DISSERTATION

WEIGHTING ADJUSTMENTS IN SURVEYS

Submitted by

Ran Fu

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Jean D. Opsomer

F. Jay Breidt
Piotr Kokoszka
David Mushinski

ABSTRACT

WEIGHTING ADJUSTMENTS IN SURVEYS

We consider three topics in this dissertation: 1) Nonresponse weighting adjustment using penalized spline regression; 2) Improving survey estimators through weight smoothing; and 3) An investigation of weight smoothing estimators under mixed model specifications.

In the first topic, we propose a new survey estimator under nonresponse, which only assumes that the response propensity is a smooth function of a known covariate, and we estimate the propensity function by fitting a nonparametric logistic model using penalized spline regression. We obtain the linearization of the nonresponse weighting adjustment estimator with respect to the sampling design and the random response mechanism, allowing us to perform asymptotically correct inference. In a simulation study, we show that the nonparametric estimator remains competitive with a linear logistic estimator when the response propensity function follows a linear logistic model, but performs significantly better when the response propensity function is nonlinear.

Beaumont (2008) proposed model-based weight smoothing as a way to improve the efficiency of survey estimators. In the second topic, we extend this work by obtaining the asymptotic properties of this approach with respect to the sampling design and the weight model. The latter is taken to be a lognormal linear regression model. We derive the asymptotic distribution of the estimator and propose a consistent estimator of the asymptotic variance. A Hájek version of the estimator is considered, as well as a replication variance estimator, both of which improve the robustness of weight smoothing against model misspecification.

In the third topic, the results from the second topic are extended to models with random effects. Two versions of the estimator are proposed, depending on whether the random effects are predicted or integrated out, and their practical performance is compared through a simulation study.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. Jean D. Opsomer for the continuous support over the years of my Ph.D study, for his patience, encouragement and guidance. He has been helping me all the time of research and in this dissertation. I would like to thank Prof. F. Jay Breidt, Prof. Piotr Kokoszka and Prof. David Mushinski in my committee. They have provided, with kindness, their insight and suggestions, which are invaluable to me. I would also like to thank my fellow graduate students, the faculty, and the staff in the Statistics Department at Colorado State University.

I would like to thank my parents, Xihua Fu and Min Wang, for their endless love, support throughout the years of my education and my life.

DEDICATION

To my Motherland.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

A large amount of the quantitative information about the economy and society comes from sample surveys. The statistics may either come from censuses or be based on a sample of the population. In Statistics, survey sampling is the process of selecting a sample of elements from a population to conduct a survey. Broadly, there are two types of survey sampling: probability sampling and non-probability sampling. Probability sampling is the commonly used procedure in academic and government survey research. Probability sampling allows design-based inference about the population.

For a pure design-based inference, only the uncertainty from the randomization of the design needs to be accounted for, with all other aspects of the population treated as fixed. Consider a finite set of elements identified by the integers $U_N = \{1, \ldots, i, \ldots, N\}$. The design-based estimator of $\bar{y}_N = N^{-1} \sum_U y_i$ is given by the Horvitz-Thompson estimator $\hat{y}_{HT} = N^{-1} \sum_{i \in S} y_i/\pi_i$ or the Hájek estimator $\hat{y}_{HA} = \hat{N}^{-1} \sum_{i \in S} y_i/\pi_i$, where $\hat{N} = \sum_{i \in S} 1/\pi_i$ and π_i is the probability that the element i will be included in a sample. The Horvitz-Thompson estimator $\hat{y}_{HT}$ is unbiased for $\bar{y}_N$. The explicit form of the variance is available when the sampling design is decided. Inferential procedures, including confidence intervals and hypothesis tests are constructed for $\bar{y}_N$ under asymptotic normality. The Hájek estimator $\hat{y}_{HA}$ is an approximately unbiased estimator of the population mean $\bar{y}_N$. An approximate variance is used instead of the explicit form of the variance. Together with the Gaussian distribution, an inference that can be made from these estimates. The estimators from the pure design-based approach can be inefficient. Improving the efficiency of estimators following sampling can be achieved in a number of ways. One can reduce variances through the use of auxiliary information or by changing the weights themselves, such as nonresponse adjustments, calibration weighting, weight trimming methods and weighting adjustment through some functions. Many of the methods used to reduce the variance often results in the introduction of some bias. There is a trade-off between the bias and the variance. If the amount of the

variance reduction is larger than the amount of the squared bias increase, it leads to a smaller mean square error. The regression estimator is a type of estimator that can improve the efficiency through using of auxiliary information for the population. Let $\boldsymbol{x}_i$ represent a vector of auxiliary variables for element i and $\bar{\boldsymbol{x}}$ its population mean. The general form of the linear regression estimator is defined as

$$\hat{y}_{reg} = \hat{y}_{HT} + (\bar{\boldsymbol{x}}_N - \hat{\boldsymbol{x}}_{HT})\hat{\boldsymbol{B}},$$

where $\hat{\boldsymbol{B}}$ is the estimated coefficients that from the specific regression analysis being considered. The regression estimator is a "model-assisted approach" as it improved the efficiency by adding an adjustment term, while the regression estimator remains design consistent. The regression estimator is a special case of a Calibration estimator. A calibration estimator use the new weights w_i as close as possible to the original inverse-probability weights with respect to the calibration constraints $\sum_{i\in S} w_i \boldsymbol{x}_i = \sum_{i\in U} \boldsymbol{x}_i$ for a given distance measure.

One may consider modeling the weights as a way to obtain more effective survey estimators. When the model for the weighting adjustment is considered, the question involved in this study is whether the model is correct for the sample data or for the population level or both. More generally, when fitting models to survey data, one needs to take care how to specify models and how the model(s) and the design relate to each other. There are four basic approaches for modeling with survey data. The first approach is to assume that the model is valid under the sample itself and make inference based on that model. This is not preferable because a non-representative sample may not provide a valid inference for the population. The second approach is to assume that the model is valid under the sample and the population simultaneously. The third approach is that there are two models valid for the sample and the population separately, the two models can be connected through an additional model setup. In addition, we can also assume the fact that the two models are connected through the design, which leads to the fourth approach. The fourth approach assumes a model for the population only, without doing so for the sample. In that case, estimation and inference are done under a combined design and model-based approach. For instance, the

ordinary least square estimator for $\boldsymbol{\beta}$ for the population level is

$$\boldsymbol{\beta} = \left(\sum_{i \in U} \boldsymbol{x}_i^T \boldsymbol{x}_i \right)^{-1} \sum_{i \in U} \boldsymbol{x}_i^T y_i.$$

Combined with Horvitz-Thompson estimation, the sample-based estimator of $\boldsymbol{\beta}$ can be written as

$$\hat{\boldsymbol{\beta}}_{HT} = \left(\sum_{i \in S} \frac{\boldsymbol{x}_i^T \boldsymbol{x}_i}{\pi_i} \right)^{-1} \sum_{i \in S} \frac{\boldsymbol{x}_i^T y_i}{\pi_i}.$$

A model-based method for the finite population is conditional on the realized sample and once the model is found, the model-based inference can be established. For example, suppose that the observations in the sample and the population follow the linear model, which is given by

$$y_i = \boldsymbol{x}_i \boldsymbol{\beta} + \varepsilon_i,$$

where the ε_i's are independent and identically distributed random variables. Then the model-based predictor is given by

$$\hat{y}_{pred} = \frac{1}{N} \left(\sum_{i \in S} y_i + \sum_{S^C} \boldsymbol{x}_i \hat{\boldsymbol{B}} \right),$$

where S^C represents the complement of the sample S and $\hat{\boldsymbol{B}}$ is the estimator obtained from the ordinary least square based on the sample. Besides linear regression model, other models can be used in the study depending on the type of the relationships.

In Chapter 2, we adjust the weights under nonresponse by estimating the response propensity through nonparametric logistic model using penalized spline regression. Even though it involves modeling, this is often still considered a design-based method, because no model is assumed for the target variables of interest, i.e. the y_i. In other words, the weights are adjusted through the auxiliary variables that are available for the population and we are not trying to adjust for the study variables. The inferential procedure like the confidence interval is obtained under asymptotic normality, while the approximate variance is derived from the Taylor linearization of the estimator. In particular, we consider a nonparametric model instead of a regular regression model since the nonparametric model is more flexible. The nonparametric model has advantages when the true relationship between the response propensity and auxiliary variables is unknown to us.

In Chapter 3, the efficiency of the proposed estimators is improved by specifying a weight model through the study variables. The smoothed weights are obtained from the weight model, which lead to large reduces in variance of the estimators, while introducing a slight bias. Beaumont [2008] refers to this approach as "generalized design-based" inference. The principles remain close to the design-based inference as we are modeling the survey weights only, not the study variables. There are two main methods to obtain the variance approximation: the linearization variance estimation approach and the replicate variance estimation approach. We consider both approaches in this chapter. For the linearization variance estimation approach, the Taylor linearization of the estimator is derived with respect to the sampling design and the weight model, the variance is computed using the standard variance estimation. The replicate variance estimation is a commonly used method to estimate a variance in survey sampling, including the jackknife and bootstrap methods. The jackknife technique was first developed by Quenouille [1949, 1956] for reducing the bias of the estimation with respect to an infinite population context. Durbin [1959] considered the use of jackknife in finite-population estimation. We investigate jackknife variance estimation as a more robust and practical alternative in this chapter. The inferential procedure like confidence interval is derived under normality.

In Chapter 4, the results from Chapter 3 are extended to models with random effects. A mixed model is a statistical model containing both fixed effects and random effects. This is useful in setting when the measurements are made on clusters of units. The estimators are obtained. Combined with the jackknife variance estimation, the confidence intervals are constructed under asymptotic normality for inferential purposes.

CHAPTER 2

NONRESPONSE WEIGHTING ADJUSTMENT USING PENALIZED SPLINE REGRESSION

2.1 Introduction

Weighting by the inverse of the estimated response probabilities is a procedure that is often applied to nonresponse in surveys. Brick [2013] reviewed the consequences of nonresponse on the bias of the estimates and the methods for its adjustment. Nargundkar and Joshi [1975] and Cassel et al. [1983] provided general descriptions of propensity weighting as an adjustment for nonresponse in survey estimators. Groves et al. [2002] and Särndal and Lundström [2006] provided overviews of nonresponse weighting adjustment (NWA) techniques in survey sampling. Under nonresponse, the set of respondents can be considered to have been obtained through two-phase sampling. In the first phase, a sample of elements is selected from the population, and the second phase is the set of respondents. Following two-phase estimation ideas, the NWA approach is to multiply the sampling weight by a response weight that is the inverse of the response probability, and then to apply the usual inverse-probability-weighted estimation approach. However, the true response probability is usually not available in practice, and an estimated response probability is used to correct for nonresponse bias. Applications of the NWA estimator can be found in Ekholm and Laaksonen [1991] and Iannacchione [2003].

Auxiliary variables are often present in surveys, either at the population or the sample level. These auxiliary variables often correlate with the study variable of interest and they can be applied in adjusting the design weights to account for nonresponse. Regression weighting is a popular method for incorporating auxiliary variables. A discussion about the regression NWA estimator can be found in Cassel et al. [1983], Bethlehem [1988] and Fuller and An [1998]. Lundström and Särndal [1999] suggested that increased auxiliary information content could reduce both variance and nonresponse bias of the point estimator.

Traditionally, the NWA estimator incorporates estimated response probability into the estimator, with the response probability 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. Beaumont [2005] gave a clear justification for reduced variance using estimated response probability from a logistic regression model in the imputation context. Kim and Kim [2007] regressed the response on auxiliary variables into a linear logistic regression model for the estimated response probability. However, if this response propensity function is misspecified, the NWA estimators are likely to be biased. Another approach is to estimate the response propensities through nonparametric methods. Giommi [1984] estimated the response probabilities by kernel smoothing. Silva and Opsomer [2009] considered the estimation of the response propensities by local polynomial regression.

In this chapter, we extend Kim and Kim [2007] through nonparametric methods. The approach we consider is the NWA estimator using the estimated response probability by fitting a nonparametric logistic model that uses penalized spline regression. We show that the nonparametric estimator remains competitive with a linear logistic estimator when the response propensity function follows a linear logistic model, but performs significantly better when the response propensity function is nonlinear. In Section 2.2, we provide a literature review for the estimator under nonresponse and the penalized spline logistic regression. In Section 2.3, the properties of the NWA estimator using the estimated response probability from the penalized spline regression are discussed. In Section 2.4, the variance estimation for the NWA estimator using the estimated response probability from the penalized spline regression are provided. In Section 2.5, we perform a simulation study to evaluate the finite sample properties of the estimator. Conclusions are given in Section 2.6.

2.2 Proposed Estimator

Let the finite population $U_N = \{1, \cdots, i, \cdots, N\}$, where N is assumed known. Let $\mathcal{F}_N = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N\}$ denote the population variables. For each individual, $\boldsymbol{u}_i = (x_i, y_i)^T$, where y_i is the study variable and x_i is the auxiliary variable for unit i. For simplicity, we consider a single

study variable y_i and a single auxiliary variable x_i. The population mean for the study variable is given by $\bar{y}_U = N^{-1} \sum_U y_i$. A probability sample of size n is drawn from U_N, the inclusion of a given element i in a sample S is a random event indicated by the random variable I_i, with $I_i = 1$ if $i \in S$, and $I_i = 0$ otherwise. Let $\bar{y}_s$ be an estimator of $\bar{y}_U$, we consider here the Horvitz-Thompson estimator,

$$\bar{y}_s = N^{-1} \sum_{i \in S} \pi_i^{-1} y_i = N^{-1} \sum_{i \in U_N} \frac{y_i I_i}{\pi_i},$$

where $\pi_i = \Pr\{i \in S\} = E_p[I_i]$ and π_i^{-1} be the sampling weight of unit i. The expectation of $\bar{y}_s$, conditional on $\mathcal{F}_N$, is $\bar{y}_U$. Therefore, $\bar{y}_s$ is unbiased for $\bar{y}_U$ with respect to the sampling design.

Under nonresponse, the desired study variable y_i may not be obtained for the entire set of elements S. To define the response model, let R_i be an indicator of the response for the study variable y_i in each unit $\boldsymbol{u}_i$. Assuming $R_i = 1$ if unit i responds, and $R_i = 0$, otherwise. We assume that, given the sample, the response indicators are independent Bernoulli random variables with $p_{i|S} = \Pr\{R_i = 1 | i \in S\}$. For simplicity and as commonly done in this context, we assume that the response probability of a unit does not depend on the characteristics of the other elements in the sample nor on the realized sample. Thus, we write $p_{i|S} = p_i$. Let $\bar{y}_{d,HT}$ be the Horvitz-Thompson type estimator of $\bar{y}_s$, so that $\bar{y}_{d,HT}$ is of the form

$$\bar{y}_{d,HT} = N^{-1} \sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{p_i} y_i.$$

Conditional on the sample S, $\bar{y}_{d,HT}$ is an unbiased estimator of $\bar{y}_s$ if the true response probability is known. Therefore, $\bar{y}_{d,HT}$ is unbiased for $\bar{y}_U$ in that case.

Särndal et al. [1992, p. 182] suggested to use the Hájek estimator instead of Horvitz-Thompson estimator since the Hájek estimator is usually more efficient, despite estimation of a priori known quantity N. Therefore, we also consider a Hájek type estimator in this paper. Under the sampling design and nonresponse, the Hájek type estimator $\bar{y}_{d,HA}$ is of the form

$$\bar{y}_{d,HA} = \frac{\sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{p_i} y_i}{\sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{p_i}},$$

and $\bar{y}_{d,HA}$ is asymptotically unbiased for $\bar{y}_U$.

The response probability p_i is usually unknown in practice. Instead of using the true probability p_i, the estimated probability $\hat{p}_i$ will be used to estimate the population mean. Then, two possible estimators of the population mean $\bar{y}_U$ are given by

$$\bar{y}_{e,HT} = N^{-1} \sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i} y_i, \tag{2.1}$$

and

$$\bar{y}_{e,HA} = \frac{\sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i} y_i}{\sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i}}. \tag{2.2}$$

In order to implement the nonresponse weighting adjustment estimators in (2.1) and (2.2), it is necessary to estimate the response probability $\hat{p}_i$. We estimate the propensity function by fitting a nonparametric logistic model using penalized spline regression. We will use cubic B-splines to create a spline basis, a common choice in practice described in Chapter 5 Appendix of Hastie et al. [2009, p. 186] and Wand and Ormerod [2008].

The spline basis and the corresponding penalty matrix are constructed as follows. Consider an interval $[a, b]$, which contains all the x_i's. For an integer $K \leq n$, let $k_1, \cdots, k_{K+8}$ be a knot sequence in $[a, b]$ such that

$$a = k_1 = k_2 = k_3 = k_4 < k_5 < \cdots < k_{K+4} < k_{K+5} = k_{K+6} = k_{K+7} = k_{K+8} = b$$

and let $B_1, \cdots, B_{K+4}$ be the cubic B-spline basis functions defined by these knots. Given the vector of auxiliary variables $(x_1, \cdots, x_n)^T$, we create the $n \times (K+4)$ design matrix $\boldsymbol{B}$ with (i, k) entry $B_{ik} = B_k(x_i)$ and the $(K+4) \times (K+4)$ penalty matrix $\boldsymbol{\Omega}$ with (k, k') entry

$$\boldsymbol{\Omega}_{kk'} = \int_a^b B_k''(x) B_{k'}''(x) dx.$$

The generalized linear model with the canonical logit link can be expressed as

$$\log\left(\frac{p_i}{1 - p_i}\right) = \eta_i.$$

It is assumed that η_i depends on the auxiliary variable through the vector $\boldsymbol{B}_{x_i}$, where $\boldsymbol{B}_{x_i} \equiv [B_1(x_i), \cdots, B_{K+4}(x_i)]$. More explicitly, it is assumed that η_i is a function of $\boldsymbol{B}_{x_i}$ with parameter $\boldsymbol{\nu}$ evaluated at $\boldsymbol{\nu} = \boldsymbol{\nu}^*$,

$$\eta_i = \boldsymbol{B}_{x_i} \boldsymbol{\nu}^*,$$

where $\boldsymbol{\nu}^*$ is the vector of unknown true values. Hence

$$\log\left(\frac{p_i}{1-p_i}\right) = \eta_i = \boldsymbol{B}_{x_i}\boldsymbol{\nu}^*. \tag{2.3}$$

Note that we ignore here any model misspecification between the logit of the true response function and the spline. For a sufficiently smooth response function and K large, any remaining bias will be small and practically negligible. While this bias could be handled explicitly and its behavior as K grows investigated [see for instance Gerda Claeskens, 2009], this would significantly complicate the theoretical investigation of the proposed procedure. We construct the sample penalized pseudo-log-likelihood function [see Binder and Roberts, 2003].

$$\begin{aligned} l_\lambda(\boldsymbol{\nu}) &= \sum_{i\in S} k_i\{R_i \text{logit}(p_i) + \log(1-p_i)\} - \frac{1}{2}\lambda\boldsymbol{\nu}^T\boldsymbol{\Omega}\boldsymbol{\nu} \\ &= \sum_{i\in S} k_i\{R_i\left(\boldsymbol{B}_{x_i}\boldsymbol{\nu}\right) - \log\left(1+e^{\boldsymbol{B}_{x_i}\boldsymbol{\nu}}\right)\} - \frac{1}{2}\lambda\boldsymbol{\nu}^T\boldsymbol{\Omega}\boldsymbol{\nu}, \end{aligned} \tag{2.4}$$

where λ is a fixed smoothing parameter and k_i is the weight of unit i in the estimating equation for $\boldsymbol{\nu}$. The choice $k_i = 1$ was suggested by Beaumont [2005] to estimate $\boldsymbol{\nu}$ under the logistic regression model for the response probability, while Fuller and An [1998] suggested to use $k_i = \pi_i^{-1}$ under the two-phase sampling approach. We will therefore consider both cases.

