}}_s$, $\hat{\mathcal{C}}$, $\tilde{\boldsymbol{\Sigma}}$ and $\boldsymbol{A}_{\boldsymbol{\mu}} = \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{\frac{1}{2}}$. Define $J_{\boldsymbol{\mu}}^1 = \{j : \boldsymbol{A}_j\boldsymbol{\mu} = 0\}$ and let*

$J_{\boldsymbol{\mu}} = J^0_{\boldsymbol{\mu}} \cup J^1_{\boldsymbol{\mu}}$, *Then, we have:*

$$\Pr(J \nsubseteq J_{\boldsymbol{\mu}}) = o(1) \quad \textit{and} \quad \Pr(J^0_{\boldsymbol{\mu}} \nsubseteq J) = o(1)$$

Proof. Firstly, consider the event $J \nsubseteq J_{\boldsymbol{\mu}}$. Define

$$\begin{aligned}
\widetilde{SSE}(\hat{\boldsymbol{\theta}}^*_1) &= (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}^*_1)^\top (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}^*_1) \\
&= \left[\hat{\boldsymbol{Z}}_s - (\boldsymbol{I} - \hat{\boldsymbol{A}}_J^\top (\hat{\boldsymbol{A}}_J \hat{\boldsymbol{A}}_J^\top)^- \hat{\boldsymbol{A}}_J)\hat{\boldsymbol{Z}}_s\right]^\top \left[\hat{\boldsymbol{Z}}_s - (\boldsymbol{I} - \hat{\boldsymbol{A}}_J^\top (\hat{\boldsymbol{A}}_J \hat{\boldsymbol{A}}_J^\top)^- \hat{\boldsymbol{A}}_J)\hat{\boldsymbol{Z}}_s\right] \\
&= \hat{\boldsymbol{Z}}_s^\top \hat{\boldsymbol{A}}_J^\top (\hat{\boldsymbol{A}}_J \hat{\boldsymbol{A}}_J^\top)^- \hat{\boldsymbol{A}}_J \hat{\boldsymbol{Z}}_s \\
&= \tilde{\boldsymbol{y}}_s^\top \boldsymbol{A}_J^\top (\boldsymbol{A}_J \tilde{\boldsymbol{\Sigma}} \boldsymbol{A}_J^\top)^- \boldsymbol{A}_J \tilde{\boldsymbol{y}}_s
\end{aligned}$$

Similarly, we define:

$$SSE(\boldsymbol{\theta}^*_{\boldsymbol{\mu}}) = (\boldsymbol{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}^*_{\boldsymbol{\mu}})^\top (\boldsymbol{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}^*_{\boldsymbol{\mu}}) = \boldsymbol{\mu}^\top \boldsymbol{A}^\top_{J^0_{\boldsymbol{\mu}}} (\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \boldsymbol{A}^\top_{J^0_{\boldsymbol{\mu}}})^- \boldsymbol{A}_{J^0_{\boldsymbol{\mu}}} \boldsymbol{\mu}$$

Note that the projection of $\boldsymbol{Z}_{\boldsymbol{\mu}}$ onto the linear space spanned by rows of $\boldsymbol{A}_{\boldsymbol{\mu}}$ in position $J^0_{\boldsymbol{\mu}}$ is the same as the projection onto the linear space spanned by rows of $\boldsymbol{A}_{\boldsymbol{\mu}}$ in position $J_{\boldsymbol{\mu}}$, so we have:

$$SSE(\boldsymbol{\theta}^*_{\boldsymbol{\mu}}) = \boldsymbol{\mu}^\top \boldsymbol{A}^\top_{J^0_{\boldsymbol{\mu}}} (\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \boldsymbol{A}^\top_{J^0_{\boldsymbol{\mu}}})^- \boldsymbol{A}_{J^0_{\boldsymbol{\mu}}} \boldsymbol{\mu} = \boldsymbol{\mu}^\top \boldsymbol{A}^\top_{J_{\boldsymbol{\mu}}} (\boldsymbol{A}_{J_{\boldsymbol{\mu}}} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \boldsymbol{A}^\top_{J_{\boldsymbol{\mu}}})^- \boldsymbol{A}_{J_{\boldsymbol{\mu}}} \boldsymbol{\mu}$$