Define $\boldsymbol{g}_i(\boldsymbol{\nu}) = \partial\{\text{logit}(p_i)\}/\partial\boldsymbol{\nu}$. Differentiating the sample penalized pseudo-log-likelihood function with respect to $\boldsymbol{\nu}$ leads to the score function

$$\boldsymbol{S}_\lambda(\boldsymbol{\nu}) = \frac{\partial l}{\partial \boldsymbol{\nu}} = \sum_{i\in S} k_i\left(R_i - p_i\right)\boldsymbol{g}_i(\boldsymbol{\nu}) - \lambda\boldsymbol{\Omega}\boldsymbol{\nu}. \tag{2.5}$$

Let $\boldsymbol{\nu}_\lambda^*$ represents the value of $\boldsymbol{\nu}$ that solves

$$E\{\boldsymbol{S}_\lambda(\boldsymbol{\nu})|\mathcal{F}_N\} = 0.$$

The "population parameter" $\boldsymbol{\nu}_\lambda^*$ can be viewed as a population level estimate of $\boldsymbol{\nu}^*$ under specific asymptotic scenario, but we do not emphasize this further here. Here, we consider $\hat{\boldsymbol{\nu}}$ as the estimator of $\boldsymbol{\nu}_\lambda^*$, and $\hat{\boldsymbol{\nu}}$ can be obtained by maximizing the sample penalized pseudo-log-likelihood function (2.4).

2.3 Asymptotic properties

In this section, we present the properties of the NWA estimators (2.1) and (2.2) under estimation of response propensities by penalized pseudo-log-likelihood maximization.

We state some assumptions before presenting the properties of the NWA estimator. We will follow the framework of Kim and Kim [2007].

A 2.1. *Assume that the sequence of finite populations of* $\boldsymbol{u_i} = (x_i, y_i)^T$ *have bounded fourth moments.*

A 2.2. *The sample moments converge to population moments, that is*

$$N^{-1}\sum_{i\in S}\pi_i^{-1}\boldsymbol{w_i}\boldsymbol{w_i}^T = N^{-1}\sum_{i=1}^{N}\boldsymbol{w_i}\boldsymbol{w_i}^T + O_p(n^{-1/2}), \tag{2.6}$$

where $\boldsymbol{w_i} = vec(\boldsymbol{u_i}\boldsymbol{u_i}^T)$ *is the column vector obtained by stacking the columns of the matrix* $\boldsymbol{u_i}\boldsymbol{u_i}^T$.

A 2.3. *As* $N \to \infty$,

$$\frac{n}{N} \to \pi \in (0,1),$$

and the sample design probabilities satisfy

$$C_1 \le n^{-1}N\pi \le C_2,$$

where C_1 *and* C_2 *are fixed positive constants.*

A 2.4. *Assume that*

$$p\lim_{n\to\infty} N^{-1}\sum_{i\in S}\pi_i^{-1}[\boldsymbol{g}_i(\boldsymbol{\nu}), \boldsymbol{g}_i(\boldsymbol{\nu})\boldsymbol{g}_i(\boldsymbol{\nu})^T, \{\partial\boldsymbol{g}_i(\boldsymbol{\nu})/\partial\boldsymbol{\nu}\}]y_i < \infty \tag{2.7}$$

uniformly in $\boldsymbol{\nu}$, *where* $\boldsymbol{g}_i(\boldsymbol{\nu}) = \partial\{logit(p_i)\}/\partial\boldsymbol{\nu}$. *That is, for the logistic parametric function* $logit(p_i) = \boldsymbol{B}_{x_i}\boldsymbol{\nu}$, *we have* $\boldsymbol{g}_i(\boldsymbol{\nu}) = \boldsymbol{B}_{x_i}$ *and* $\boldsymbol{g}_i(\boldsymbol{\nu})$ *satisties (2.7) when* $\boldsymbol{B}_{x_i}$ *has finite second moments.*

In addition to the assumptions on the sampling design and population distribution of $\boldsymbol{u}_i$, we assume the following conditions on the response mechanism.

B 2.1. *The responses R_i and R_j are independent random variables for $i \neq j$, and*

$$E(R_i|\mathcal{F}_N) = p_i,$$

$$V(R_i|\mathcal{F}_N) = p_i(1-p_i).$$

B 2.2. *Let $\boldsymbol{B}_{x_i}$ be the cubic B-spline basis function obtained from x_i. There exists a vector $\boldsymbol{\nu}^*$ such that the model (2.3) is true.*

B 2.3. *The inverse of the response probability p_i^{-1} is bounded by a fixed constant C; that is, $p_i^{-1} < C$.*

In addition to the assumptions of sampling design and response mechanism, we also need the following assumptions for the penalized spline regression:

C 2.1. *The smoothing parameter λ is a positive scalar and satisfies*

$$\lambda = O(\frac{N}{\sqrt{n}}),$$

where n is the sample size.

C 2.2. *The penalty matrix $\boldsymbol{\Omega}$ is a symmetric positive semidefinite matrix.*

C 2.3. *We assume that the matrices*

$$\sum_{i \in U_N} \pi_i k_i p_i (1-p_i) \boldsymbol{g}_i(\boldsymbol{\nu}) \boldsymbol{g}_i^T(\boldsymbol{\nu}) + \lambda \boldsymbol{\Omega}$$

and

$$\sum_{i \in S} k_i p_i (1-p_i) \boldsymbol{g}_i(\boldsymbol{\nu}) \boldsymbol{g}_i^T(\boldsymbol{\nu}) + \lambda \boldsymbol{\Omega}$$

are nonsingular for all possible samples S, thus invertible.

Theorem 2.1 below provides a linearized expression for the estimator of model parameters. This will be used subsequently to linearize the NWA estimator in Theorem 2.2.

Theorem 2.1. *Assume that (A2.1)-(A2.4), (B2.1)-(B2.3), (C2.1)-(C2.3) hold. The estimator* $\hat{\boldsymbol{\nu}}$ *satisfies*

$$\hat{\boldsymbol{\nu}} = \boldsymbol{\nu}_\lambda^* + \{\boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*)\}^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*) + o_p(n^{-1/2}), \tag{2.8}$$

where $\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*)$ *is the score function at* $\boldsymbol{\nu}_\lambda^*$ *given by*

$$\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*) = \sum_{i\in S} k_i\left(R_i - p_i\right)\boldsymbol{g}_i(\boldsymbol{\nu}_\lambda^*) - \lambda\boldsymbol{\Omega}\boldsymbol{\nu}_\lambda^*,$$

and $\boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*)$ *is the Fisher information matrix at* $\boldsymbol{\nu}_\lambda^*$ *given by*

$$\boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*) = -E\left\{\frac{\partial}{\partial\boldsymbol{\nu}^T}\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*)\middle|\mathcal{F}_N\right\} = \sum_{i\in U_N}\pi_i k_i p_i(1-p_i)\boldsymbol{g}_i(\boldsymbol{\nu}_\lambda^*)\boldsymbol{g}_i^T(\boldsymbol{\nu}_\lambda^*) + \lambda\boldsymbol{\Omega}. \tag{2.9}$$

Proof of Theorem 2.1: To prove (2.8), we apply a Taylor expansion to obtain

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

Evaluating the preceding at $\boldsymbol{\nu} = \hat{\boldsymbol{\nu}}$ yields an approximation which holds provided that $\boldsymbol{\nu}_\lambda^*$ is close to $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\nu}}$ satisfies

$$\boldsymbol{S}_\lambda(\hat{\boldsymbol{\nu}}) = 0.$$

That is,

$$\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^* = \left\{-\frac{\partial\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*)}{\partial\boldsymbol{\nu}^T}\right\}^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*) + o_p(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*), \tag{2.10}$$

and let

$$\boldsymbol{\mathcal{I}}_\lambda(\boldsymbol{\nu}_\lambda^*) = -\frac{\partial\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*)}{\partial\boldsymbol{\nu}^T} = \sum_{i\in S} k_i p_i(1-p_i)\boldsymbol{g}_i(\boldsymbol{\nu}_\lambda^*)\boldsymbol{g}_i^T(\boldsymbol{\nu}_\lambda^*) + \lambda\boldsymbol{\Omega}.$$

Then (2.10) can be written as

$$\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^* = \{\boldsymbol{\mathcal{I}}_\lambda(\boldsymbol{\nu}_\lambda^*)\}^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*) + o_p(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*). \tag{2.11}$$

Now, let $\boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*)$ be the Fisher Information matrix at $\boldsymbol{\nu}_\lambda^*$ and $\boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*)$ of the form

$$\boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*) = -E\left\{\frac{\partial}{\partial\boldsymbol{\nu}^T}\boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*)\middle|\mathcal{F}_N\right\} = \sum_{i\in U_N}\pi_i k_i p_i(1-p_i)\boldsymbol{g}_i(\boldsymbol{\nu}_\lambda^*)\boldsymbol{g}_i^T(\boldsymbol{\nu}_\lambda^*) + \lambda\boldsymbol{\Omega}.$$

Consider the normalized form of (2.11), we have

$$\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^* = \left\{ \frac{1}{N} \boldsymbol{\mathcal{I}}_\lambda(\boldsymbol{\nu}_\lambda^*) \right\}^{-1} \left\{ \frac{1}{N} \boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*) \right\} + o_p(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*). \tag{2.12}$$

Under the stated design and model assumptions, we have

$$\frac{1}{N} \boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*) = \frac{1}{N} \sum_{i \in S} k_i \left(R_i - p_i\right) \boldsymbol{g}_i(\boldsymbol{\nu}_\lambda^*) - \frac{1}{N} \lambda \boldsymbol{\Omega} \boldsymbol{\nu}_\lambda^* = O_p(n^{-1/2}), \tag{2.13}$$

and

$$\frac{1}{N} \boldsymbol{\mathcal{I}}_\lambda(\boldsymbol{\nu}_\lambda^*) = \frac{1}{N} \boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*) + O_p(n^{-1/2}). \tag{2.14}$$

Apply Taylor expansion to equation (2.14) and get

$$\left\{ \frac{1}{N} \boldsymbol{\mathcal{I}}_\lambda(\boldsymbol{\nu}_\lambda^*) \right\}^{-1} = \left\{ \frac{1}{N} \boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*) \right\}^{-1} + O_p(n^{-1/2}). \tag{2.15}$$

Inserting (2.15) and (2.13) into (2.12), we therefore obtain

$$\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^* = O_p(n^{-1/2}). \tag{2.16}$$

Inserting (2.15) and (2.16) into (2.12), we have

$$\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^* = \left\{ \frac{1}{N} \boldsymbol{I}_\lambda(\boldsymbol{\nu}_\lambda^*) \right\}^{-1} \left\{ \frac{1}{N} \boldsymbol{S}_\lambda(\boldsymbol{\nu}_\lambda^*) \right\} + o_p(n^{-1/2}).$$

□

Theorem 2.2. *Assume that (A2.1)-(A2.4), (B2.1)-(B2.3), (C2.1)-(C2.3) hold. Consider the estimation of the population mean $\bar{y}_U$ by the Horvitz-Thompson type NWA estimator $\bar{y}_{e,HT}$ defined in (2.1). Estimate the response propensity under model (2.3), where the parameter $\boldsymbol{\nu}^*$ is estimated by $\hat{\boldsymbol{\nu}}$, the maximizer the penalized pseudo-log-likelihood function (2.4). Then the Horvitz-Thompson type NWA estimator satisfies*

$$\bar{y}_{e,HT} = \bar{y}_{el,HT} + o_p(n^{-1/2}), \tag{2.17}$$

where

$$\bar{y}_{el,HT} = \frac{1}{N} \sum_{i \in S} \frac{1}{\pi_i} \left\{ k_i \pi_i p_i \boldsymbol{g}_{i0}^T \boldsymbol{\gamma}_N + \frac{R_i}{p_i} (y_i - k_i \pi_i p_i \boldsymbol{g}_{i0}^T \boldsymbol{\gamma}_N) \right\} + N^{-1} \lambda \boldsymbol{\nu}_\lambda^{*T} \boldsymbol{\Omega} \boldsymbol{\gamma}_N,$$

$\boldsymbol{g}_{i0}$ is the value of $\boldsymbol{g}_i(\boldsymbol{\nu}) = \partial\{logit(p_i)\}/\partial\boldsymbol{\nu}$ evaluated at $\boldsymbol{\nu} = \boldsymbol{\nu}_\lambda^*$ and

$$\boldsymbol{\gamma}_N = \left\{ \sum_{i \in U_N} \pi_i k_i p_i (1-p_i) \boldsymbol{g}_{i0} \boldsymbol{g}_{i0}^T + \lambda \boldsymbol{\Omega} \right\}^{-1} \sum_{i \in U_N} (1-p_i) \boldsymbol{g}_{i0} y_i.$$

Proof of Theorem 2.2: To prove (2.17), we apply a Taylor expansion to the estimated response probability to obtain

$$\hat{p}_i^{-1} - p_i^{-1} = \left(\frac{\partial p_i^{-1}}{\partial \boldsymbol{\nu}} \bigg|_{\boldsymbol{\nu}=\boldsymbol{\nu}_\lambda^*} \right)^T (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*) + 0.5 (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*)^T \left(\frac{\partial^2 p_i^{-1}}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^T} \bigg|_{\boldsymbol{\nu}=\ddot{\boldsymbol{\nu}}} \right) (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*),$$

where $\ddot{\boldsymbol{\nu}}$ is on the line segment joining $\hat{\boldsymbol{\nu}}$ and $\boldsymbol{\nu}_\lambda^*$. Considering now

$$\bar{y}_{d,HT} = N^{-1} \sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{p_i} y_i$$

and

$$\bar{y}_{e,HT} = N^{-1} \sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i} y_i,$$

we have

$$\bar{y}_{e,HT} = \bar{y}_{d,HT} + A_n^T (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*) + 0.5 (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*)^T B_n (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*), \tag{2.18}$$

where

$$A_n = N^{-1} \sum_{i \in S} \pi_i^{-1} R_i \left(\frac{\partial p_i^{-1}}{\partial \boldsymbol{\nu}} \bigg|_{\boldsymbol{\nu}=\boldsymbol{\nu}_\lambda^*} \right) y_i,$$

$$B_n = N^{-1} \sum_{i \in S} \pi_i^{-1} R_i \left(\frac{\partial^2 p_i^{-1}}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^T} \bigg|_{\boldsymbol{\nu}=\ddot{\boldsymbol{\nu}}} \right) y_i.$$

From the definition of $\boldsymbol{g}_i$, we have

$$\boldsymbol{g}_i = \frac{\partial \text{logit}(p_i)}{\partial \boldsymbol{\nu}} = \frac{1}{p_i(1-p_i)} \frac{\partial p_i}{\partial \boldsymbol{\nu}}.$$

Thus, we obtain

$$\frac{\partial p_i^{-1}}{\partial \boldsymbol{\nu}} = -(p_i^{-1} - 1) \boldsymbol{g}_i,$$

and

$$\frac{\partial^2 p_i^{-1}}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^T} = (p_i^{-1} - 1) \left(\boldsymbol{g}_i \boldsymbol{g}_i^T - \frac{\partial \boldsymbol{g}_i}{\partial \boldsymbol{\nu}} \right).$$

Hence,

$$A_n = N^{-1} \sum_{i \in S} \pi_i^{-1} R_i p_i^{-1} (p_i - 1) \boldsymbol{g}_{i0} y_i,$$

where $\boldsymbol{g}_{i0}$ is the value of $\boldsymbol{g}_i(\boldsymbol{\nu})$ evaluated at $\boldsymbol{\nu} = \boldsymbol{\nu}_\lambda^*$ and

$$B_n = N^{-1} \sum_{i \in S} \pi_i^{-1} R_i p_i^{-1} (1 - p_i) \left(\boldsymbol{g}_{i1} \boldsymbol{g}_{i1}^T - \frac{\partial \boldsymbol{g}_{i1}}{\partial \boldsymbol{\nu}} \right) y_i,$$

where $\boldsymbol{g}_{i1}$ is the value of $\boldsymbol{g}_i(\boldsymbol{\nu})$ evaluated at $\boldsymbol{\nu} = \ddot{\boldsymbol{\nu}}$. Thus,

$$A_n = N^{-1} \sum_{i \in S} \pi_i^{-1} (p_i - 1) \boldsymbol{g}_{i0} y_i + O_p(n^{-1/2}), \tag{2.19}$$

and under (2.6) and (2.7),

$$B_n = O_p(1). \tag{2.20}$$

Thus, from (2.18) and (2.20), we have

$$\bar{y}_{e,HT} = \bar{y}_{d,HT} + A_n^T (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^*) + O_p(n^{-1}). \tag{2.21}$$

For (2.8), plug in (2.5) and (2.9), we obtain

$$\begin{aligned} \hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_\lambda^* &= \left\{ \sum_{i \in U_N} \pi_i k_i p_i (1 - p_i) \boldsymbol{g}_i(\boldsymbol{\nu}_\lambda^*) \boldsymbol{g}_i^T(\boldsymbol{\nu}_\lambda^*) + \lambda \boldsymbol{\Omega} \right\}^{-1} \\ &\quad \left[\sum_{i \in S} k_i \left\{ (R_i - p_i) \boldsymbol{g}_i(\boldsymbol{\nu}_\lambda^*) \right\} - \lambda \boldsymbol{\Omega} \boldsymbol{\nu}_\lambda^* \right] + o_p(n^{-1/2}). \end{aligned} \tag{2.22}$$

Inserting (2.19) and (2.22) into (2.21), we have

$$\bar{y}_{e,HT} = \bar{y}_{el,HT}^* + o_p(n^{-1/2}), \tag{2.23}$$

where

$$\begin{aligned} \bar{y}_{el,HT}^* &= N^{-1} \sum_{i \in S} \pi_i^{-1} R_i p_i^{-1} y_i + \left\{ N^{-1} \sum_{i \in S} \pi_i^{-1} (p_i - 1) \boldsymbol{g}_{i0} y_i \right\}^T \\ &\quad \left\{ \sum_{i \in U_N} \pi_i k_i p_i (1 - p_i) \boldsymbol{g}_{i0} \boldsymbol{g}_{i0}^T + \lambda \boldsymbol{\Omega} \right\}^{-1} \left\{ \sum_{i \in S} k_i (R_i - p_i) \boldsymbol{g}_{i0} - \lambda \boldsymbol{\Omega} \boldsymbol{\nu}_\lambda^* \right\} \\ &= N^{-1} \sum_{i \in S} \pi_i^{-1} R_i p_i^{-1} y_i + N^{-1} \left\{ \sum_{i \in S} k_i (p_i - R_i) \boldsymbol{g}_{i0}^T + \lambda \boldsymbol{\nu}_\lambda^{*T} \boldsymbol{\Omega} \right\} \\ &\quad \left\{ \sum_{i \in U_N} \pi_i k_i p_i (1 - p_i) \boldsymbol{g}_{i0} \boldsymbol{g}_{i0}^T + \lambda \boldsymbol{\Omega} \right\}^{-1} \left\{ \sum_{i \in S} \pi_i^{-1} (1 - p_i) \boldsymbol{g}_{i0} y_i \right\}. \end{aligned}$$

Thus, $\bar{y}^*_{el,HT}$ can be written as

$$\bar{y}^*_{el,HT} = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_n + \frac{R_i}{p_i}(y_i - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_n)\right\} + \frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_n, \tag{2.24}$$

where

$$\boldsymbol{\gamma}_n = \left\{\sum_{i\in U_N}\pi_i k_i p_i(1-p_i)\boldsymbol{g}_{i0}\boldsymbol{g}_{i0}^T + \lambda\boldsymbol{\Omega}\right\}^{-1}\left\{\sum_{i\in S}\pi_i^{-1}(1-p_i)\boldsymbol{g}_{i0}y_i\right\}.$$

Let $\boldsymbol{\gamma}_N$ to be the variable $\boldsymbol{\gamma}$ under the population, i.e.

$$\boldsymbol{\gamma}_N = \left\{\sum_{i\in U_N}\pi_i k_i p_i(1-p_i)\boldsymbol{g}_{i0}\boldsymbol{g}_{i0}^T + \lambda\boldsymbol{\Omega}\right\}^{-1}\left\{\sum_{i\in U_N}(1-p_i)\boldsymbol{g}_{i0}y_i\right\},$$

and we have

$$\boldsymbol{\gamma}_n = \boldsymbol{\gamma}_N + O_p(n^{-1/2}). \tag{2.25}$$

For (2.24), plug-in (2.25), it follows

$$\bar{y}^*_{el,HT} = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\frac{R_i}{p_i}y_i + \left\{\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}(1-\frac{R_i}{p_i})\left(k_i\pi_i p_i\boldsymbol{g}_{i0}^T\right) + \frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\right\}\left\{\boldsymbol{\gamma}_N + O_p(n^{-1/2})\right\}.$$

From assumption (C2.1), which states $\lambda = O\left(\frac{N}{\sqrt{n}}\right)$, then $\frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega} = O(n^{-1/2})$, $\bar{y}^*_{el,HT}$ can be written as

$$\bar{y}^*_{el,HT} = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\frac{R_i}{p_i}y_i + \left\{\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}(1-\frac{R_i}{p_i})\left(k_i\pi_i p_i\boldsymbol{g}_{i0}^T\right) + \frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\right\}\boldsymbol{\gamma}_N + o_p(n^{-1/2}). \tag{2.26}$$

Combining (2.23) and (2.26), we obtain the desired result. □

Theorem 2.2 states that the equation (2.17) is the Taylor linearization of $\bar{y}_{e,HT}$ under the sampling design and response mechanism. The Horvitz-Thompson type NWA estimator $\bar{y}_{e,HT}$ is asymptotically equivalent to a random variable $\bar{y}_{el,HT}$. This is similar to the linearization obtained in Kim and Kim [2007], except that our result includes penalization, allowing for a penalized spline approach, and the fact the linearization in Kim and Kim [2007] was conditional on the sampling design. By doing the linearization with respect to the design as well, we are able to conduct the variance estimation through variance decomposition either conditionally on the response mechanism or on the sampling design given the population, as will be further discussed in the next section. Theorem 2.3 derives the same linearization as in Theorem 2.2, for the Hájek type estimator.