Further, denote:

$$\begin{aligned}
\widetilde{SSE}(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}) &= (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}})^\top (\hat{\boldsymbol{Z}}_s - \hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}) = \tilde{\boldsymbol{y}}_s^\top \boldsymbol{A}^\top_{J_{\boldsymbol{\mu}}} (\boldsymbol{A}_{J_{\boldsymbol{\mu}}} \tilde{\boldsymbol{\Sigma}} \boldsymbol{A}^\top_{J_{\boldsymbol{\mu}}})^- \boldsymbol{A}_{J_{\boldsymbol{\mu}}} \tilde{\boldsymbol{y}}_s \\
SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J}) &= (\boldsymbol{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_{\boldsymbol{\mu},J})^\top (\boldsymbol{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_{\boldsymbol{\mu},J}) = \boldsymbol{\mu}^\top \boldsymbol{A}_J^\top (\boldsymbol{A}_J \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \boldsymbol{A}_J^\top)^- \boldsymbol{A}_J \boldsymbol{\mu}
\end{aligned}$$

where $\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}} = (\boldsymbol{I} - \hat{\boldsymbol{A}}_{J_{\boldsymbol{\mu}}}^{\top}(\hat{\boldsymbol{A}}_{J_{\boldsymbol{\mu}}}\hat{\boldsymbol{A}}_{J_{\boldsymbol{\mu}}}^{\top})^{-}\hat{\boldsymbol{A}}_{J_{\boldsymbol{\mu}}})\hat{\boldsymbol{Z}}_s$ and $\boldsymbol{\theta}_{\boldsymbol{\mu},J} = (\boldsymbol{I} - \boldsymbol{A}_{\boldsymbol{\mu},J}^{\top}(\boldsymbol{A}_{\boldsymbol{\mu},J}\boldsymbol{A}_{\boldsymbol{\mu},J}^{\top})^{-}\boldsymbol{A}_{\boldsymbol{\mu},J})\boldsymbol{Z}_{\boldsymbol{\mu}}$. Then, we must have

$$SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*) < SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J}) \quad \text{and} \quad \widetilde{SSE}(\hat{\boldsymbol{\theta}}_1^*) < \widetilde{SSE}(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}})$$

and due to the consistency of $\tilde{\boldsymbol{y}}_s$ and $\tilde{\boldsymbol{\Sigma}}$, respectively, we also have:

$$\widetilde{SSE}(\hat{\boldsymbol{\theta}}_1^*) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J}) = o_p(1) \quad \text{and} \quad \widetilde{SSE}(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*) = o_p(1)$$

Finally, by Markov's inequality, we get:

$$\begin{aligned}
&\Pr(J \nsubseteq J_{\boldsymbol{\mu}}) \\
\leq &\Pr\left(\widetilde{SSE}(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}) - \widetilde{SSE}(\hat{\boldsymbol{\theta}}_1^*) + SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J}) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*) > SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J}) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*)\right) \\
\leq &\frac{E\left(\widetilde{SSE}(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}) - \widetilde{SSE}(\hat{\boldsymbol{\theta}}_1^*) + SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J}) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*)\right)}{SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J}) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*)} \\
= &\frac{E\left(\widetilde{SSE}(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*)\right) - E\left(\widetilde{SSE}(\hat{\boldsymbol{\theta}}_1^*) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J})\right)}{SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J}) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*)} \\
&\to 0
\end{aligned}$$

since $E\left(\widetilde{SSE}(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*)\right) = o(1)$ and $E\left(\widetilde{SSE}(\hat{\boldsymbol{\theta}}_1^*) - SSE(\boldsymbol{\theta}_{\boldsymbol{\mu},J})\right) = o(1)$.

Using the similar argument, we can also show that:

$$\Pr(J_{\boldsymbol{\mu}}^0 \nsubseteq J) = o(1)$$

This completes the proof. □

By the same argument as in Lemme 10, we also have the following result.