Theorem 2.3. *Assume that (A2.1)-(A2.4), (B2.1)-(B2.3), (C2.1)-(C2.3) hold. Consider the estimation of the population mean $\bar{y}_U$ by the Hájek type NWA estimator $\bar{y}_{e,HA}$ defined in (2.2). Estimate the response propensity under model (2.3), where the parameter $\boldsymbol{\nu}^*$ is estimated by $\hat{\boldsymbol{\nu}}$, the maximizer the penalized pseudo-log-likelihood function (2.4). Then the Hájek type NWA estimator satisfies*

$$\bar{y}_{e,HA} = \bar{y}_{el,HA} + o_p(n^{-1/2}),$$

where

$$\bar{y}_{el,HA} = \bar{y}_U + \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_C} + \frac{R_i}{p_i}(y_i - \bar{y}_U - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_C})\right\} + N^{-1}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_{N_C},$$

$\boldsymbol{g}_{i0}$ is the value of $\boldsymbol{g}_i(\boldsymbol{\nu}) = \partial\{logit(p_i)\}/\partial\boldsymbol{\nu}$ evaluated at $\boldsymbol{\nu} = \boldsymbol{\nu}_\lambda^*$ and

$$\boldsymbol{\gamma}_{N_C} = \left\{\sum_{i\in U_N}\pi_i k_i p_i(1-p_i)\boldsymbol{g}_{i0}\boldsymbol{g}_{i0}^T + \lambda\boldsymbol{\Omega}\right\}^{-1}\sum_{i\in U_N}(1-p_i)\boldsymbol{g}_{i0}(y_i - \bar{y}_U).$$

Proof of Theorem 2.3: The estimation of the population mean $\bar{y}_U$ by the Hájek type NWA estimator $\bar{y}_{e,HA}$ defined in (2.2) can be written as

$$\bar{y}_{e,HA} = \frac{\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\frac{R_i}{\hat{p}_i}y_i}{\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\frac{R_i}{\hat{p}_i}}.$$

Let $\bar{z}_e = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\frac{R_i}{\hat{p}_i}$, then

$$\bar{y}_{e,HA} = \frac{\bar{y}_{e,HT}}{\bar{z}_e}.$$

Using the result from Theorem 2.2, we have

$$\bar{y}_{e,HT} = \bar{y}_{el,HT} + o_p(n^{-1/2}),$$

where

$$\bar{y}_{el,HT} = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N + \frac{R_i}{p_i}(y_i - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N)\right\} + N^{-1}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_N,$$

$\boldsymbol{g}_{i0}$ is the value of $\boldsymbol{g}_i(\boldsymbol{\nu}) = \partial\{\text{logit}(p_i)\}/\partial\boldsymbol{\nu}$ evaluated at $\boldsymbol{\nu} = \boldsymbol{\nu}_\lambda^*$ and

$$\boldsymbol{\gamma}_N = \left\{\sum_{i\in U_N}\pi_i k_i p_i(1-p_i)\boldsymbol{g}_{i0}\boldsymbol{g}_{i0}^T + \lambda\boldsymbol{\Omega}\right\}^{-1}\sum_{i\in U_N}(1-p_i)\boldsymbol{g}_{i0}y_i.$$

Similarly, applying Theorem 2.2 to $\bar{z}_e$, we obtain

$$\bar{z}_e = \bar{z}_{el} + o_p(n^{-1/2}),$$

where

$$\bar{z}_{el} = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_{i|S}\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_1} + \frac{R_i}{p_i}(1 - k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_1})\right\} + N^{-1}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_{N_1},$$

$\boldsymbol{g}_{i0}$ is the value of $\boldsymbol{g}_i(\boldsymbol{\nu}) = \partial\{\text{logit}(p_i)\}/\partial\boldsymbol{\nu}$ evaluated at $\boldsymbol{\nu} = \boldsymbol{\nu}_\lambda^*$ and

$$\boldsymbol{\gamma}_{N_1} = \left\{\sum_{i\in U_N}\pi_i k_i p_i(1-p_i)\boldsymbol{g}_{i0}\boldsymbol{g}_{i0}^T + \lambda\boldsymbol{\Omega}\right\}^{-1}\sum_{i\in U_N}(1-p_i)\boldsymbol{g}_{i0}.$$

The approximate Hájek type estimator $\bar{y}_{e,HA}$ can be written as

$$\bar{y}_{e,HA} = \frac{\bar{y}_{el_0} + N^{-1}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_N}{\bar{z}_{el_0} + N^{-1}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_{N_1}} + o_p(n^{-1/2}), \tag{2.27}$$

where

$$\bar{y}_{el_0} = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N + \frac{R_i}{p_i}(y_i - k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N)\right\}$$

and

$$\bar{z}_{el_0} = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_1} + \frac{R_i}{p_i}(1 - k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_1})\right\}.$$

Then, apply Taylor expansion to (2.27) about $E(\bar{y}_{el_0}|\mathcal{F}_N)$ and $E(\bar{z}_{el_0}|\mathcal{F}_N)$ for the numerator and the denominator, respectively. Notice that $E(\bar{y}_{el_0}|\mathcal{F}_N) = \bar{y}_U$, $E(\bar{z}_{el_0}|\mathcal{F}_N) = 1$, and $\frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_N = O(n^{-1/2})$, $\frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_{N_1} = O(n^{-1/2})$. Then the approximate Hájek type estimator is

$$\begin{aligned}
\bar{y}_{e,HA} &= \bar{y}_{el_0} + \frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_N - \bar{y}_U(\bar{z}_{el_0} - 1) - \bar{y}_U\frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_{N_1} + o_p(n^{-1/2}) \\
&= \bar{y}_U + \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N + \frac{R_i}{p_i}(y_i - k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N)\right\} \\
&\quad -\bar{y}_U\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_1} + \frac{R_i}{p_i}(1 - k_i\pi_i p_i\boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_1})\right\} \\
&\quad +\frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_N - \bar{y}_U\frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_{N_1} + o_p(n^{-1/2}).
\end{aligned}$$

Let

$$\boldsymbol{\gamma}_{N_C} = \boldsymbol{\gamma}_N - \bar{y}_U\boldsymbol{\gamma}_{N_1} = \left\{\sum_{i\in U_N}\pi_i k_i p_i(1-p_i)\boldsymbol{g}_{i0}\boldsymbol{g}_{i0}^T + \lambda\boldsymbol{\Omega}\right\}^{-1}\sum_{i\in U_N}(1-p_i)\boldsymbol{g}_{i0}(y_i - \bar{y}_U),$$

thus, we obtain

$$\bar{y}_{e,HA} = \bar{y}_{el,HA} + o_p(n^{-1/2}),$$

where

$$\bar{y}_{el,HA} = \bar{y}_U + \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_C} + \frac{R_i}{p_i}(y_i - \bar{y}_U - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_{N_C})\right\} + \frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_{N_C}.$$

□

Theorems 2.2 and 2.3 have shown that the Horvitz-Thompson and Hájek type NWA estimators are asymptotically equivalent to estimators that are linear with respect to the design and the response mechanism. Therefore, the variances of their respective asymptotic distributions can be directly obtained from those of the linearized approximations, as is standard practice in survey asymptotic theory. In the next section, we discuss consistent estimation of these variances.

2.4 Variance estimation

Variance estimation for the NWA estimator can be conducted under a two-phase sampling [Beaumont, 2005] or under the so-called “reverse approach" [Fay, 1991, Shao and Steel, 1999]. For the two-phase sampling, we consider a random sample S is selected from population according to the design first, and then apply the response mechanism to the sample S. The form of the two-phase variance estimation can be found in Särndal et al. [1992, p. 348, Result 9.3.1]. In the reverse approach, the response variable R_i is extended to the entire finite population. First, the response mechanism is applied to the population, $R_i = 1$ if unit i responds and $R_i = 0$ if unit i do not respond. Second, a random sample S is selected from the population according to the design. By reversing the order of sampling and response, the response is explicitly treated as independent from the sample.

In this section, we discuss the variance estimation for the Horvitz-Thompson type NWA estimator $\bar{y}_{e,HT}$ and Hájek type NWA estimator $\bar{y}_{e,HA}$. As noted above, since we proved that the estimators $\bar{y}_{el,HT}$ and $\bar{y}_{el,HA}$ have the same asymptotic distributions as $\bar{y}_{e,HT}$ and $\bar{y}_{e,HA}$ respectively in Section 3, we will consider the variance estimation for the asymptotic equivalent estimators $\bar{y}_{el,HT}$

and $\bar{y}_{el,HA}$. The penalty terms $N^{-1}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_N$ and $N^{-1}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_{N_C}$ are non-random and they do not contribute to the variance of $\bar{y}_{el,HT}$ and $\bar{y}_{el,HA}$, so that we will ignore them in the following variance estimation calculation. The variance estimators will be plug-in type estimators and we will not prove their consistency. Instead, we will evaluate and compare them via simulations in the next section.

2.4.1 Variance estimation of the Horvitz-Thompson type estimator

First, we consider the reverse approach variance estimation of the Horvitz-Thompson type NWA estimator $\bar{y}_{e,HT}$. The linearized term $\bar{y}_{el,HT}$ in (2.17) can be written as

$$\bar{y}_{el,HT} = \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\delta_i + \frac{1}{N}\lambda\boldsymbol{\nu}_\lambda^{*T}\boldsymbol{\Omega}\boldsymbol{\gamma}_N,$$

where

$$\delta_i = k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N + \frac{R_i}{p_i}(y_i - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N).$$

The total variance of $\bar{y}_{el,HT}$ is

$$V(\bar{y}_{el,HT}|\mathcal{F}_N) = E\left\{V(\bar{y}_{el,HT}|\boldsymbol{R}_N,\mathcal{F}_N)|\mathcal{F}_N\right\} + V\left\{E(\bar{y}_{el,HT}|\boldsymbol{R}_N,\mathcal{F}_N)|\mathcal{F}_N\right\}, \tag{2.28}$$

where $\boldsymbol{R}_N = (R_1,\cdots,R_N)$. Given the response for the finite populations, the variance of $\bar{y}_{el,HT}$ is

$$\begin{aligned} V(\bar{y}_{el,HT}|\boldsymbol{R}_N,\mathcal{F}_N) &= V\left(N^{-1}\sum_{i\in S}\pi_i^{-1}\delta_i\Big|\boldsymbol{R}_N,\mathcal{F}_N\right) \\ &= \frac{1}{N^2}\sum_{i\in U_N}\sum_{j\in U_N}\Delta_{ij}\frac{\delta_i}{\pi_i}\frac{\delta_j}{\pi_j}, \end{aligned}$$

where $\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$, and then

$$E\left\{V(\bar{y}_{el,HT}\Big|\boldsymbol{R}_N,\mathcal{F}_N)\Big|\mathcal{F}_N\right\} = E\left\{\frac{1}{N^2}\sum_{i\in U_N}\sum_{j\in U_N}\Delta_{ij}\frac{\delta_i}{\pi_i}\frac{\delta_j}{\pi_j}\Big|\mathcal{F}_N\right\}.$$

After plugging in the estimated $\hat{p}_i = p(x_i;\boldsymbol{B}_{x_i},\hat{\boldsymbol{\nu}})$ and $\hat{\boldsymbol{g}}_i = \boldsymbol{g}(x_i;\boldsymbol{B}_{x_i},\hat{\boldsymbol{\nu}})$, an estimator of the first term in the total variance (2.28) is

$$\hat{V}_{e1,HT} = \frac{1}{N^2}\sum_{i\in S}\sum_{j\in S}\frac{\Delta_{ij}}{\pi_{ij}}\frac{\hat{\delta}_i}{\pi_i}\frac{\hat{\delta}_j}{\pi_j},$$

where

$$\hat{\delta}_i = k_i\pi_i\hat{p}_i\hat{\boldsymbol{g}}_i^T\hat{\boldsymbol{\gamma}}_n + \frac{R_i}{\hat{p}_i}(y_i - k_i\pi_i\hat{p}_i\hat{\boldsymbol{g}}_i^T\hat{\boldsymbol{\gamma}}_n), \tag{2.29}$$

$$\hat{\boldsymbol{\gamma}}_n = \left\{\sum_{i\in S_R} k_i(1-\hat{p}_i)\hat{\boldsymbol{g}}_i\hat{\boldsymbol{g}}_i^T + \lambda\boldsymbol{\Omega}\right\}^{-1}\left\{\sum_{i\in S_R}\pi_i^{-1}(\hat{p}_i^{-1}-1)\hat{\boldsymbol{g}}_i y_i\right\},$$

and $S_R = \{i \in S; R_i = 1\}$ is the set of respondents in the sample.

Particularly, under the stratified simple random sampling without replacement (STSI), which is used in the simulation below, the first term $\hat{V}_{e1,HT}$ can be written as

$$\hat{V}_{e1,HT,STSI} = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2(1-\frac{n_h}{N_h})\frac{1}{n_h}\frac{1}{n_h-1}\sum_{S_h}\left(\hat{\delta}_i - \bar{\hat{\delta}}_{S_h}\right)^2,$$

where $\hat{\delta}_i$ is given in (2.29), N_h is the number of elements in stratum h, n_h is the size of the sampled elements in stratum h and $\bar{\hat{\delta}}_{S_h}$ is the averaged $\hat{\delta}_i$ within each strata h.

To estimate the second term of the total variance in (2.28), note that

$$\begin{aligned} V\left\{E(\bar{y}_{el,HT}\big|\boldsymbol{R}_N,\mathcal{F}_N)\big|\mathcal{F}_N\right\} &= V\left\{E\left(N^{-1}\sum_{i\in S}\pi^{-1}\delta_i\Big|\boldsymbol{R}_N,\mathcal{F}_N\right)\Big|\mathcal{F}_N\right\} \\ &= V\left(\frac{1}{N}\sum_{i\in U_N}\delta_i\Big|\mathcal{F}_N\right) \\ &= \frac{1}{N^2}\sum_{i\in U_N}\frac{p_i(1-p_i)}{p_i^2}(y_i - k_i\pi_i p_i\boldsymbol{g}_i^T\boldsymbol{\gamma}_N)^2. \end{aligned}$$

So an estimator for the second term of the total variance is

$$\hat{V}_{e2,HT} = \frac{1}{N^2}\sum_{i\in S_R}\pi_i^{-1}\hat{p}_i^{-2}(1-\hat{p}_i)(y_i - k_i\pi_i\hat{p}_i\hat{\boldsymbol{g}}_{i0}^T\hat{\boldsymbol{\gamma}}_n)^2.$$

Then, the total variance estimator is given by

$$\hat{V}_{e,HT} = \hat{V}_{e1,HT} + \hat{V}_{e2,HT}, \tag{2.30}$$

where $\hat{V}_{e1,HT}$ and $\hat{V}_{e2,HT}$ are given above.

Second, consider the variance estimation by using the two-phase sampling. Under the finite populations, the variance of $\bar{y}_{el,HT}$ can be written as

$$V(\bar{y}_{el,HT}\big|\mathcal{F}_N) = V\left\{E(\bar{y}_{el,HT}\big|\boldsymbol{I}_N,\mathcal{F}_N)\big|\mathcal{F}_N\right\} + E\left\{V(\bar{y}_{el,HT}\big|\boldsymbol{I}_N,\mathcal{F}_N)\big|\mathcal{F}_N\right\}. \tag{2.31}$$

Given the sample, the expected value of $\bar{y}_{el,HT}$ is

$$\begin{aligned} & E(\bar{y}_{el,HT}\big|\boldsymbol{I}_N,\mathcal{F}_N) \\ &= E\left[\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N + \frac{R_i}{p_i}(y_i - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N)\right\}\Bigg|\boldsymbol{I}_N,\mathcal{F}_N\right] \\ &= \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}y_i. \end{aligned}$$

So the first term in (2.31) is given by

$$\begin{aligned} V\left(E\left(\bar{y}_{el,HT}\big|\boldsymbol{I}_N,\mathcal{F}_N\right)\big|\mathcal{F}_N\right) &= V\left(\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}y_i\Big|\mathcal{F}_N\right) \\ &= \frac{1}{N^2}\sum_{i\in U_N}\sum_{j\in U_N}\Delta_{ij}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j} \\ &= \frac{1}{N^2}\sum_{i\in U_N}\Delta_{ii}\left(\frac{y_i}{\pi_i}\right)^2 + \frac{1}{N^2}\sum_{i,j\in U_N}\sum_{i\neq j}\Delta_{ij}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}. \end{aligned}$$

Since p_i is unknown, we use the estimate $\hat{p}_i = p(x_i;\boldsymbol{B}_{x_i},\hat{\boldsymbol{\nu}})$ to obtain a plug-in estimator

$$\hat{V}_{sam,HT} = \frac{1}{N^2}\sum_{i\in S_R}\hat{p}_i^{-1}\frac{\Delta_{ii}}{\pi_i}\left(\frac{y_i}{\pi_i}\right)^2 + \frac{1}{N^2}\sum_{i,j\in S_R}\sum_{i\neq j}\hat{p}_i^{-1}\hat{p}_j^{-1}\frac{\Delta_{ij}}{\pi_{ij}}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}.$$

Particularly, under the stratified simple random sampling without replacement (STSI), the first term $V_{sam,HT}$ is

$$V_{sam,HT,STSI} = \frac{1}{N^2}\sum_{h=1}^{H}N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{1}{n_h}\left\{\frac{1}{N_h-1}\sum_{U_h}(y_i-\bar{y}_{U_h})^2\right\}.$$

An estimator for $V_{sam,HT,STSI}$ can be written as

$$\hat{V}_{sam,HT,STSI} = \frac{1}{N^2}\sum_{h=1}^{H}N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{1}{n_h}\left\{\frac{1}{\sum_{S_{hR}}\frac{1}{\hat{p}_i}-1}\sum_{S_{hR}}\frac{1}{\hat{p}_i}\left(y_i - \frac{\sum_{S_{hR}}\frac{y_i}{\hat{p}_i}}{\sum_{S_{hR}}\frac{1}{\hat{p}_i}}\right)^2\right\},$$

where N_h is the number of elements in stratum h, n_h is the size of the sampled elements in stratum h. Given the sample, the variance of $\bar{y}_{el,HT}$ in the second term of (2.31) is

$$\begin{aligned} & V(\bar{y}_{el,HT}\big|\boldsymbol{I}_N, \mathcal{F}_N) \\ &= V\left[\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N + \frac{R_i}{p_i}\left(y_i - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N\right)\right\}\Big|\boldsymbol{I}_N, \mathcal{F}_N\right] \\ &= \frac{1}{N^2}\sum_{i\in S}\frac{1}{\pi_i^2}\frac{1-p_i}{p_i}\left(y_i - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N\right)^2, \end{aligned}$$

so that

$$\begin{aligned} V_{res,HT} &= E\left\{V(\bar{y}_{el,HT}\big|I_N, \mathcal{F}_N)\big|\mathcal{F}_N\right\} \\ &= E\left\{\frac{1}{N^2}\sum_{i\in S}\frac{1}{\pi_i^2}\frac{1-p_i}{p_i}\left(y_i - k_i\pi_i p_i \boldsymbol{g}_{i0}^T\boldsymbol{\gamma}_N\right)^2\Big|\mathcal{F}_N\right\}. \end{aligned}$$

After plugging-in the estimates of $\hat{p}_i = p(x_i; \boldsymbol{B}_{x_i}, \hat{\boldsymbol{\nu}})$ and $\hat{\boldsymbol{g}}_i = \boldsymbol{g}(x_i; \boldsymbol{B}_{x_i}, \hat{\boldsymbol{\nu}})$, we obtain the estimator for $V_{res,HT}$ is

$$\hat{V}_{res,HT} = \frac{1}{N^2}\sum_{i\in S_R}\frac{1}{\pi_i^2}\frac{1-\hat{p}_i}{\hat{p}_i^2}\left(y_i - k_i\pi_i\hat{p}_i\hat{\boldsymbol{g}}_i^T\hat{\boldsymbol{\gamma}}_n\right)^2.$$

Then, the total two phase variance estimator is given by

$$\hat{V}_{HT} = \hat{V}_{sam,HT} + \hat{V}_{res,HT}, \tag{2.32}$$

where $\hat{V}_{sam,HT}$ and $\hat{V}_{res,HT}$ are given above.

2.4.2 Variance estimation of the Hájek type estimator

Similarly, we obtain the variance estimator of the Hájek type estimator by the reverse approach and the two-phase sampling methods. We consider two types of Hájek variance estimators on the basis of whether the true population total N or the estimated population total $\hat{N}$ is used, where $\hat{N} = \sum_{i\in S}\frac{1}{\pi_i}\frac{R_i}{\hat{p}_i}$.

First, consider the reverse approach. We consider the asymptotic equivalent linearization term $\bar{y}_{el,HA}$ of the Hájek type NWA estimator $\bar{y}_{e,HA}$. The total variance of $\bar{y}_{el,HA}$ is

$$V(\bar{y}_{el,HA}|\mathcal{F}_N) = E\left\{V(\bar{y}_{el,HA}|\boldsymbol{R}_N, \mathcal{F}_N)|\mathcal{F}_N\right\} + V\left\{E(\bar{y}_{el,HA}|\boldsymbol{R}_N, \mathcal{F}_N)|\mathcal{F}_N\right\}, \tag{2.33}$$

where $\boldsymbol{R}_N = (R_1, \cdots, R_N)$. Given the response for the finite populations, the variance of $\bar{y}_{el,HA}$ is

$$
\begin{aligned}
& V(\bar{y}_{el,HA}|\boldsymbol{R}_N, \mathcal{F}_N) \\
&= V\left[\bar{y}_U + \frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\left\{k_i\pi_i p_i \boldsymbol{g}_i^T \boldsymbol{\gamma}_{N_C} + \frac{R_i}{p_i}(y_i - \bar{y}_U - k_i\pi_i p_i \boldsymbol{g}_i^T \boldsymbol{\gamma}_{N_C})\right\} \middle| \boldsymbol{R}_N, \mathcal{F}_N\right].
\end{aligned}
$$

Let

$$
\delta_{ic} = k_i\pi_i p_i \boldsymbol{g}_i^T \boldsymbol{\gamma}_{N_C} + \frac{R_i}{p_i}(y_i - \bar{y}_U - k_i\pi_i p_i \boldsymbol{g}_i^T \boldsymbol{\gamma}_{N_C}),
$$

where

$$
\boldsymbol{\gamma}_{N_C} = \boldsymbol{\gamma}_N - \bar{y}_U \cdot \boldsymbol{\gamma}_{N_1} = \left\{\sum_{i\in U_N} \pi_i k_i p_i (1-p_i)\boldsymbol{g}_i \boldsymbol{g}_i^T + \lambda\boldsymbol{\Omega}\right\}^{-1} \sum_{i\in U_N}(1-p_i)\boldsymbol{g}_i(y_i - \bar{y}_U).
$$

Then, it follows

$$
V(\bar{y}_{el,HA}|\boldsymbol{R}_N, \mathcal{F}_N) = V\left(\frac{1}{N}\sum_{i\in S}\frac{1}{\pi_i}\delta_{ic}\middle|\boldsymbol{R}_N, \mathcal{F}_N\right) = \frac{1}{N^2}\sum_{i\in U_N}\sum_{j\in U_N}\Delta_{ij}\frac{\delta_{ic}}{\pi_i}\frac{\delta_{jc}}{\pi_j},
$$

so the first variance component of (2.33) is

$$
E\left\{V(\bar{y}_{el,HA}|\boldsymbol{R}_N, \mathcal{F}_N)|\mathcal{F}_N\right\} = E\left(\frac{1}{N^2}\sum_{i\in U_N}\sum_{j\in U_N}\Delta_{ij}\frac{\delta_{ic}}{\pi_i}\frac{\delta_{jc}}{\pi_j}\middle|\mathcal{F}_N\right).
$$

If the true population totals N is known, then the plug-in Hájek type I variance estimator of the first term in (2.33) is

$$
\hat{V}_{e1,HA(I)} = \frac{1}{N^2}\sum_{i\in S}\sum_{j\in S}\frac{\Delta_{ij}}{\pi_{ij}}\frac{\hat{\delta}_{ic}}{\pi_i}\frac{\hat{\delta}_{jc}}{\pi_j},
$$

where

$$
\hat{\delta}_{ic} = k_i\pi_i\hat{p}_i\hat{\boldsymbol{g}}_i^T\hat{\boldsymbol{\gamma}}_{n_c} + \frac{R_i}{\hat{p}_i}(y_i - \bar{y}_{e,HA} - k_i\pi_i\hat{p}_i\hat{\boldsymbol{g}}_i^T\hat{\boldsymbol{\gamma}}_{n_c}), \tag{2.34}
$$

$$
\hat{\boldsymbol{\gamma}}_{n_c} = \left\{\sum_{i\in S_R} k_i(1-\hat{p}_i)\hat{\boldsymbol{g}}_i\hat{\boldsymbol{g}}_i^T + \lambda\boldsymbol{\Omega}\right\}^{-1}\sum_{i\in S_R}\pi_i^{-1}(\hat{p}_i^{-1}-1)\hat{\boldsymbol{g}}_i(y_i - \bar{y}_{e,HA}).
$$

Then, if we use a population estimator $\hat{N}$ instead of a priori known quantity N, the plug-in Hájek type II variance estimator of the first term in (2.33) is

$$
\hat{V}_{e1,HA(II)} = \frac{1}{\hat{N}^2}\sum_{i\in S}\sum_{j\in S}\frac{\Delta_{ij}}{\pi_{ij}}\frac{\hat{\delta}_{ic}}{\pi_i}\frac{\hat{\delta}_{jc}}{\pi_j},
$$

where

$$\hat{N} = \sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i},$$

$$\hat{\delta}_{ic} = k_i \pi_i \hat{p}_i \hat{\boldsymbol{g}}_i^T \hat{\boldsymbol{\gamma}}_{n_c} + \frac{R_i}{\hat{p}_i}(y_i - \bar{y}_{e,HA} - k_i \pi_i \hat{p}_i \hat{\boldsymbol{g}}_i^T \hat{\boldsymbol{\gamma}}_{n_c}),$$

$$\hat{\boldsymbol{\gamma}}_{n_c} = \left\{ \sum_{i \in S_R} k_i (1 - \hat{p}_i) \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^T + \lambda \boldsymbol{\Omega} \right\}^{-1} \sum_{i \in S_R} \pi_i^{-1} (\hat{p}_i^{-1} - 1) \hat{\boldsymbol{g}}_i (y_i - \bar{y}_{e,HA}).$$

Particularly, under STSI, $\hat{V}_{e1,HA(I)}$ and $\hat{V}_{e1,HA(II)}$ are

$$\begin{aligned}
\hat{V}_{e1,HA(I),STSI} &= \frac{1}{N^2} \sum_{h=1}^{H} N_h^2 (1 - \frac{n_h}{N_h}) \frac{1}{n_h} \frac{1}{n_h - 1} \sum_{S_h} \left(\hat{\delta}_{ic} - \bar{\hat{\delta}}_{c_{S_h}} \right)^2, \\
\hat{V}_{e1,HA(II),STSI} &= \frac{1}{\hat{N}^2} \sum_{h=1}^{H} N_h^2 (1 - \frac{n_h}{N_h}) \frac{1}{n_h} \frac{1}{n_h - 1} \sum_{S_h} \left(\hat{\delta}_{ic} - \bar{\hat{\delta}}_{c_{S_h}} \right)^2,
\end{aligned}$$

where $\hat{\delta}_{ic}$ is given in (2.34), N_h is the number of elements in stratum h, n_h is the size of the sampled elements in stratum h and $\bar{\hat{\delta}}_{c_{S_h}}$ is the averaged $\hat{\delta}_{ic}$ within each strata h.

Notice that the expected value of $\bar{y}_{el,HA}$ given $\boldsymbol{R}_N$ in the second term of (2.33),

$$E(\bar{y}_{el,HA} | \boldsymbol{R}_N, \mathcal{F}_N) = \bar{y}_U + \frac{1}{N} \sum_{i \in U_N} \delta_{ic},$$

and then

$$\begin{aligned}
V\{E(\bar{y}_{el,HA} | \boldsymbol{R}_N, \mathcal{F}_N) | \mathcal{F}_N\} &= V\left(\frac{1}{N} \sum_{i \in U_N} \delta_{ic} \middle| \mathcal{F}_N \right) \\
&= \frac{1}{N^2} \sum_{i \in U_N} \frac{1 - p_i}{p_i} (y_i - \bar{y}_U - k_i \pi_i p_i \boldsymbol{g}_{i0}^T \boldsymbol{\gamma}_{N_C})^2,
\end{aligned}$$

so an estimator for the second variance component of Hájek type I is

$$\hat{V}_{e2,HA(I)} = \frac{1}{N^2} \sum_{i \in S_R} \pi_i^{-1} \hat{p}_i^{-2} (1 - \hat{p}_i)(y_i - \bar{y}_{e,HA} - k_i \pi_i \hat{p}_i \hat{\boldsymbol{g}}_i^T \hat{\boldsymbol{\gamma}}_{n_c})^2,$$

where

$$\hat{\boldsymbol{\gamma}}_{n_c} = \left\{ \sum_{i \in S_R} k_i (1 - \hat{p}_i) \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^T + \lambda \boldsymbol{\Omega} \right\}^{-1} \sum_{i \in S_R} \pi_i^{-1} (\hat{p}_i^{-1} - 1) \hat{\boldsymbol{g}}_i (y_i - \bar{y}_{e,HA}).$$

Similarly, using $\hat{N}$ instead of N, the second variance components of Hájek type II estimator is

$$\hat{V}_{e2,HA(II)} = \frac{1}{\hat{N}^2} \sum_{i \in S_R} \pi_i^{-1} \hat{p}_i^{-2} (1 - \hat{p}_i)(y_i - \bar{y}_{e,HA} - k_i \pi_i \hat{p}_i \hat{\boldsymbol{g}}_i^T \hat{\boldsymbol{\gamma}}_{n_c})^2,$$

where

$$\hat{N} = \sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i},$$

$$\hat{\boldsymbol{\gamma}}_{n_c} = \left\{ \sum_{i \in S_R} k_i (1 - \hat{p}_i) \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^T + \lambda \boldsymbol{\Omega} \right\}^{-1} \sum_{i \in S_R} \pi_i^{-1} (\hat{p}_i^{-1} - 1) \hat{\boldsymbol{g}}_i (y_i - \bar{y}_{e,HA}).$$

Then, the total variance estimator of the Hájek type I and Hájek type II are given by

$$\hat{V}_{e,HA(I)} = \hat{V}_{e1,HA(I)} + \hat{V}_{e2,HA(I)}, \tag{2.35}$$

$$\hat{V}_{e,HA(II)} = \hat{V}_{e1,HA(II)} + \hat{V}_{e2,HA(II)}, \tag{2.36}$$

where $\hat{V}_{e1,HA(I)}$, $\hat{V}_{e2,HA(I)}$, $\hat{V}_{e1,HA(II)}$, $\hat{V}_{e2,HA(II)}$ are given above.

Second, consider the two-phase sampling variance estimation of the Hájek type estimator $\bar{y}_{e,HA}$. Under the finite population, the variance of $\bar{y}_{el}$ can be written

$$V(\bar{y}_{el,HA} \big| \mathcal{F}_N) = V\left\{ E(\bar{y}_{el,HA} \big| \boldsymbol{I}_N, \mathcal{F}_N) \big| \mathcal{F}_N \right\} + E\left\{ V(\bar{y}_{el,HA} \big| \boldsymbol{I}_N, \mathcal{F}_N) \big| \mathcal{F}_N \right\}. \tag{2.37}$$

The first term of two-phase variance is

$$V\left\{ E\left(\bar{y}_{el,HA} \big| \boldsymbol{I}_N, \mathcal{F}_N \right) \big| \mathcal{F}_N \right\} \;=\; \frac{1}{N^2} \sum_{i \in U_N} \sum_{j \in U_N} \Delta_{ij} \frac{y_i - \bar{y}_U}{\pi_i} \frac{y_j - \bar{y}_U}{\pi_j}.$$

Then the plug-in estimator of Hájek type I for the first term is

$$\begin{aligned} &\hat{V}_{sam,HA(I)} \\ &= \frac{1}{N^2} \sum_{i \in S_R} \hat{p}_i^{-1} \frac{\Delta_{ii}}{\pi_i} \left(\frac{y_i - \bar{y}_{el,HA}}{\pi_i} \right)^2 + \frac{1}{N^2} \sum_{i,j \in S_R} \sum_{i \neq j} \hat{p}_i^{-1} \hat{p}_j^{-1} \frac{\Delta_{ij}}{\pi_{ij}} \left(\frac{y_i - \bar{y}_{e,HA}}{\pi_i} \right) \left(\frac{y_j - \bar{y}_{e,HA}}{\pi_j} \right). \end{aligned}$$

If we use $\hat{N}$ instead of N in $\hat{V}_{sam,HA(I)}$, the Hájek type II variance estimator of the first component in (2.37) is

$$\begin{aligned} &\hat{V}_{sam,HA(II)} \\ &= \frac{1}{\hat{N}^2} \sum_{i \in S_R} \hat{p}_i^{-1} \frac{\Delta_{ii}}{\pi_i} \left(\frac{y_i - \bar{y}_{el,HA}}{\pi_i} \right)^2 + \frac{1}{\hat{N}^2} \sum_{i,j \in S_R} \sum_{i \neq j} \hat{p}_i^{-1} \hat{p}_j^{-1} \frac{\Delta_{ij}}{\pi_{ij}} \left(\frac{y_i - \bar{y}_{e,HA}}{\pi_i} \right) \left(\frac{y_j - \bar{y}_{e,HA}}{\pi_j} \right), \end{aligned}$$

where

$$\hat{N} = \sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i}.$$

Under STSI in the simulation study for the following Chapter, $\hat{V}_{sam,HA(I)}$ and $\hat{V}_{sam,HA(II)}$ can be written as

$$\begin{aligned}
&\hat{V}_{sam,HA(I),STSI} \\
&= \frac{1}{N^2} \sum_{h=1}^{H} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} \left\{ \frac{1}{\sum_{S_{hR}} \frac{1}{\hat{p}_i} - 1} \sum_{S_{hR}} \frac{1}{\hat{p}_i} \left(y_i - \bar{y}_{e,HA} - \frac{\sum_{S_{hR}} \frac{y_i - \bar{y}_{e,HA}}{\hat{p}_i}}{\sum_{S_{hR}} \frac{1}{\hat{p}_i}} \right)^2 \right\}, \\
&\hat{V}_{sam,HA(II),STSI} \\
&= \frac{1}{\hat{N}^2} \sum_{h=1}^{H} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} \left\{ \frac{1}{\sum_{S_{hR}} \frac{1}{\hat{p}_i} - 1} \sum_{S_{hR}} \frac{1}{\hat{p}_i} \left(y_i - \bar{y}_{e,HA} - \frac{\sum_{S_{hR}} \frac{y_i - \bar{y}_{e,HA}}{\hat{p}_i}}{\sum_{S_{hR}} \frac{1}{\hat{p}_i}} \right)^2 \right\}.
\end{aligned}$$

Consider the variance of $\bar{y}_{el,HA}$ given the sample in the second term of (2.37),

$$\begin{aligned}
&V(\bar{y}_{el,HA} | \boldsymbol{I}_N, \mathcal{F}_N) \\
&= V \left[\bar{y}_U + \frac{1}{N} \sum_{i \in S} \frac{1}{\pi_i} \left\{ k_i \pi_i p_i \boldsymbol{g}_{i0}^T \boldsymbol{\gamma}_{N_C} + \frac{R_i}{p_i} (y_i - \bar{y}_U - k_i \pi_i p_i \boldsymbol{g}_{i0}^T \boldsymbol{\gamma}_{N_C}) \right\} \middle| \boldsymbol{I}_N, \mathcal{F}_N \right] \\
&= \frac{1}{N^2} \sum_{i \in S} \frac{1}{\pi_i^2} \frac{1 - p_i}{p_i} (y_i - \bar{y}_U - k_i \pi_i p_i \boldsymbol{g}_{i0}^T \boldsymbol{\gamma}_{N_C})^2,
\end{aligned}$$

so that

$$\begin{aligned}
V_{res,HA} &= E \left\{ V(\bar{y}_{el,HA} | \boldsymbol{I}_N, \mathcal{F}_N) | \mathcal{F}_N \right\} \\
&= E \left\{ \frac{1}{N^2} \sum_{i \in S} \frac{1}{\pi_i^2} \frac{1 - p_i}{p_i} (y_i - \bar{y}_U - k_i \pi_i p_i \boldsymbol{g}_{i0}^T \boldsymbol{\gamma}_{N_C})^2 \middle| \mathcal{F}_N \right\}.
\end{aligned}$$

Then, the estimator of V_{res} of type I and type II are

$$\begin{aligned}
\hat{V}_{res,HA(I)} &= \tfrac{1}{N^2} \textstyle\sum_{i \in S_R} \tfrac{1}{\pi_i^2} \tfrac{1 - \hat{p}_i}{\hat{p}_i^2} (y_i - \bar{y}_{e,HA} - k_i \pi_i \hat{p}_i \hat{\boldsymbol{g}}_i^T \hat{\boldsymbol{\gamma}}_{n_c})^2, \\
\hat{V}_{res,HA(II)} &= \tfrac{1}{\hat{N}^2} \textstyle\sum_{i \in S_R} \tfrac{1}{\pi_i^2} \tfrac{1 - \hat{p}_i}{\hat{p}_i^2} (y_i - \bar{y}_{e,HA} - k_i \pi_i \hat{p}_i \hat{\boldsymbol{g}}_i^T \hat{\boldsymbol{\gamma}}_{n_c})^2,
\end{aligned}$$

where

$$\hat{\boldsymbol{\gamma}}_{n_c} = \left\{ \sum_{i \in S_R} k_i (1 - \hat{p}_i) \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^T + \lambda \boldsymbol{\Omega} \right\}^{-1} \sum_{i \in S_R} \pi_i^{-1} (\hat{p}_i^{-1} - 1) \hat{\boldsymbol{g}}_i (y_i - \bar{y}_{e,HA}),$$

and

$$\hat{N} = \sum_{i \in S} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i}.$$

Then, the total two phase variance estimator of Hájek type I and Hájek type II are given by

$$\hat{V}_{HA(I)} = \hat{V}_{sam,HA(I)} + \hat{V}_{res,HA(I)}, \tag{2.38}$$

$$\hat{V}_{HA(II)} = \hat{V}_{sam,HA(II)} + \hat{V}_{res,HA(II)}, \tag{2.39}$$

where $\hat{V}_{sam,HA(I)}$, $\hat{V}_{res,HA(I)}$, $\hat{V}_{sam,HA(II)}$, $\hat{V}_{res,HA(II)}$ are given above.

In the next section, we compare these two variance estimators through a simulation experiment.

2.5 Simulation study

The simulation study follows the structure of Kim and Kim [2007]. Suppose the finite populations are from a multivariate normal distribution

$$\begin{pmatrix} y_{hi} \\ x_{hi} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N\left[\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right], \; h = 1, 2, 3, 4; \; i = 1, \cdots, N_h,$$

where y_{hi} is the study variable and x_{hi} is the auxiliary variable available for both responses and nonresponses, and $N_h = 1000, 2000, 3000, 4000$ for $h = 1, 2, 3, 4$, respectively. Three different finite populations are generated from different levels of ρ, that is, $\rho = (0, 0.3, 0.6)$.

For each population, two sets of independent stratified random samples of size $n = 100$ and $n = 400$ are generated without replacement and the sample sizes are all equal ($n_h = n/4$, for $h = 1, 2, 3, 4$) in each stratum. The response indicator variable R_{hi} is generated from the Bernoulli distribution with probability p_{hi}, where p_{hi} is considered in two logistic response functions of the auxiliary variable x_{hi}. These two functions consider a linear and a nonlinear predictor of x_{hi}, which are similar to Silva and Opsomer [2009] as follows

Linear predictor: $p_{hi} = \{1 + \exp[-(x_{hi} - 1)]\}^{-1}$

Nonlinear predictor: $p_{hi} = \left\{ 1 + \exp\left[-\left(-1 + \frac{x_{hi}}{3} + \frac{x_{hi}^3}{5} + \cos\left(-\frac{x_{hi}^3}{\pi} \right) \sin\left(-\frac{2x_{hi}^2}{\pi} \right) \right) \right] \right\}^{-1}$.