Lemma 11. *Let $J_{\boldsymbol{\Sigma}}$ (unknown) be the corresponding set of the solution $\tilde{\boldsymbol{\theta}}_1^*$ that minimizes $(\tilde{\boldsymbol{Z}}_s - \boldsymbol{\theta}_1)^{\top}(\tilde{\boldsymbol{Z}}_s - \boldsymbol{\theta}_1)$ subject to $\boldsymbol{\theta}_1 \in \tilde{\mathcal{C}}$. Then, we have:*

$$\Pr(J_{\boldsymbol{\Sigma}} \nsubseteq J_{\boldsymbol{\mu}}) = o(1) \quad \textit{and} \quad \Pr(J_{\boldsymbol{\mu}}^0 \nsubseteq J_{\boldsymbol{\Sigma}}) = o(1)$$

where $J_{\boldsymbol{\mu}}$ and $J^0_{\boldsymbol{\mu}}$ are defined in Lemma 10.

Lemma 12. *For the selected sample s, we have:*

$$\min_{\boldsymbol{\theta}_1\in\mathcal{C}}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)^\top\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)-\min_{\boldsymbol{\theta}_1\in\mathcal{C}}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)^\top\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)=o_p(1)$$

Proof. Let J be the observed set for a given sample s. We can write the difference as follows:

$$\begin{aligned}
&\min_{\boldsymbol{\theta}_1\in\mathcal{C}}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)^\top\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)-\min_{\boldsymbol{\theta}_1\in\mathcal{C}}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)^\top\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)\\
=&\hat{\boldsymbol{Z}}_s^\top\hat{\boldsymbol{A}}_J^\top(\hat{\boldsymbol{A}}_J\hat{\boldsymbol{A}}_J^\top)^-\hat{\boldsymbol{A}}_J\hat{\boldsymbol{Z}}_s-\tilde{\boldsymbol{Z}}_s^\top\tilde{\boldsymbol{A}}_{J_{\boldsymbol{\Sigma}}}^\top(\tilde{\boldsymbol{A}}_{J_{\boldsymbol{\Sigma}}}\tilde{\boldsymbol{A}}_{J_{\boldsymbol{\Sigma}}}^\top)^-\tilde{\boldsymbol{A}}_{J_{\boldsymbol{\Sigma}}}\tilde{\boldsymbol{Z}}_s\\
=&\tilde{\boldsymbol{y}}_s^\top\boldsymbol{A}_J^\top(\boldsymbol{A}_J\tilde{\boldsymbol{\Sigma}}\boldsymbol{A}_J^\top)^-\boldsymbol{A}_J\tilde{\boldsymbol{y}}_s-\tilde{\boldsymbol{y}}_s^\top\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}^\top(\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}\boldsymbol{\Sigma}\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}^\top)^-\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}\tilde{\boldsymbol{y}}_s\\
=&\tilde{\boldsymbol{y}}_s^\top\boldsymbol{A}_J^\top(\boldsymbol{A}_J\tilde{\boldsymbol{\Sigma}}\boldsymbol{A}_J^\top)^-\boldsymbol{A}_J\tilde{\boldsymbol{y}}_s\left(I_{(J^0_{\boldsymbol{\mu}}\subseteq J\subseteq J_{\boldsymbol{\mu}})}+I_{(J\not\subseteq J_{\boldsymbol{\mu}}\text{ or }J^0_{\boldsymbol{\mu}}\not\subseteq J)}\right)\\
&-\tilde{\boldsymbol{y}}_s^\top\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}^\top(\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}\boldsymbol{\Sigma}\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}^\top)^-\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}\tilde{\boldsymbol{y}}_s\left(I_{(J^0_{\boldsymbol{\mu}}\subseteq J_{\boldsymbol{\Sigma}}\subseteq J_{\boldsymbol{\mu}})}+I_{(J_{\boldsymbol{\Sigma}}\not\subseteq J_{\boldsymbol{\mu}}\text{ or }J^0_{\boldsymbol{\mu}}\not\subseteq J_{\boldsymbol{\Sigma}})}\right)
\end{aligned}$$