The study variable y_{hi} is observed if and only if the response indicator variable $R_{hi} = 1$, and the auxiliary variable x_{hi} is observed for the sample. The finite populations of (y_{hi}, x_{hi}) are fixed in the Monte Carlo sampling. The average response rates for both linear predictor and nonlinear predictor are about 70% in the simulation. The Monte Carlo sample sizes are all $B = 10,000$.

For each of three different populations and two different sets of stratified random samples of size $n = 100$ and $n = 400$, we compare the means and variances of the nine NWA estimators 1-9 for both linear and nonlinear predictors in the following:

1. $\bar{y}_d$: NWA estimator using the true response probability p_{hi},

2. $\bar{y}_{e_p}(1)$: NWA estimator using the logistic regression estimated response probability with $k_i = 1$,

3. $\bar{y}_{e_p}(\pi^{-1})$: NWA estimator using the logistic regression estimated response probability with $k_i = \pi_i^{-1}$,

4. $\bar{y}_{e_{np1}}(1)$: NWA estimator using the penalized spline logistic regression estimated response probability with $k_i = 1$ and the smoothing parameter $\lambda = 1$,

5. $\bar{y}_{e_{np1}}(\pi^{-1})$: NWA estimator using the penalized spline logistic regression estimated response probability with $k_i = \pi^{-1}$ and the smoothing parameter $\lambda = 1$,

6. $\bar{y}_{e_{np10}}(1)$: NWA estimator using the penalized spline logistic regression estimated response probability with $k_i = 1$ and the smoothing parameter $\lambda = 10$,

7. $\bar{y}_{e_{np10}}(\pi^{-1})$: NWA estimator using the penalized spline logistic regression estimated response probability with $k_i = \pi^{-1}$ and the smoothing parameter $\lambda = 10$,

8. $\bar{y}_{e_{np200}}(1)$: NWA estimator using the penalized spline logistic regression estimated response probability with $k_i = 1$ and the smoothing parameter $\lambda = 200$,

9. $\bar{y}_{e_{np200}}(\pi^{-1})$: NWA estimator using the penalized spline logistic regression estimated response probability with $k_i = \pi^{-1}$ and the smoothing parameter $\lambda = 200$.

The simulation results are given in Table 2.1 through Table 2.16. Notice that for the nonlinear predictor function, when the sample size is small, say $n = 100$, if the model is misspecified, the estimated response probability will be close to 0. In that case, the inverse of the estimated response probability may become very large, and the means and the variances of the NWA estimators will be extremely large because of these. Since the Horvitz-Thompson type estimators are not stable in some replicates in the simulation, we correct the results by deleting such particular extreme replicates from $B = 10,000$ Monte Carlo sample sizes. These corrections are made only for the Horvitz-Thompson type estimators of the nonlinear predictor function when the sample size $n = 100$, as shown in Table 2.2, Table 2.6 and Table 2.12.

Table 2.1 though Table 2.4 give the Monte Carlo percentage relative biases and variances of the Horvitz-Thompson type estimators and Hájek type estimators for both linear predictor and nonlinear predictor, respectively. The Monte Carlo percentage relative bias and scaled variances are computed by

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

and

$$\text{Scaled Variance} = \frac{Var(\star)}{Var(\bar{y}_d)},$$

where the notation $\cdot$ stands for each of the NWA estimators 1-9 and the notation $\star$ stands for each of the NWA estimators 2-9. For the Horvitz-Thompson estimators of the linear predictor in Table 2.1, the relative biases of the estimators 2 and 3 from linear logistic regression are all small with absolute values less than 0.2%. The absolute values of the relative biases of the estimators 4-9 from the penalized spline logistic regression are all less than 3%. As the smoothing parameter increases from 1 to 200, the relative biases gets smaller. The scaled variances of the estimators 2-9 are all less than 1, which means the variance of the estimator using the estimated response probability is smaller than the variance of the estimator using the true response probability, regardless of the correlation coefficient between y_{hi} and x_{hi}.

For the Horvitz-Thompson estimators of the nonlinear predictor in Table 2.2, we recognize that the relative biases of the estimators 2 and 3 from unweighted and weighted linear logistic

Table 2.1: Monte Carlo relative biases and variances of the NWA estimators (Horvitz-Thompson type), the variances are scaled with respect to $\bar{y}_d$, based on 10,000 samples. (Linear case)

n	Estimator	Relative Bias (%)			Variance (Scaled)		
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	-0.03	0.13	0.14	–	–	–
	$\bar{y}_{e_p}(1)$	-0.01	0.07	0.11	0.58	0.57	0.60
	$\bar{y}_{e_p}(\pi^{-1})$	-0.07	0.03	0.09	0.48	0.48	0.52
	$\bar{y}_{e_{np1}}(1)$	-0.91	-0.44	-0.20	0.50	0.55	0.58
	$\bar{y}_{e_{np1}}(\pi^{-1})$	-2.45	-1.76	-1.18	0.40	0.44	0.49
	$\bar{y}_{e_{np10}}(1)$	-0.34	-0.07	0.03	0.54	0.56	0.59
	$\bar{y}_{e_{np10}}(\pi^{-1})$	-1.63	-1.07	-0.64	0.40	0.44	0.49
	$\bar{y}_{e_{np200}}(1)$	-0.11	0.07	0.11	0.57	0.57	0.60
	$\bar{y}_{e_{np200}}(\pi^{-1})$	-0.79	-0.43	-0.19	0.41	0.45	0.50
400	$\bar{y}_d$	0.02	0.00	0.10	–	–	–
	$\bar{y}_{e_p}(1)$	-0.04	0.00	0.06	0.52	0.55	0.57
	$\bar{y}_{e_p}(\pi^{-1})$	-0.05	-0.02	0.07	0.43	0.45	0.48
	$\bar{y}_{e_{np1}}(1)$	-0.39	-0.20	-0.04	0.49	0.54	0.57
	$\bar{y}_{e_{np1}}(\pi^{-1})$	-0.70	-0.46	-0.18	0.39	0.44	0.47
	$\bar{y}_{e_{np10}}(1)$	-0.21	-0.10	0.01	0.50	0.54	0.57
	$\bar{y}_{e_{np10}}(\pi^{-1})$	-0.51	-0.31	-0.07	0.39	0.44	0.47
	$\bar{y}_{e_{np200}}(1)$	-0.06	-0.01	0.05	0.52	0.55	0.57
	$\bar{y}_{e_{np200}}(\pi^{-1})$	-0.27	-0.15	0.01	0.40	0.44	0.47

Table 2.2: Monte Carlo relative biases and variances of the NWA estimators (Horvitz-Thompson type), the variances are scaled with respect to $\bar{y}_d$, based on 10,000 samples. (Nonlinear case)

n	Estimator	Relative Bias (%)			Variance (Scaled)		
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	-0.26	0.07	-0.21	–	–	–
	$\bar{y}_{e_p}(1)$	19.46	13.96	6.58	18.25	13.50	8.48
	$\bar{y}_{e_p}(\pi^{-1})$	21.88	14.51	6.78	18.18	15.24	12.51
	$\bar{y}_{e_{np1}}(1)$	4.04	3.16	1.28	1.57	1.62	1.17
	$\bar{y}_{e_{np1}}(\pi^{-1})$	-3.23	-2.48	-1.82	0.37	0.42	0.47
	$\bar{y}_{e_{np10}}(1)$	15.51	11.24	5.39	10.96	8.90	4.97
	$\bar{y}_{e_{np10}}(\pi^{-1})$	-1.46	-1.17	-1.01	0.41	0.50	0.51
	$\bar{y}_{e_{np200}}(1)$	19.23	13.80	6.51	17.75	13.18	8.21
	$\bar{y}_{e_{np200}}(\pi^{-1})$	9.47	6.29	3.06	4.19	4.21	2.62
400	$\bar{y}_d$	-0.01	-0.02	0.00	–	–	–
	$\bar{y}_{e_p}(1)$	20.14	14.48	5.78	12.67	8.56	6.63
	$\bar{y}_{e_p}(\pi^{-1})$	21.01	14.08	5.78	15.12	8.37	7.99
	$\bar{y}_{e_{np1}}(1)$	0.47	0.31	-0.24	0.53	0.59	0.58
	$\bar{y}_{e_{np1}}(\pi^{-1})$	-0.87	-0.72	-0.48	0.35	0.39	0.44
	$\bar{y}_{e_{np10}}(1)$	10.14	7.32	2.95	2.75	2.60	1.64
	$\bar{y}_{e_{np10}}(\pi^{-1})$	-0.27	-0.39	-0.43	0.36	0.41	0.46
	$\bar{y}_{e_{np200}}(1)$	19.33	13.90	5.57	11.40	7.84	5.90
	$\bar{y}_{e_{np200}}(\pi^{-1})$	9.34	6.09	2.51	2.24	1.86	1.33

regression are quite large for both sample sizes $n = 100$ and $n = 400$. As the correlation coefficient increase from $\rho = 0$ to $\rho = 0.6$, the relative biases get smaller. Examining the relative biases of the estimators 4-9 from penalized spline logistic regression, we see the penalized spline logistic regression estimators are all smaller than the linear logistic regression estimators, regardless the choice of the weight. The relative biases from the penalized spline logistic regression estimators become closer to the relative biases of the linear logistic regression estimators by increasing the smoothing parameter from $\lambda = 1$ to $\lambda = 200$. The scaled variances of the estimators 4-9 from penalized spline logistic regression are all smaller than the scaled variances of the estimators 2 and 3 from the logistic regression. It is seen that the estimators which have small absolute relative bias (less than 2%) also have small scaled variance (less than 1).

The Monte Carlo percentage relative biases and variances of the Hájek type NWA estimators for linear predictor and nonlinear predictor are also provided here for comparison, as shown in Table 2.3 and Table 2.4, respectively. The percentage relative biases are all small with absolute

values less than 0.2% when $\rho = 0$ for both linear and nonlinear predictors. As ρ gets larger, the penalized spline logistic regression estimators 4-9 give nearly unbiased estimates compared with the linear logistic estimators 2 and 3 when the response propensity function follows a linear logistic relationship, as shown in Table 2.3. For the nonlinear predictor, where the form of the response propensity function do not follow a linear logistic relationship, as shown in Table 2.4, the linear logistic estimators 2 and 3 yield large biases for the nonlinear model, but the penalized spline logistic regression estimators 4-9 present smaller biases. For example, when $\rho = 0.6$ with the sample size $n = 400$, the absolute percentage relative bias from the linear logistic regression is around 11%, but the nonparametric estimator with $w_i = \pi^{-1}$ and the smoothing parameter $\lambda = 10$ presents a much smaller absolute percentage relative bias, which is around 0.1%. Again, it is seen that the relative biases from the penalized spline logistic regression estimators become closer to the relative biases of the linear logistic regression estimators by increasing the smoothing parameter from $\lambda = 1$ to $\lambda = 200$. By using Hájek type estimators instead of Horvitz-Thompson type estimators, we correct the percentage relative bias to be more stable and closer to 0% when $\rho = 0$ for both linear predictor and nonlinear predictor.

The variance results of the Hájek type estimators in Table 2.3 and Table 2.4 present the scaled variances of the linear logistic regression estimators and the penalized spline logistic regression estimators (i.e. estimators 2 to 9) with respect to the variance of $\bar{y}_d$. It shows the penalized spline logistic regression estimators (i.e. estimators 4 to 9) are more efficient than the linear logistic regression estimators (i.e. estimators 2 and 3), regardless the choice of the weight. Given the same smoothing parameter λ, the penalized spline logistic regression estimators using $k_i = 1$ tend to be less efficient than the estimators using $k_i = \pi_i^{-1}$.

Table 2.5 through Table 2.10 present the percent relative biases of the variance estimators for NWA estimators 1-9 from both linear predictor and nonlinear predictor using the Horvitz-Thompson type, Hájek type I and Hájek type II estimators, respectively. The relative bias of the estimated variance is the Monte Carlo bias divided by the Monte Carlo variance of the point estimator. We compute two variance estimators for each of the estimators 2-9. One is the variance

Table 2.3: Monte Carlo relative biases and variances of the NWA estimators (Hájek type), the variances are scaled with respect to $\bar{y}_d$, based on 10,000 samples. (Linear case)

n	Estimator	Relative Bias (%)			Variance (Scaled)		
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	-0.06	0.15	0.34	–	–	–
	$\bar{y}_{e_p}(1)$	-0.06	0.16	0.20	1.03	0.99	0.90
	$\bar{y}_{e_p}(\pi^{-1})$	-0.08	0.13	0.19	1.03	0.99	0.86
	$\bar{y}_{e_{np1}}(1)$	-0.06	0.38	0.63	0.99	0.95	0.83
	$\bar{y}_{e_{np1}}(\pi^{-1})$	-0.02	0.63	1.28	0.98	0.94	0.79
	$\bar{y}_{e_{np10}}(1)$	-0.06	0.22	0.32	1.01	0.98	0.88
	$\bar{y}_{e_{np10}}(\pi^{-1})$	-0.04	0.48	0.95	0.98	0.94	0.79
	$\bar{y}_{e_{np200}}(1)$	-0.06	0.16	0.21	1.03	0.99	0.90
	$\bar{y}_{e_{np200}}(\pi^{-1})$	-0.07	0.31	0.54	0.99	0.95	0.80
400	$\bar{y}_d$	-0.04	0.00	0.11	–	–	–
	$\bar{y}_{e_p}(1)$	-0.04	0.00	0.10	1.01	0.97	0.89
	$\bar{y}_{e_p}(\pi^{-1})$	-0.04	-0.01	0.09	1.01	0.97	0.86
	$\bar{y}_{e_{np1}}(1)$	-0.04	0.13	0.33	1.00	0.95	0.81
	$\bar{y}_{e_{np1}}(\pi^{-1})$	-0.04	0.20	0.49	0.99	0.94	0.77
	$\bar{y}_{e_{np10}}(1)$	-0.04	0.06	0.20	1.00	0.95	0.83
	$\bar{y}_{e_{np10}}(\pi^{-1})$	-0.04	0.15	0.39	0.99	0.94	0.77
	$\bar{y}_{e_{np200}}(1)$	-0.04	0.00	0.11	1.01	0.97	0.88
	$\bar{y}_{e_{np200}}(\pi^{-1})$	-0.04	0.07	0.23	1.00	0.94	0.80

Table 2.4: Monte Carlo relative biases and variances of the NWA estimators (Hájek type), the variances are scaled with respect to $\bar{y}_d$, based on 10,000 samples. (Nonlinear case)

n	Estimator	Relative Bias (%)			Variance (Scaled)		
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	-0.18	0.26	0.34	–	–	–
	$\bar{y}_{e_p}(1)$	-0.02	-4.23	-8.66	2.71	3.27	3.77
	$\bar{y}_{e_p}(\pi^{-1})$	0.11	-4.57	-8.90	2.77	3.57	3.71
	$\bar{y}_{e_{np1}}(1)$	-0.09	-1.26	-2.68	1.46	1.56	1.40
	$\bar{y}_{e_{np1}}(\pi^{-1})$	-0.11	0.72	1.43	1.02	0.97	0.78
	$\bar{y}_{e_{np10}}(1)$	-0.02	-3.60	-7.41	2.39	2.84	3.12
	$\bar{y}_{e_{np10}}(\pi^{-1})$	-0.11	0.19	0.41	1.06	1.04	0.81
	$\bar{y}_{e_{np200}}(1)$	-0.02	-4.20	-8.59	2.69	3.25	3.73
	$\bar{y}_{e_{np200}}(\pi^{-1})$	0.05	-2.51	-4.67	1.81	2.28	1.86
400	$\bar{y}_d$	-0.02	0.01	0.10	–	–	–
	$\bar{y}_{e_p}(1)$	-0.14	-6.64	-11.74	4.25	6.58	8.00
	$\bar{y}_{e_p}(\pi^{-1})$	-0.07	-6.63	-11.59	4.24	6.51	7.56
	$\bar{y}_{e_{np1}}(1)$	0.03	-0.46	-0.83	1.07	1.04	0.82
	$\bar{y}_{e_{np1}}(\pi^{-1})$	0.04	0.20	0.45	0.99	0.93	0.71
	$\bar{y}_{e_{np10}}(1)$	-0.02	-3.80	-6.90	2.25	2.95	3.19
	$\bar{y}_{e_{np10}}(\pi^{-1})$	0.04	-0.10	-0.12	1.01	0.95	0.71
	$\bar{y}_{e_{np200}}(1)$	-0.13	-6.44	-11.40	4.07	6.26	7.58
	$\bar{y}_{e_{np200}}(\pi^{-1})$	0.07	-3.37	-6.01	2.00	2.51	2.42

estimator using the reverse approach, as defined in (2.30) for Horvitz-Thompson type estimator, in (2.35) for Hájek type I estimator and in (2.36) for Hájek type II estimator. The other one is the two-phase variance estimator, as defined in (2.32) for Horvitz-Thompson type, in (2.38) for Hájek type I estimator and in (2.39) for Hájek type II estimator. We also provide the two-phase variance estimators of $\bar{y}_d$ for comparison. The reverse approach and two-phase variance estimators show similar results of percent relative biases. We see the absolute percentage relative biases of the variance estimators using the estimated response probability tend to be smaller as n increases from $n = 100$ to $n = 400$. For linear predictor in Table 2.5, Table 2.7 and Table 2.9, the Horvitz-Thompson type variance estimators show similar performances as the Hájek type variance estimators, and there is a negative relative bias implying that the two variance estimators may underestimate the true variance. For nonlinear predictor in Table 2.6, Table 2.8 and Table 2.10, when n is as large as 400, the Hájek type I variance estimators from the linear logistic regression estimators (i.e. estimators 2 and 3) show positive relative bias values, while the Horvitz-Thompson type estimators and Hájek type II variance estimators from the linear logistic NWA estimators show negative relative bias values. The changes in the sign can be explained by the fact that the linear logistic regression is misspecified for a nonlinear predictor. For n is as large as 400, the absolute relative biases of the Hájek type variance estimators from the penalized spline logistic regression are smaller than the variance estimators from linear logistic regression in general. The relative biases of the variance estimators from the penalized spline logistic regression become closer to the variance estimators from the linear logistic regression by increasing the smoothing parameter from $\lambda = 1$ to $\lambda = 200$.

Table 2.11 through Table 2.16 present the mean lengths and coverages of 95% confidence intervals of linear predictor and nonlinear predictor for each of the Horvitz-Thompson type estimator, Hájek type I estimator and Hájek type II estimator, respectively. The mean length is $2 \times 1.96\sqrt{\hat{V}}$ and the confidence intervals are $\left(\hat{\theta} - 1.96\sqrt{\hat{V}}, \hat{\theta} + 1.96\sqrt{\hat{V}}\right)$, where $\hat{\theta}$ is a point estimator and $\hat{V}$ is its estimated variance. The same as the variance estimators, there are no significant differences between the interval estimators computed from the reverse method and that computed from the

Table 2.5: Percent relative biases of the variance estimators (Horvitz-Thompson type), based on 10,000 samples. (Linear Case)

n	Parameter	Method	Relative Bias (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	Variance of $\bar{y}_d$	Two-phase	2.20	-0.04	-0.12
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-6.25	-2.33	-2.03
		Two-phase	-8.46	-5.22	-5.38
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-8.99	-4.05	-5.57
		Two-phase	-10.08	-6.16	-8.19
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-5.36	-4.25	-2.74
		Two-phase	-8.78	-8.12	-7.06
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-13.36	-10.90	-8.36
		Two-phase	-15.89	-14.75	-13.67
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-5.45	-2.62	-2.19
		Two-phase	-8.25	-5.82	-5.78
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-9.75	-7.73	-6.78
		Two-phase	-11.69	-10.75	-10.95
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-6.22	-2.70	-2.04
		Two-phase	-8.48	-5.61	-5.41
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-6.07	-4.44	-5.61
		Two-phase	-7.83	-6.97	-8.83
400	Variance of $\bar{y}_d$	Two-phase	-2.80	-3.90	-2.01
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-3.73	-3.31	-1.85
		Two-phase	-4.38	-4.02	-2.67
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-4.99	-3.58	-0.76
		Two-phase	-5.39	-4.10	-1.45
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-6.29	-4.94	-3.18
		Two-phase	-7.39	-6.13	-4.48
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-9.23	-7.90	-4.07
		Two-phase	-9.76	-8.88	-5.52
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-4.97	-3.96	-2.43
		Two-phase	-5.84	-4.87	-3.43
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-7.19	-5.87	-2.93
		Two-phase	-7.60	-6.66	-4.10
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-3.88	-3.34	-1.89
		Two-phase	-4.56	-4.08	-2.72
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-5.81	-4.36	-1.59
		Two-phase	-6.31	-5.04	-2.47

Table 2.6: Percent relative biases of the variance estimators (Horvitz-Thompson type), based on 10,000 samples. (Nonlinear Case)

n	Parameter	Method	Relative Bias (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	Variance of $\bar{y}_d$	Two-phase	0.65	-0.87	-1.05
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-68.32	-65.61	-65.51
		Two-phase	-64.46	-62.05	-62.60
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-52.58	-53.08	-59.35
		Two-phase	-46.12	-46.03	-54.41
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-52.89	-53.36	-41.02
		Two-phase	-53.33	-53.84	-42.35
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-25.64	-23.80	-16.24
		Two-phase	-27.83	-27.71	-22.40
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-69.00	-66.69	-63.39
		Two-phase	-66.17	-63.90	-61.33
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-24.38	-27.44	-18.21
		Two-phase	-26.10	-30.03	-22.09
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-68.43	-65.70	-65.46
		Two-phase	-64.63	-62.19	-62.61
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-50.80	-55.42	-55.27
		Two-phase	-48.24	-51.57	-52.84
400	Variance of $\bar{y}_d$	Two-phase	0.35	-1.37	-1.96
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-28.51	-23.93	-28.66
		Two-phase	-25.33	-20.73	-25.30
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-37.62	-31.59	-36.82
		Two-phase	-33.83	-28.05	-33.43
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-10.51	-10.58	-6.23
		Two-phase	-11.34	-11.47	-6.60
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-9.19	-7.38	-4.74
		Two-phase	-9.51	-8.28	-6.18
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-32.65	-30.36	-28.44
		Two-phase	-31.42	-28.93	-26.31
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-8.34	-7.20	-5.40
		Two-phase	-8.48	-7.70	-5.88
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-29.40	-24.75	-29.63
		Two-phase	-26.38	-21.71	-26.47
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-41.39	-36.86	-33.21
		Two-phase	-40.27	-35.66	-31.14

Table 2.7: Percent relative biases of the variance estimators (Hájek type I), based on 10,000 samples. (Linear Case)

n	Parameter	Method	Relative Bias (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	Variance of $\bar{y}_d$	Two-phase	-2.15	-1.56	-3.74
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-4.39	-4.65	-7.91
		Two-phase	-5.65	-6.16	-9.66
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-4.91	-5.22	-7.86
		Two-phase	-5.95	-6.53	-9.16
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-6.58	-7.25	-10.40
		Two-phase	-7.33	-8.21	-11.53
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-14.92	-15.49	-16.95
		Two-phase	-14.26	-15.17	-16.57
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-4.99	-5.34	-8.69
		Two-phase	-6.13	-6.71	-10.35
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-10.76	-11.13	-12.83
		Two-phase	-10.74	-11.40	-13.10
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-4.43	-4.69	-7.97
		Two-phase	-5.69	-6.19	-9.72
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-6.74	-7.26	-9.37
		Two-phase	-7.36	-8.13	-10.29
400	Variance of $\bar{y}_d$	Two-phase	-2.33	-0.95	1.33
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-2.86	-2.15	-1.79
		Two-phase	-3.20	-2.56	-2.28
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-3.17	-2.42	-2.86
		Two-phase	-3.46	-2.75	-3.22
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-5.61	-5.43	-4.55
		Two-phase	-5.72	-5.60	-4.70
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-8.54	-8.63	-7.21
		Two-phase	-8.36	-8.57	-7.07
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-4.17	-3.67	-3.10
		Two-phase	-4.40	-3.97	-3.45
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-6.66	-6.37	-5.33
		Two-phase	-6.61	-6.41	-5.30
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-2.99	-2.34	-2.08
		Two-phase	-3.32	-2.74	-2.56
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-4.69	-4.06	-3.76
		Two-phase	-4.83	-4.25	-3.96

Table 2.8: Percent relative biases of the variance estimators (Hájek type I), based on 10,000 samples. (Nonlinear Case)

n	Parameter	Method	Relative Bias (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	Variance of $\bar{y}_d$	Two-phase	-3.13	-4.55	-4.73
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-18.52	-25.83	-28.20
		Two-phase	-20.23	-27.00	-28.44
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-15.12	-10.53	-22.89
		Two-phase	-16.17	-9.27	-21.46
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-25.35	-30.10	-30.98
		Two-phase	-27.33	-32.16	-33.03
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-27.49	-28.41	-25.76
		Two-phase	-26.91	-28.11	-25.47
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-22.85	-29.80	-33.07
		Two-phase	-24.95	-31.50	-34.19
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-22.64	-25.08	-20.92
		Two-phase	-22.87	-25.53	-21.47
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-18.82	-26.12	-28.57
		Two-phase	-20.56	-27.32	-28.88
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-22.24	-27.78	-28.53
		Two-phase	-24.02	-28.20	-29.63
400	Variance of $\bar{y}_d$	Two-phase	-0.99	-0.55	-1.24
	Variance of $\bar{y}_{e_p}(1)$	Reverse	22.30	20.35	15.04
		Two-phase	20.90	21.01	17.05
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	16.72	17.84	11.30
		Two-phase	15.55	19.22	14.21
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-4.90	-5.05	-3.57
		Two-phase	-5.44	-5.64	-3.31
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-9.13	-8.45	-7.50
		Two-phase	-8.98	-8.40	-7.26
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	1.94	-4.98	-9.59
		Two-phase	-0.30	-6.39	-9.78
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-6.81	-6.51	-4.35
		Two-phase	-6.92	-6.61	-3.87
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	20.17	17.42	12.33
		Two-phase	18.66	17.86	14.12
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-2.56	-6.48	-13.56
		Two-phase	-4.68	-7.56	-13.27

Table 2.9: Percent relative biases of the variance estimators (Hájek type II), based on 10,000 samples. (Linear Case)

n	Parameter	Method	Relative Bias (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	Variance of $\bar{y}_d$	Two-phase	-1.53	-1.71	-4.51
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-4.91	-5.26	-9.40
		Two-phase	-5.86	-6.46	-10.86
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-5.56	-5.60	-8.72
		Two-phase	-6.55	-6.86	-10.06
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-5.10	-5.91	-9.28
		Two-phase	-5.57	-6.60	10.11
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-10.59	-11.34	-12.81
		Two-phase	-9.84	-10.94	-12.33
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-4.79	-5.26	-9.17
		Two-phase	-5.63	-6.35	-10.51
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-7.83	-8.33	-10.05
		Two-phase	-7.76	-8.56	-10.28
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-4.91	-5.27	-9.39
		Two-phase	-5.85	-6.46	-10.84
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-5.41	-5.98	-8.16
		Two-phase	-6.00	-6.82	-9.05
400	Variance of $\bar{y}_d$	Two-phase	-2.28	-1.10	0.78
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-3.00	-2.50	-2.98
		Two-phase	-3.26	-2.83	-3.38
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-3.26	-2.64	-3.68
		Two-phase	-3.54	-2.96	-4.03
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-4.98	-4.86	-3.96
		Two-phase	-5.03	-4.96	-4.04
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-7.30	-7.42	-5.98
		Two-phase	-7.11	-7.36	-5.82
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-3.90	-3.47	-2.95
		Two-phase	-4.06	-3.70	-3.22
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-5.78	-5.51	-4.47
		Two-phase	-5.72	-5.54	-4.43
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-3.09	-2.61	-3.03
		Two-phase	-3.35	-2.93	-3.43
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-4.28	-3.67	-3.42
		Two-phase	-4.41	-3.86	-3.62

Table 2.10: Percent relative biases of the variance estimators (Hájek type II), based on 10,000 samples. (Nonlinear Case)

n	Parameter	Method	Relative Bias (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	Variance of $\bar{y}_d$	Two-phase	-3.00	-4.45	-4.75
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-51.71	-58.66	-64.06
		Two-phase	-52.85	-59.57	-64.62
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-52.93	-60.55	-63.60
		Two-phase	-53.94	-61.00	-63.80
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-33.49	-39.11	-40.45
		Two-phase	-34.78	-40.48	-41.75
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-22.66	-23.65	-20.83
		Two-phase	-21.96	-23.24	-20.41
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-48.91	-56.15	-61.15
		Two-phase	-50.23	-57.28	-62.02
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-20.54	-23.26	-18.99
		Two-phase	-20.72	-23.62	-19.39
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-51.55	-58.54	-63.91
		Two-phase	-52.71	-59.46	-64.49
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-42.07	-52.74	-49.48
		Two-phase	-43.44	-53.69	-50.50
400	Variance of $\bar{y}_d$	Two-phase	-1.03	-0.81	-2.65
	Variance of $\bar{y}_{e_p}(1)$	Reverse	-33.46	-46.64	-50.28
		Two-phase	-34.52	-46.84	-49.76
	Variance of $\bar{y}_{e_p}(\pi^{-1})$	Reverse	-35.07	-46.66	-51.44
		Two-phase	-36.03	-46.57	-50.50
	Variance of $\bar{y}_{e_{np1}}(1)$	Reverse	-6.11	-6.98	-5.20
		Two-phase	-6.54	-7.47	-4.84
	Variance of $\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	-7.46	-6.76	-5.76
		Two-phase	-7.31	-6.71	-5.51
	Variance of $\bar{y}_{e_{np10}}(1)$	Reverse	-22.98	-34.09	-39.49
		Two-phase	-24.68	-35.20	-39.72
	Variance of $\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	-6.25	-6.01	-3.79
		Two-phase	-6.36	-6.10	-3.30
	Variance of $\bar{y}_{e_{np200}}(1)$	Reverse	-32.85	-46.02	-49.84
		Two-phase	-33.97	-46.28	-49.38
	Variance of $\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	-22.68	-31.57	-38.39
		Two-phase	-24.40	-32.51	-38.29

two-phase method. For the linear predictor, the mean lengths and coverages of 95% confidence intervals from linear logistic regression estimators and the penalized spline logistic regression estimators with a large smoothing parameter perform similarly for each of the three type estimators, as shown in Table 2.11, Table 2.13 and Table 2.15. For the mean lengths and coverages of the nonlinear predictor as shown in Table 2.12, Table 2.14 and Table 2.16, the penalized spline logistic regression estimators (i.e. estimators 4-9) are more efficient than the linear logistic regression estimators (i.e. estimators 2 and 3) for each of the three populations, which the penalized spline logistic regression estimators provide shorter mean lengths and coverages closer to 95% than the logistic linear regression estimators, regardless the choice of the weight k_i.

2.6 Conclusions

In this chapter, we studied the properties of the nonresponse weighting adjustment estimators when the response probability was estimated by maximizing the penalized pseudo-log-likelihood function. Two types of the estimators were considered. The first type was the Horvitz-Thompson type estimator, and the second type was the Hájek type estimator. Both estimators were seen to be consistent. However, the Hájek type estimator was more efficient in reducing the bias and variance.

The penalized spline logistic estimator with $k_i = 1$ tended to be less efficient in bias than $k_i = \pi_i^{-1}$ with the same smoothing parameter. It was seen that in the penalized pseudo-log-likelihood function (2.4), the pseudo-log-likelihood function $\sum_{i \in S} k_i \{R_i \text{logit}(p_i) + \log(1 - p_i)\}$ was much larger as $k_i = \pi_i^{-1}$ compared with $k_i = 1$. Therefore, given the same smoothing parameter λ, the penalty term $\frac{1}{2}\lambda \boldsymbol{\nu}^T \boldsymbol{\Omega} \boldsymbol{\nu}$ had larger penalty on the term $\sum_{i \in S} k_i \{R_i \text{logit}(p_i) + \log(1 - p_i)\}$ with $k_i = 1$ than $k_i = \pi_i^{-1}$ in practice. Therefore, for a given choice of k_i, the penalty value can be selected that gives the desired amount of smoothing.

When the response propensity function followed a linear logistic relationship, the penalized spline logistic regression estimator with a large smoothing parameter remained competitive in bias and variance with a linear logistic regression estimator. When the response propensity function did not follow a linear logistic relationship, the penalized spline logistic regression estimated response

Table 2.11: Mean lengths and coverages of 95% confidence interval estimators (Horvitz-Thompson type), based on 10,000 samples. (Linear case)

n	Point Estimator	Var. Estimation Method	Mean Length			Coverage (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	Two-phase	0.833	0.778	0.709	94.23	94.46	94.78
	$\bar{y}_{e_p}(1)$	Reverse	0.599	0.579	0.541	94.19	93.91	94.17
		Two-phase	0.591	0.570	0.531	93.97	93.63	93.65
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.540	0.523	0.492	93.72	93.81	93.97
		Two-phase	0.536	0.517	0.485	93.62	93.49	93.67
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.571	0.564	0.532	93.43	93.75	93.86
		Two-phase	0.561	0.552	0.520	93.04	93.20	93.33
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.492	0.490	0.473	90.51	92.28	93.06
		Two-phase	0.484	0.480	0.458	89.85	91.58	92.23
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.589	0.574	0.537	93.96	93.86	94.07
		Two-phase	0.580	0.565	0.528	93.70	93.52	93.57
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.500	0.498	0.477	92.33	92.99	93.65
		Two-phase	0.495	0.490	0.466	91.86	92.56	93.00
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.598	0.579	0.540	94.18	93.91	94.16
		Two-phase	0.590	0.570	0.531	93.96	93.65	93.64
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.516	0.510	0.485	93.38	93.75	93.90
		Two-phase	0.511	0.503	0.476	93.06	93.34	93.37
400	$\bar{y}_d$	Two-phase	0.421	0.390	0.356	94.52	94.33	94.61
	$\bar{y}_{e_p}(1)$	Reverse	0.302	0.290	0.269	94.10	94.51	94.72
		Two-phase	0.301	0.289	0.268	94.00	94.43	94.60
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.272	0.263	0.247	94.20	94.62	94.66
		Two-phase	0.271	0.262	0.246	94.15	94.59	94.57
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.289	0.285	0.266	93.71	94.16	94.54
		Two-phase	0.288	0.283	0.264	93.61	93.98	94.39
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.255	0.254	0.241	93.14	93.64	94.25
		Two-phase	0.254	0.253	0.239	93.08	93.54	94.02
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.294	0.287	0.267	93.87	94.42	94.62
		Two-phase	0.293	0.286	0.266	93.75	94.27	94.57
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.257	0.256	0.243	93.64	94.20	94.31
		Two-phase	0.256	0.255	0.241	93.57	94.09	94.25
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.301	0.290	0.269	94.03	94.49	94.72
		Two-phase	0.300	0.289	0.268	93.94	94.41	94.61
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.263	0.259	0.245	94.07	94.43	94.62
		Two-phase	0.262	0.258	0.244	94.02	94.36	94.51

Table 2.12: Mean lengths and coverages of 95% confidence interval estimators (Horvitz-Thompson type), based on 10,000 samples. (Nonlinear case)

n	Point Estimator	Var. Estimation Method	Mean Length			Coverage (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	Two-phase	0.993	0.893	0.774	92.39	93.29	93.69
	$\bar{y}_{e_p}(1)$	Reverse	1.370	1.156	0.873	88.61	89.28	90.12
		Two-phase	1.400	1.176	0.883	89.12	89.67	90.48
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	1.432	1.139	0.851	85.01	85.77	87.58
		Two-phase	1.477	1.174	0.870	86.36	86.93	88.61
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.763	0.714	0.616	89.47	89.53	90.94
		Two-phase	0.753	0.703	0.605	88.80	88.88	90.35
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.522	0.510	0.488	86.85	88.88	91.12
		Two-phase	0.514	0.496	0.469	86.20	87.68	89.70
	$\bar{y}_{e_{np10}}(1)$	Reverse	1.184	1.024	0.792	88.25	88.71	89.73
		Two-phase	1.200	1.035	0.795	88.40	88.92	90.00
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.549	0.533	0.501	89.79	90.17	91.70
		Two-phase	0.544	0.524	0.489	89.54	89.64	90.90
	$\bar{y}_{e_{np200}}(1)$	Reverse	1.359	1.148	0.868	88.59	89.30	90.10
		Two-phase	1.388	1.167	0.877	89.07	89.58	90.46
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.877	0.769	0.628	85.82	86.90	89.03
		Two-phase	0.883	0.773	0.628	86.08	86.96	89.01
400	$\bar{y}_d$	Two-phase	0.505	0.453	0.393	94.73	94.34	94.58
	$\bar{y}_{e_p}(1)$	Reverse	0.959	0.816	0.552	68.98	77.92	85.89
		Two-phase	0.977	0.832	0.564	70.94	79.72	87.29
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.924	0.750	0.518	59.44	72.97	82.42
		Two-phase	0.946	0.766	0.531	61.50	74.87	84.20
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.346	0.330	0.293	93.64	93.69	93.64
		Two-phase	0.345	0.328	0.292	93.46	93.47	93.57
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.284	0.276	0.257	92.86	93.56	94.03
		Two-phase	0.284	0.274	0.255	92.81	93.34	93.84
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.575	0.517	0.386	78.07	84.52	90.00
		Two-phase	0.579	0.522	0.392	78.66	85.02	90.73
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.290	0.281	0.261	93.52	93.87	93.96
		Two-phase	0.290	0.281	0.260	93.57	93.80	93.90
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.923	0.789	0.536	69.14	78.28	86.19
		Two-phase	0.940	0.803	0.548	71.11	79.89	87.48
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.486	0.429	0.338	70.45	80.83	88.02
		Two-phase	0.488	0.432	0.344	70.87	81.30	88.82

Table 2.13: Mean lengths and coverages of 95% confidence interval estimators (Hájek type I), based on 10,000 samples. (Linear case)

n	Point Estimator	Var. Estimation Method	Mean Length $\rho=0.0$	Mean Length $\rho=0.3$	Mean Length $\rho=0.6$	Coverage (%) $\rho=0.0$	Coverage (%) $\rho=0.3$	Coverage (%) $\rho=0.6$
100	$\bar{y}_d$	Two-phase	0.518	0.528	0.544	93.90	93.87	93.36
	$\bar{y}_{e_p}(1)$	Reverse	0.518	0.518	0.507	93.67	93.57	93.27
		Two-phase	0.515	0.515	0.503	93.61	93.43	93.19
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.517	0.516	0.499	93.72	93.49	93.55
		Two-phase	0.515	0.513	0.495	93.42	93.39	93.42
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.505	0.503	0.486	93.41	93.31	92.75
		Two-phase	0.503	0.501	0.483	93.24	93.19	92.57
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.481	0.478	0.457	92.19	92.07	91.22
		Two-phase	0.483	0.479	0.458	92.22	92.11	91.35
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.514	0.513	0.500	93.57	93.49	93.12
		Two-phase	0.511	0.510	0.496	93.49	93.43	92.98
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.492	0.490	0.467	92.77	92.78	92.26
		Two-phase	0.492	0.489	0.467	92.81	92.76	92.31
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.518	0.518	0.507	93.67	93.56	93.27
		Two-phase	0.515	0.514	0.503	93.58	93.44	93.19
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.505	0.503	0.481	93.42	93.27	93.10
		Two-phase	0.504	0.501	0.479	93.27	93.05	92.98
400	$\bar{y}_d$	Two-phase	0.262	0.269	0.281	94.62	94.91	94.50
	$\bar{y}_{e_p}(1)$	Reverse	0.262	0.264	0.260	94.56	94.61	94.15
		Two-phase	0.262	0.264	0.260	94.54	94.59	94.13
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.262	0.263	0.255	94.46	94.60	94.14
		Two-phase	0.262	0.263	0.255	94.47	94.59	94.08
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.257	0.258	0.248	94.19	94.37	93.73
		Two-phase	0.257	0.257	0.248	94.16	94.31	93.76
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.253	0.253	0.239	93.72	93.94	93.31
		Two-phase	0.254	0.253	0.240	93.72	93.93	93.26
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.260	0.260	0.253	94.41	94.56	93.94
		Two-phase	0.260	0.260	0.253	94.41	94.51	93.96
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.256	0.255	0.242	93.96	94.13	93.65
		Two-phase	0.256	0.255	0.242	93.92	94.15	93.65
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.262	0.264	0.259	94.55	94.63	94.14
		Two-phase	0.262	0.263	0.259	94.53	94.60	94.12
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.259	0.259	0.247	94.29	94.43	94.07
		Two-phase	0.259	0.258	0.247	94.26	94.36	94.02