By Lemma 10 and Lemma 11, we have that $I_{(J\not\subseteq J_{\boldsymbol{\mu}}\text{ or }J^0_{\boldsymbol{\mu}}\not\subseteq J)}=o_p(1)$ and $I_{(J_{\boldsymbol{\Sigma}}\not\subseteq J_{\boldsymbol{\mu}}\text{ or }J^0_{\boldsymbol{\mu}}\not\subseteq J_{\boldsymbol{\Sigma}})}=o_p(1)$. Then, we have:

$$\begin{aligned}
&\min_{\boldsymbol{\theta}_1\in\mathcal{C}}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)^\top\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)-\min_{\boldsymbol{\theta}_1\in\mathcal{C}}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)^\top\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{y}}_s-\boldsymbol{\theta}_1)\\
=&\tilde{\boldsymbol{y}}_s^\top\boldsymbol{A}_J^\top(\boldsymbol{A}_J\tilde{\boldsymbol{\Sigma}}\boldsymbol{A}_J^\top)^-\boldsymbol{A}_J\tilde{\boldsymbol{y}}_sI_{(J^0_{\boldsymbol{\mu}}\subseteq J\subseteq J_{\boldsymbol{\mu}})}-\tilde{\boldsymbol{y}}_s^\top\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}^\top(\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}\boldsymbol{\Sigma}\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}^\top)^-\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}\tilde{\boldsymbol{y}}_sI_{(J^0_{\boldsymbol{\mu}}\subseteq J_{\boldsymbol{\Sigma}}\subseteq J_{\boldsymbol{\mu}})}+o_p(1)\\
=&\boldsymbol{\mu}^\top\boldsymbol{A}_J^\top(\boldsymbol{A}_J\boldsymbol{\Sigma}_{\boldsymbol{\mu}}\boldsymbol{A}_J^\top)^-\boldsymbol{A}_J\boldsymbol{\mu}I_{(J^0_{\boldsymbol{\mu}}\subseteq J\subseteq J_{\boldsymbol{\mu}})}-\boldsymbol{\mu}^\top\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}^\top(\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}^\top)^-\boldsymbol{A}_{J_{\boldsymbol{\Sigma}}}\boldsymbol{\mu}I_{(J^0_{\boldsymbol{\mu}}\subseteq J_{\boldsymbol{\Sigma}}\subseteq J_{\boldsymbol{\mu}})}+o_p(1)\\
=&\boldsymbol{\mu}^\top\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}^\top(\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}^\top)^-\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}\boldsymbol{\mu}-\boldsymbol{\mu}^\top\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}^\top(\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}^\top)^-\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}\boldsymbol{\mu}+o_p(1)\\
=&o_p(1)
\end{aligned}$$

where we use the fact that for any set J with $J^0_{\boldsymbol{\mu}} \subseteq J \subseteq J_{\boldsymbol{\mu}}$, we have that:

$$
\begin{aligned}
SSE(\boldsymbol{\theta}^*_{\boldsymbol{\mu}}) =& \boldsymbol{\mu}^\top \boldsymbol{A}_J^\top (\boldsymbol{A}_J \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \boldsymbol{A}_J^\top)^- \boldsymbol{A}_J \boldsymbol{\mu} \\
=& \boldsymbol{\mu}^\top \boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}^\top (\boldsymbol{A}_{J^0_{\boldsymbol{\mu}}} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \boldsymbol{A}_{J^0_{\boldsymbol{\mu}}}^\top)^- \boldsymbol{A}_{J^0_{\boldsymbol{\mu}}} \boldsymbol{\mu} \\
=& \boldsymbol{\mu}^\top \boldsymbol{A}_{J_{\boldsymbol{\mu}}}^\top (\boldsymbol{A}_{J_{\boldsymbol{\mu}}} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \boldsymbol{A}_{J_{\boldsymbol{\mu}}}^\top)^- \boldsymbol{A}_{J_{\boldsymbol{\mu}}} \boldsymbol{\mu}
\end{aligned}
$$

□