Table 2.14: Mean lengths and coverages of 95% confidence interval estimators (Hájek type I), based on 10,000 samples. (Nonlinear case)

n	Point Estimator	Var. Estimation Method	Mean Length			Coverage (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	Two-phase	0.576	0.571	0.587	93.10	92.53	92.26
	$\bar{y}_{e_p}(1)$	Reverse	0.765	0.789	0.832	91.60	90.64	89.65
		Two-phase	0.755	0.780	0.824	91.35	90.36	89.45
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.778	0.822	0.836	91.06	90.14	88.83
		Two-phase	0.769	0.817	0.832	90.62	89.83	88.32
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.594	0.590	0.574	90.36	89.43	89.85
		Two-phase	0.588	0.583	0.567	90.14	89.17	89.39
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.506	0.491	0.465	89.42	88.90	88.86
		Two-phase	0.508	0.492	0.467	89.59	89.11	88.94
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.720	0.736	0.761	91.11	90.11	89.07
		Two-phase	0.710	0.726	0.752	90.88	89.85	88.71
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.531	0.517	0.488	90.48	89.88	90.46
		Two-phase	0.531	0.516	0.487	90.45	89.78	90.33
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.762	0.786	0.827	91.59	90.64	89.63
		Two-phase	0.753	0.776	0.820	91.34	90.32	89.37
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.643	0.658	0.633	90.25	89.39	88.86
		Two-phase	0.635	0.652	0.627	89.81	89.10	88.43
400	$\bar{y}_d$	Two-phase	0.295	0.294	0.306	94.59	94.62	93.75
	$\bar{y}_{e_p}(1)$	Reverse	0.563	0.648	0.698	96.22	93.49	83.55
		Two-phase	0.557	0.645	0.700	95.98	93.24	83.45
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.560	0.641	0.667	95.91	93.19	80.77
		Two-phase	0.554	0.640	0.671	95.66	93.10	81.19
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.298	0.292	0.276	94.18	94.04	94.02
		Two-phase	0.297	0.291	0.277	94.13	94.00	93.95
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.282	0.273	0.254	93.59	93.72	93.42
		Two-phase	0.282	0.273	0.254	93.65	93.72	93.42
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.414	0.440	0.452	94.97	92.28	84.81
		Two-phase	0.409	0.436	0.450	94.69	91.99	84.65
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.288	0.279	0.259	93.88	93.78	94.09
		Two-phase	0.288	0.279	0.260	93.85	93.80	94.12
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.550	0.630	0.677	96.13	93.32	83.46
		Two-phase	0.544	0.627	0.679	95.85	93.06	83.35
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.392	0.410	0.396	94.43	91.30	82.73
		Two-phase	0.387	0.407	0.396	94.14	91.07	82.86

Table 2.15: Mean lengths and coverages of 95% confidence interval estimators (Hájek type II), based on 10,000 samples. (Linear case)

n	Point Estimator	Var. Estimation Method	Mean Length			Coverage (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	Two-phase	0.520	0.529	0.545	93.93	93.95	93.59
	$\bar{y}_{e_p}(1)$	Reverse	0.518	0.518	0.506	93.76	93.56	93.40
		Two-phase	0.516	0.515	0.502	93.53	93.40	93.23
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.517	0.516	0.498	93.69	93.54	93.63
		Two-phase	0.514	0.513	0.495	93.51	93.41	93.45
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.509	0.507	0.489	93.55	93.54	93.01
		Two-phase	0.508	0.505	0.487	93.37	93.38	92.94
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.493	0.489	0.468	92.91	92.76	92.09
		Two-phase	0.495	0.490	0.470	92.96	92.70	92.30
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.515	0.514	0.500	93.70	93.62	93.30
		Two-phase	0.513	0.512	0.497	93.46	93.39	93.16
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.500	0.497	0.475	93.25	93.20	92.81
		Two-phase	0.501	0.497	0.474	93.27	93.11	92.78
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.518	0.518	0.506	93.75	93.56	93.38
		Two-phase	0.515	0.514	0.502	93.53	93.40	93.24
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.509	0.507	0.485	93.65	93.45	93.32
		Two-phase	0.507	0.505	0.482	93.47	93.30	93.22
400	$\bar{y}_d$	Two-phase	0.262	0.269	0.281	94.59	94.96	94.53
	$\bar{y}_{e_p}(1)$	Reverse	0.262	0.264	0.260	94.57	94.65	94.22
		Two-phase	0.262	0.264	0.259	94.47	94.63	94.10
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.262	0.263	0.255	94.47	94.62	94.14
		Two-phase	0.262	0.263	0.255	94.40	94.62	94.13
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.258	0.258	0.249	94.22	94.40	93.85
		Two-phase	0.258	0.258	0.249	94.19	94.28	93.87
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.255	0.254	0.241	93.82	94.11	93.50
		Two-phase	0.255	0.254	0.241	93.81	94.14	93.55
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.260	0.261	0.254	94.45	94.60	94.10
		Two-phase	0.260	0.260	0.253	94.30	94.54	93.95
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.257	0.256	0.243	94.04	94.23	93.73
		Two-phase	0.257	0.256	0.243	94.05	94.23	93.83
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.262	0.264	0.259	94.58	94.63	94.19
		Two-phase	0.262	0.263	0.259	94.46	94.60	94.10
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.259	0.259	0.248	94.30	94.40	94.09
		Two-phase	0.259	0.259	0.247	94.25	94.39	94.07

Table 2.16: Mean lengths and coverages of 95% confidence interval estimators (Hájek type II), based on 10,000 samples. (Nonlinear case)

n	Point Estimator	Var. Estimation Method	Mean Length			Coverage (%)		
			$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.6$
100	$\bar{y}_d$	Two-phase	0.578	0.572	0.590	93.31	93.05	92.63
	$\bar{y}_{e_p}(1)$	Reverse	0.635	0.642	0.660	85.30	82.40	79.22
		Two-phase	0.627	0.634	0.654	84.59	81.86	78.63
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.634	0.649	0.657	84.28	81.72	78.67
		Two-phase	0.626	0.643	0.653	83.70	81.35	78.27
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.569	0.561	0.544	88.91	87.60	87.95
		Two-phase	0.564	0.555	0.539	88.44	87.19	87.48
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.522	0.507	0.481	90.40	89.94	90.00
		Two-phase	0.525	0.508	0.482	90.62	90.15	90.10
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.619	0.622	0.632	85.83	83.29	80.78
		Two-phase	0.611	0.614	0.625	85.33	82.87	80.07
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.538	0.524	0.494	90.84	90.21	90.87
		Two-phase	0.538	0.523	0.494	90.89	90.26	90.69
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.634	0.641	0.659	85.30	82.45	79.23
		Two-phase	0.626	0.633	0.652	84.62	82.00	78.67
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.584	0.586	0.568	86.89	85.51	84.97
		Two-phase	0.577	0.580	0.562	86.50	84.86	84.55
400	$\bar{y}_d$	Two-phase	0.295	0.294	0.306	94.59	94.69	93.93
	$\bar{y}_{e_p}(1)$	Reverse	0.450	0.492	0.532	91.57	82.88	65.66
		Two-phase	0.445	0.489	0.533	91.18	82.57	65.58
	$\bar{y}_{e_p}(\pi^{-1})$	Reverse	0.448	0.490	0.512	90.72	82.10	63.07
		Two-phase	0.443	0.488	0.515	90.30	81.88	63.57
	$\bar{y}_{e_{np1}}(1)$	Reverse	0.296	0.289	0.275	94.16	93.86	93.70
		Two-phase	0.296	0.289	0.275	94.12	93.77	93.77
	$\bar{y}_{e_{np1}}(\pi^{-1})$	Reverse	0.285	0.276	0.256	93.86	93.96	93.68
		Two-phase	0.285	0.276	0.257	93.88	93.95	93.69
	$\bar{y}_{e_{np10}}(1)$	Reverse	0.371	0.385	0.395	92.40	87.29	76.45
		Two-phase	0.366	0.381	0.394	92.01	86.89	76.12
	$\bar{y}_{e_{np10}}(\pi^{-1})$	Reverse	0.289	0.280	0.260	93.88	93.90	94.19
		Two-phase	0.289	0.280	0.260	93.89	93.93	94.29
	$\bar{y}_{e_{np200}}(1)$	Reverse	0.444	0.484	0.522	91.54	83.11	66.22
		Two-phase	0.439	0.481	0.523	91.19	82.79	66.15
	$\bar{y}_{e_{np200}}(\pi^{-1})$	Reverse	0.355	0.366	0.355	91.65	87.13	76.00
		Two-phase	0.351	0.363	0.355	91.27	86.85	76.10

probability improved efficiency in reducing bias and variance to a misspecified parametric response model. As the smoothing parameter increased, the bias and the variance of the penalized spline logistic regression estimator were closer to the linear logistic regression estimator.

CHAPTER 3

IMPROVING SURVEY ESTIMATORS THROUGH WEIGHT SMOOTHING

3.1 Introduction

In academic and government surveys, probability-based sampling continues to be commonly used. Probability-based sampling allows design-based inference about finite population quantities such as population totals and means. In particular, the classical Horvitz and Thompson [1952] estimator is unbiased with respect to the sampling design distribution, for any variable of interest. The variance of this estimator is small if the variable of interest is positively correlated with the inclusion probabilities. However, in most practical surveys, achieving high correlation across many variables is unfeasible because a single design (and hence a single set of weights) is used to produce estimates for all variables in the survey. In addition, while achieving high precision in the survey estimators is desirable, this is often only one of several considerations when selecting a sampling design. Hence, while survey weights are used to account for the sampling design and ensure that the estimators are at least approximately unbiased, the resulting estimators can be inefficient.

Improving the efficiency of estimators following sampling can be achieved in a number of ways, including calibration and model-assisted estimation [Särndal et al., 1992]. These methods can be expressed as weight adjustments and take advantage of relationships between auxiliary variables in the sample and the population to increase efficiency. Survey agencies frequently apply these methods to adjust survey weights prior to release. However, these methods do not target weight stability directly and can result in individual weights that are even larger or smaller than the original design weights, which can still lead to unstable estimators. The problem can be particularly acute for domains, where weight discrepancies can have an overwhelming effect on the stability of the estimators.

The most common approach for dealing directly with large survey weights is weight trimming [Elliot and Little, 2000, Potter, 1990], in which individual weights larger than some value w_0

are reduced to w_0 and the remaining weights are ratio-adjusted so that the sum of the weights is unchanged. This winsorization approach reduces variance at the cost of introducing some bias. This method requires the determination of the winsorization threshold w_0 [Martinoz et al., 2015]. The selection of the value for the threshold is not straightforward in practice and requires some care. Beaumont et al. [2013] proposed robust estimators to downweight influential sample units, but this method again requires the determination of a tuning constant.

Another approach is to reduce the variability of design weights through the application of a functional adjustment to all weights. Chambers [1996] considered a method of ridge shrinkage regression to ensure positive weights. Beaumont and Bocci [2008] and Kim [2010] also discussed ridge regression to avoid extreme weights by considering different objective functions. Beaumont [2008] considered a "generalized design-based approach." It is similar to the other proposed functional adjustments, in the sense that all the weights are jointly modified according to an objective function, but it departs from them by explicitly assuming a model for the relationship between the weight and the survey variable. That is, the survey weights are parametrically modeled as a function of the survey variables and it leads to a single set of smoothed weights. To the extent that this model is correctly specified, it can lead to substantial gains in precision.

While Beaumont [2008] proposed the approach and showed how it can reduce the variance of the estimator under correct model specification, he did not fully explore a number of important aspects. The aim of the current work is to extend Beaumont's results, by completing some of the theoretical investigation of model-based weight smoothing. We also consider a Hájek-type extension of his estimator as well as a replication variance estimation, both of which are widely used in practice.

One particular application where this model-based weight smoothing is particularly promising is in surveys that are subject to "stratum jumpers." Beaumont and Rivest [2009] discuss this in the context of business surveys, where the stratum jumping is the result of inaccurate size information at the time of sampling. The application motivating the current work is the National Survey of College Graduates (NSCG), in which respondents are selected from the American Community Survey

(ACS) sample. The ACS is stratified based on several variables that are thought to be important for the NSCG. The NSCG sample is selected through a stratified design with widely different stratum sampling fractions, depending on the level of interest in that stratum and the stratum size. However, the ACS stratification variables are recorded with error, so that some of the selected respondents will end up in an estimation domain that differs from their ACS stratum. When such a "stratum jumper" with a large weight ends up in a domain in which most respondents have much smaller weights, the quality of the resulting domain estimates can be significantly degraded, in the sense that the responses from the stratum jumper(s) end up dominating the estimates. The application of weight smoothing to address the stratum jumper issue in the NSCG will not be addressed in this chapter, focusing instead on the general properties of the estimator.

In Section 3.2, the smoothed Horvitz-Thompson estimator is introduced. In Section 3.3, the smoothed weight is estimated under the model assumptions. Section 3.4 derives the theoretical properties of the smoothed Horvitz-Thompson estimator, and Section 3.5 describes the associated variance estimation. In Section 3.6, the Hájek version of the weight smoothing estimator is discussed. Section 3.7 describes simulation studies of the smoothed Horvitz-Thompson and Hájek estimators, under both correctly and incorrectly specified weight models. In Section 3.8, replicate variance estimation is discussed. Conclusions are given in Section 3.9.

3.2 The smoothed Horvitz-Thompson estimator

We denote the finite population by $U_N = \{1, \cdots, i, \cdots, N\}$. Let $\mathcal{F}_N = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N\}$ represent the population variables. For each individual, $\boldsymbol{u}_i = (y_i, z_i)^{'}$, where y_i is the study variable of interest for population unit i, and z_i is the design variable for population unit i. For simplicity, we consider a single study variable y_i and a single design variable z_i for now. Suppose that an estimate is needed for the finite population total $T_y = \sum_{i \in U_N} y_i$. A sample S of size n is drawn from U_N according to a probability sampling design $p(\boldsymbol{I}_N | \boldsymbol{Z}_N, \boldsymbol{Y}_N) = p(\boldsymbol{I}_N | \boldsymbol{Z}_N)$, where $\boldsymbol{Z}_N$ is a population vector containing z_i for $i = 1, \cdots, N$, $\boldsymbol{Y}_N$ is a vector containing y_i for $i = 1 \cdots N$ and $\boldsymbol{I}_N = (I_1, \cdots, I_N)$ is a vector of the sample inclusion indicators, $I_i = 1$ if $i \in S$, and $I_i = 0$

otherwise. Let $\hat{Y}^{HT}$ be the Horvitz-Thompson estimator,

$$\hat{Y}^{HT} = \frac{1}{N}\sum_{i\in S} w_i y_i,$$

where $w_i = 1/\pi_i$ is the design weight of unit i and $\pi_i = \mathrm{E}(I_i|\boldsymbol{Z}_N, \boldsymbol{Y}_N) = \mathrm{E}(I_i|\boldsymbol{Z}_N)$. The conditional expectation of $\hat{Y}^{HT}$, conditioning on $\mathcal{F}_N$ is $\bar{Y}$, therefore, the Horvitz-Thompson estimator $\hat{Y}^{HT}$ is an unbiased estimator of $\bar{Y}$.

When design variable z is strongly correlated with the variable of interest y, it is possible to construct an efficient Horvitz-Thompson estimator, using the principle of PPS sampling. However, such a strongly correlated design variable may not be available in many surveys, so that the Horvitz-Thompson estimator, while unbiased, is highly variable. To solve the inefficient Horvitz-Thompson estimator, a smoothed random variable $\tilde{Y}^{SHT}$ was proposed by Beaumont [2008]. This smoothed random variable is of the form

$$\tilde{Y}^{SHT} = \mathrm{E}(\hat{Y}^{HT}\big|\boldsymbol{I}_N, \boldsymbol{Y}_N) = \mathrm{E}\left(\frac{1}{N}\sum_{i\in S} w_i y_i \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N\right) = \frac{1}{N}\sum_{i\in S}\tilde{w}_i y_i,$$

where $\tilde{w}_i = \mathrm{E}(w_i|\boldsymbol{I}_N, \boldsymbol{Y}_N)$ is the smoothed weight for unit $i \in S$. The $\tilde{w}_i$ is not observable, but it can be estimated after specifying a model for the conditional expectation. Once we obtain an estimator $\hat{w}_i$ of $\tilde{w}_i$, we can construct a smoothed Horvitz-Thompson estimator $\hat{Y}_{SHT} = \frac{1}{N}\sum_{i\in S}\hat{w}_i y_i$.

As explained in Beaumont [2008], assuming correct specification of the weight smoothing model, $\tilde{Y}^{SHT}$ continues to be unbiased but will have smaller variance than $\hat{Y}^{HT}$. Speaking somewhat loosely, this is because the portion of the weight w_i that is unrelated to y_i is removed, reducing variance, while the portion of the weight that depends on y_i and is necessary to provide unbiasedness is preserved.

3.3 Estimation of the smoothed weight

To estimate $\tilde{w}_i$, it is assumed that $\tilde{w}_i = f_s(y_i)$, where the function f_s is some function that needs to be estimated from (y_i, w_i) under the sample. Following Beaumont [2008], we consider a shifted log-normal regression model,

$$w_i = 1 + \exp(\boldsymbol{B}_i\boldsymbol{\nu} + \varepsilon_i) \tag{3.1}$$

for $i \in S$, where the vector $\boldsymbol{B}_i = [B_1(y_i), \cdots, B_r(y_i)]$ is a set of basis functions depending on y_i, the vector $\boldsymbol{\nu} = [\nu_1, \nu_2, \cdots, \nu_r]'$ is a vector of unknown coefficients, the ε_i given $\boldsymbol{I}_N$ and $\boldsymbol{Y}_N$ are independently and identically normally distributed with $\mathrm{E}(\varepsilon_i|\boldsymbol{I}_N, \boldsymbol{Y}_N) = 0$ and $\mathrm{Var}(\varepsilon_i|\boldsymbol{I}_N, \boldsymbol{Y}_N) = \sigma^2 > 0$. The shift of 1 is added to the model to ensure that $\tilde{w}_i \geq 1$. The unknown model parameters to be estimated are $\boldsymbol{\nu}$ and σ^2. Note that $\boldsymbol{B}_i = [1, y_i]$ is a special case of this model, corresponding to a simple loglinear regression model. If y_i is itself multivariate, then a multiple loglinear regression model is likewise a special case of model (3.1).

Under this model, the smoothed weight is

$$\tilde{w}_i = \mathrm{E}(w_i|\boldsymbol{I}_N, \boldsymbol{Y}_N) = 1 + \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right).$$

The lognormal model implies a weight distribution that is quite variable and which has a long upper tail. While that distribution will not be appropriate for all surveys, it is a reasonable model for large-scale surveys with complex designs and potential weight stability problems, such as the NSCG.

To obtain an estimator $\hat{w}_i$ of the smoothed weight $\tilde{w}_i$, we estimate $\boldsymbol{\nu}$ and σ^2 by their least squares estimators. However, it will be convenient to consider their maximum likelihood (ML) estimators first, to enable generalization to more complicated models in the future. The log-likelihood function is

$$l(\boldsymbol{\nu}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{\sum_{i\in S}\left\{\log(w_i - 1) - \boldsymbol{B}_i\boldsymbol{\nu}\right\}^2}{2\sigma^2}.$$

Conditional on the sample, the least squares or ML estimator of the parameter $\boldsymbol{\nu}$ given the $\boldsymbol{B}_i$ is

$$\hat{\boldsymbol{\nu}} = \left(\sum_{i\in S}\boldsymbol{B}_i^T\boldsymbol{B}_i\right)^{-1}\left\{\sum_{i\in S}\boldsymbol{B}_i^T\log\left(w_i - 1\right)\right\}.$$

The ML variance estimator of σ^2 is

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i\in S}\left\{\log(w_i - 1) - \boldsymbol{B}_i\hat{\boldsymbol{\nu}}\right\}^2.$$

The estimator $\hat{\boldsymbol{\nu}}$ is unbiased for $\boldsymbol{\nu}$, but $\hat{\sigma}^2_{ML}$ is biased:

$$\begin{aligned}
\mathrm{E}(\hat{\sigma}^2_{ML}|\boldsymbol{I}_N, \boldsymbol{Y}_N) &= \mathrm{E}\left\{ \frac{1}{n}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} - \frac{1}{n}\boldsymbol{\varepsilon}^T\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{\varepsilon} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right\} \\
&= \frac{1}{n}\left[\mathrm{tr}(\boldsymbol{I}_n) - \mathrm{tr}\left\{\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\right\}\right]\sigma^2 \\
&= \frac{n-r}{n}\sigma^2.
\end{aligned}$$

However, the least squares estimator $\hat{\sigma}^2 = \frac{n}{n-r}\hat{\sigma}^2_{ML}$ is unbiased.

The plug-in estimator of $\tilde{w}_i$ is therefore defined as

$$\hat{w}_i = 1 + \exp(\boldsymbol{B}_i\hat{\boldsymbol{\nu}} + \frac{\hat{\sigma}^2}{2}).$$

In the next section, we study the properties of the smoothed HT estimator, with the weights $\hat{w}_i$ created as just described.

3.4 Theoretical properties

Before presenting the properties of the smoothed HT estimator, we state the assumptions we will be using.

A 3.1. *Let $\boldsymbol{B}_i$ be a vector composed of fixed functions of the study variable y_i. Assume that the sequence of $\boldsymbol{B}_i$ and the sequence of y_i are bounded.*

A 3.2. *The inclusion probabilities satisfy $\min_{i\in U}\pi_i \geq \pi^* > 0$ and $\max_{i\neq j}|\pi_{ij} - \pi_i\pi_j| < C_\pi/n$ for all N for some constant C_π. The sample size n is non-random and satisfies $n/N \to f$ with $0 < f < 1$.*

A 3.3. *Assume that the matrix $\sum_{i\in S}\boldsymbol{B}_i^T\boldsymbol{B}_i$ is nonsingular for all samples S. Assume that the matrix $\sum_{i\in U}\pi_i\boldsymbol{B}_i^T\boldsymbol{B}_i$ is nonsingular.*

A 3.4. *Assume that weight model parameters $\boldsymbol{\nu}$ and σ^2 satisfy $\|\boldsymbol{\nu}\| < C_\nu$ and $0 < \sigma^2 < C_\sigma$ for two fixed constants C_ν and C_σ, for all samples S.*

Theorem 3.1. *Assume that (A3.1)-(A3.4) hold. The smoothed estimator $\hat{Y}^{SHT}$ can be linearized, in the sense that*

$$\hat{Y}^{SHT} = \hat{Y}_l^{SHT} + o_p(n^{-1/2}),$$

where

$$\hat{Y}_l^{SHT} = \frac{1}{N}\sum_{i\in S} y_i + \frac{1}{N}\sum_{i\in S} y_i \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right)\left\{1 + \boldsymbol{B}_i\left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\right) + \frac{1}{2}\left(\hat{\sigma}^2 - \sigma^2\right)\right\}.$$

Proof of Theorem 3.1: We first show that $\hat{\boldsymbol{\nu}}$ and $\hat{\sigma}^2$ are consistent for $\boldsymbol{\nu}$ and σ^2, respectively, under the weight model. For $\hat{\boldsymbol{\nu}}$,

$$\hat{\boldsymbol{\nu}} - \boldsymbol{\nu} = \left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\boldsymbol{B}_i\right)^{-1}\left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\varepsilon_i\right).$$

Applying assumption (A3.1), (A3.2) and (A3.3), we find that

$$\left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\boldsymbol{B}_i\right)^{-1} = \left(\frac{1}{n}\sum_{i\in U}\pi_i\boldsymbol{B}_i^T\boldsymbol{B}_i\right)^{-1} + O_p(n^{-1/2}). \tag{3.2}$$

Consider

$$\mathrm{E}\left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\varepsilon_i\middle|\boldsymbol{Y}_N\right) = \mathrm{E}\left\{\mathrm{E}\left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\varepsilon_i\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} = 0$$

and

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\varepsilon_i\middle|\boldsymbol{Y}_N\right) = \mathrm{E}\left(\frac{1}{n^2}\sum_{i\in S}\boldsymbol{B}_i^T\boldsymbol{B}_i\sigma^2\middle|\boldsymbol{Y}_N\right).$$

Applying assumption (A3.1), (A3.2) and (A3.4), we immediately have $\mathrm{Var}\left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\varepsilon_i\middle|\boldsymbol{Y}_N\right) = O_p(n^{-1})$. By Chebyshev's inequality, we have

$$\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\varepsilon_i = O_p\left(n^{-1/2}\right) \tag{3.3}$$

with respect to the model and the sampling design. Therefore, from (3.2) and (3.3), it follows

$$\hat{\boldsymbol{\nu}} - \boldsymbol{\nu} = O_p(n^{-1/2}). \tag{3.4}$$

For $\hat{\sigma}^2$, consider first

$$\begin{aligned}\hat{\sigma}^2_{ML} &= \frac{1}{n}\sum_{i\in S}\left\{\varepsilon_i - \boldsymbol{B}_i(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\right\}^2 \\ &= \frac{1}{n}\sum_{i\in S}\varepsilon_i^2 - \left(\frac{1}{n}\sum_{i\in S}\varepsilon_i^T\boldsymbol{B}_i\right)\left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\boldsymbol{B}_i\right)^{-1}\left(\frac{1}{n}\sum_{i\in S}\boldsymbol{B}_i^T\varepsilon_i\right).\end{aligned}$$

By Chebyshev's inequality again, we have $\frac{1}{n}\sum_{i\in S}\varepsilon_i^2 = \sigma^2 + O_p(n^{-1/2})$. Applying (3.2) and (3.3), it follows

$$\hat{\sigma}^2_{ML} - \sigma^2 = O_p(n^{-1/2})$$

and hence,

$$\hat{\sigma}^2 - \sigma^2 = O_p(n^{-1/2}). \tag{3.5}$$

Applying a Taylor expansion of $\hat{Y}^{SHT}$ as a function of $\hat{\boldsymbol{\nu}}, \hat{\sigma}^2$ in the neighborhood of $\boldsymbol{\nu}, \sigma^2$, we obtain

$$\begin{aligned}\hat{Y}^{SHT} &= \tilde{Y}^{SHT} + \left\{\left.\frac{\partial\hat{Y}^{SHT}}{\partial\hat{\boldsymbol{\nu}}}\right|_{\left(\begin{smallmatrix}\hat{\boldsymbol{\nu}}\\ \hat{\sigma}^2\end{smallmatrix}\right)=\left(\begin{smallmatrix}\boldsymbol{\nu}\\ \sigma^2\end{smallmatrix}\right)}\right\}^T(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}) \\ &\quad + \left\{\left.\frac{\partial\hat{Y}^{SHT}}{\partial\hat{\sigma}^2}\right|_{\left(\begin{smallmatrix}\hat{\boldsymbol{\nu}}\\ \hat{\sigma}^2\end{smallmatrix}\right)=\left(\begin{smallmatrix}\boldsymbol{\nu}\\ \sigma^2\end{smallmatrix}\right)}\right\}^T\left(\hat{\sigma}^2-\sigma^2\right) + o_p\left(\left|\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}, \hat{\sigma}^2-\sigma^2\right|\right) \\ &= \tilde{Y}^{SHT} + \frac{1}{N}\sum_{i\in S} y_i \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right)\boldsymbol{B}_i(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}) \\ &\quad + \frac{1}{N}\sum_{i\in S} y_i \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right)\times\frac{1}{2}\left(\hat{\sigma}^2-\sigma^2\right) + o_p\left(\left|\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}, \hat{\sigma}^2-\sigma^2\right|\right).\end{aligned}$$

From this, (3.4) and (3.5), we immediately obtain

$$\hat{Y}^{SHT} = \hat{Y}_l^{SHT} + o_p(n^{-1/2})$$

as desired. □

In Theorem 3.1, we approximated the estimator $\hat{Y}^{SHT}$ by $\hat{Y}_l^{SHT}$, with both estimators sharing the same asymptotic distribution. We now consider the bias and variance properties of $\hat{Y}_l^{SHT}$ with respect to the sampling design and the weight model.

Theorem 3.2. *Assume that (A3.1)-(A3.4) hold. The linearized estimator $\hat{Y}_l^{SHT}$ is unbiased for the population mean and*

$$\begin{aligned} Var\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{Y}_N\right) &= E\left\{Var\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} \\ &\quad +E\left[\left\{Var\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)-Var\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\right\}\middle|\boldsymbol{Y}_N\right], \end{aligned}$$

where

$$\begin{aligned} Var\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right) &= \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}(\pi_{ij}-\pi_i\pi_j)w_iw_jy_iy_j \\ Var\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right) &= \frac{1}{N^2}\sum_{i\in S}\sum_{j\in S}y_iy_j\exp\left\{\left(\boldsymbol{B}_i+\boldsymbol{B}_j\right)\boldsymbol{\nu}+\sigma^2\right\}\left\{\boldsymbol{B}_i\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}_j^T\sigma^2\right. \\ &\quad \left.+\frac{1}{2\left(n-r\right)}\left(\sigma^2\right)^2\right\} \\ Var\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right) &= \frac{1}{N^2}\sum_{i\in S}y_i^2\left(e^{\sigma^2}-1\right)e^{2\boldsymbol{B}_i\boldsymbol{\nu}+\sigma^2}. \end{aligned}$$

Proof of Theorem 3.2: The unbiasedness of $\hat{Y}_l^{SHT}$ follows directly from the conditional unbiasedness of $\hat{\boldsymbol{\nu}}$ and $\hat{\sigma}^2$ and the unbiasedness of $\tilde{Y}^{SHT}$. The variance of $\hat{Y}_l^{SHT}$ is written

$$\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{Y}_N\right)=\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}+\mathrm{Var}\left\{\mathrm{E}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}. \quad (3.6)$$

The second component of (3.6) is

$$\mathrm{Var}\left\{\mathrm{E}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}=\mathrm{Var}\left(\tilde{Y}^{SHT}\middle|\boldsymbol{Y}_N\right). \quad (3.7)$$

Taking advantage of the fact that

$$\begin{aligned} \mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Y}_N\right) &= \mathrm{Var}\left\{\mathrm{E}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}+\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} \\ &= \mathrm{Var}\left(\tilde{Y}^{SHT}\middle|\boldsymbol{Y}_N\right)+\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} \\ &= \mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}, \end{aligned} \quad (3.8)$$

we have

$$\mathrm{Var}\left(\tilde{Y}^{SHT}\middle|\boldsymbol{Y}_N\right)=\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}-\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}. \quad (3.9)$$

The total variance of $\hat{Y}_l^{SHT}$ can therefore be written as

$$\begin{aligned}\text{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{Y}_N\right) &= \text{E}\left\{\text{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} \\ &\quad +\text{E}\left[\left\{\text{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)-\text{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\right\}\middle|\boldsymbol{Y}_N\right]. \quad (3.10)\end{aligned}$$

The variance of Horviz-Thompson estimator conditional on the design in (3.10) is

$$\text{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)=\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}(\pi_{ij}-\pi_i\pi_j)w_iw_jy_iy_j.$$

Under the stated weight model, the conditional variance $\text{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)$ is

$$\begin{aligned}&\text{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right) \\ &= \text{Var}\left[\frac{1}{N}\sum_{i\in S}y_i+\frac{1}{N}\sum_{i\in S}y_i\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\frac{\sigma^2}{2}\right)\left\{1+\boldsymbol{B}_i\left(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}\right)+\frac{1}{2}\left(\hat{\sigma}^2-\sigma^2\right)\right\}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right] \\ &= \frac{1}{N^2}\left\{\sum_{i\in S}y_i\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\frac{\sigma^2}{2}\right)\boldsymbol{B}_i\right\}\text{Var}\left(\hat{\boldsymbol{\nu}}|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\left\{\sum_{j\in S}y_j\exp\left(\boldsymbol{B}_j\boldsymbol{\nu}+\frac{\sigma^2}{2}\right)\boldsymbol{B}_j\right\}^T \\ &\quad +\frac{1}{N^2}\times\frac{1}{4}\left\{\sum_{i\in S}y_i\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\frac{\sigma^2}{2}\right)\right\}\text{Var}\left(\hat{\sigma}^2|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\left\{\sum_{j\in S}y_j\exp\left(\boldsymbol{B}_j\boldsymbol{\nu}+\frac{\sigma^2}{2}\right)\right\}^T \\ &= \frac{1}{N^2}\sum_{i\in S}\sum_{j\in S}y_iy_j\exp\left\{\left(\boldsymbol{B}_i+\boldsymbol{B}_j\right)\boldsymbol{\nu}+\sigma^2\right\}\left\{\boldsymbol{B}_i\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}_j^T\sigma^2+\frac{1}{2\left(n-r\right)}\left(\sigma^2\right)^2\right\}.\end{aligned}$$

Finally, since $w_i-1=\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\varepsilon_i\right)$ has a log-normal distribution with mean $\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\frac{\sigma^2}{2}\right)$ and variance $\left\{\exp\left(\sigma^2\right)-1\right\}\times\exp\left(2\boldsymbol{B}_i\boldsymbol{\nu}+\sigma^2\right)$, the variance of Horviz-Thompson estimator conditional on the model in (3.10) is

$$\begin{aligned}\text{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right) &= \text{Var}\left(\frac{1}{N}\sum_{i\in S}w_iy_i\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right) \\ &= \frac{1}{N^2}\sum_{i\in S}y_i^2\left(e^{\sigma^2}-1\right)e^{2\boldsymbol{B}_i\boldsymbol{\nu}+\sigma^2}.\end{aligned}$$

□

The next theorem establishes the $\sqrt{n}$-consistency of the estimator with respect to the sampling design and the weight model.

Theorem 3.3. *Assume that (A3.1)-(A3.4) hold. The smoothed estimator $\hat{Y}^{SHT}$ satisfies*

$$\hat{Y}^{SHT} = \bar{Y} + O_p\left(n^{-1/2}\right).$$

Proof of Theorem 3.3: Under assumptions (A1) and (A2), the design variance

$$\text{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N, \boldsymbol{Y}_N\right) < C_1/n$$

for some constant C. Hence, it immediately follows that

$$\text{E}\left\{\text{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} = O(n^{-1}).$$

Considering now Var $\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)$ and Var $\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)$, assumption (A3.1)–(A3.4) ensure that all the terms inside the sample summations remain bounded and hence that

$$\text{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right) < C_2/n$$

and

$$\text{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right) < C_3/n$$

for all samples. We conclude

$$\text{E}\left\{\text{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} = O(n^{-1}),$$

and

$$\text{E}\left\{\text{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} = O(n^{-1}).$$

Together with the unbiasedness obtained in Theorem 3.2, these lead to

$$\hat{Y}_l^{SHT} = \bar{Y} + O_p(n^{-1/2}),$$

Finally, combined with Theorem 3.1, we obtain that

$$\hat{Y}^{SHT} = \bar{Y} + O_p(n^{-1/2}),$$

that is, $\hat{Y}^{SHT}$ is a $\sqrt{n}$-consistent estimator for $\bar{Y}$. □

3.5 Variance estimation

Starting from Theorem 2, a plug-in estimator for the total variance can be constructed as

$$\hat{\mathrm{V}}\left(\hat{Y}_l^{SHT}\right) = \hat{\mathrm{V}}_1 + \left\{\hat{\mathrm{V}}_2 - \hat{\mathrm{V}}_3\right\}, \tag{3.11}$$

where

$$\begin{aligned}
\hat{\mathrm{V}}_1 &= \frac{1}{N^2}\sum_{i\in S}\sum_{j\in S}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_{ij}} w_i w_j y_i y_j, && (3.12)\\
\hat{\mathrm{V}}_2 &= \frac{1}{N^2}\sum_{i\in S}\sum_{j\in S} y_i y_j \exp\left\{\left(\boldsymbol{B}_i+\boldsymbol{B}_j\right)\hat{\boldsymbol{\nu}}+\hat{\sigma}^2\right\} \\
&\quad \times\left\{\boldsymbol{B}_i\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}_j^T\hat{\sigma}^2+\frac{1}{2\left(n-r\right)}\left(\hat{\sigma}^2\right)^2\right\}, && (3.13)
\end{aligned}$$

and

$$\hat{\mathrm{V}}_3 = \frac{1}{N^2}\sum_{i\in S} y_i^2\left(e^{\hat{\sigma}^2}-1\right)e^{2\boldsymbol{B}_i\hat{\boldsymbol{\nu}}+\hat{\sigma}^2}. \tag{3.14}$$

Note that $\hat{V}_1$ is the standard estimator for the variance of a HT estimator, so it can be simplified for most sampling design into a standard form, or approximated by a with-replacement version.

In order to use $\hat{\mathrm{V}}\left(\hat{Y}_l^{SHT}\right)$ to construct asymptotically correct confidence intervals, one would normally try to show that it is a consistent estimator of $\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{Y}_N\right)$ with respect to the sampling design and weight model. However, that is not possible without specifying the distribution of $\boldsymbol{W} = (w_1, w_2, \cdots, w_N)^T$ under the population, i.e. conditioning on $\boldsymbol{Y}_N$ but no longer on $\boldsymbol{Z}_N$ and $\boldsymbol{I}_N$. In specific cases where a model for this distribution can be formulated, then it would indeed be possible to show the consistency of the variance estimator with respect to this model. Here, we will restrict ourselves here to show that $\hat{\mathrm{V}}\left(\hat{Y}_l^{SHT}\right)$ is asymptotically unbiased for $\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{Y}_N\right)$. We will do this by showing that $\hat{\mathrm{V}}_1$, $\hat{\mathrm{V}}_2$ and $\hat{\mathrm{V}}_3$ are asymptotically unbiased for $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$, $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$, $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$, respectively. This is not sufficient for consistency, because we do not address the second asymptotic moment of $\hat{\mathrm{V}}\left(\hat{Y}_l^{SHT}\right)$.

We will use some moment results for the normal distribution in the proof of the main theorem in this Section. They are given in the following two lemmas.

Lemma 3.1. *Suppose that* $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$*, let* $\boldsymbol{A}$*:* $k \times n$ *be matrix of constants, then*

$$E\{\exp(\boldsymbol{A}'\boldsymbol{\varepsilon})\} = \exp\left\{\frac{1}{2}\boldsymbol{A}'\left(\sigma^2\boldsymbol{I}\right)\boldsymbol{A}\right\}.$$

Proof of Lemma 3.1: By the definition of expectation,

$$\begin{aligned}
&\mathrm{E}\{\exp(\boldsymbol{A}'\boldsymbol{\varepsilon})\} \\
&= \int_{-\infty}^{\infty} e^{\boldsymbol{A}'\boldsymbol{\varepsilon}} f(\boldsymbol{\varepsilon})\, d\boldsymbol{\varepsilon} \\
&= \int_{-\infty}^{\infty} \frac{\exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}'\left(\sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{\varepsilon} + \boldsymbol{A}'\boldsymbol{\varepsilon}\right\}}{\left|\sigma^2\boldsymbol{I}\right|^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}} d\boldsymbol{\varepsilon} \\
&= \int_{-\infty}^{\infty} \frac{\exp\left[-\frac{1}{2}\left\{\boldsymbol{\varepsilon} - \left(\sigma^2\boldsymbol{I}\right)\boldsymbol{A}\right\}'\left(\sigma^2\boldsymbol{I}\right)^{-1}\left\{\boldsymbol{\varepsilon} - \left(\sigma^2\boldsymbol{I}\right)\boldsymbol{A}\right\}\right] \cdot \exp\left\{\frac{1}{2}\boldsymbol{A}'\left(\sigma^2\boldsymbol{I}\right)\boldsymbol{A}\right\}}{\left|\sigma^2\boldsymbol{I}\right|^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}} d\boldsymbol{\varepsilon} \\
&= \exp\left\{\frac{1}{2}\boldsymbol{A}'\left(\sigma^2\boldsymbol{I}\right)\boldsymbol{A}\right\},
\end{aligned}$$

because $\frac{\exp\left[-\frac{1}{2}\left\{\boldsymbol{\varepsilon}-\left(\sigma^2\boldsymbol{I}\right)\boldsymbol{A}\right\}'\left(\sigma^2\boldsymbol{I}\right)^{-1}\left\{\boldsymbol{\varepsilon}-\left(\sigma^2\boldsymbol{I}\right)\boldsymbol{A}\right\}\right]}{\left|\sigma^2\boldsymbol{I}\right|^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}}$ is the pdf of $N(\sigma^2\boldsymbol{I}\boldsymbol{A}, \sigma^2\boldsymbol{I})$. □

Lemma 3.2. *Suppose that* $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$*, the matrix* $\boldsymbol{A}$ *is a symmetric matrix of constant and idempotent. Let* $Q = \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}\boldsymbol{\varepsilon}}{\sigma^2}$*, then* $Q \sim \chi^2(r)$*, where* $r = rank(\boldsymbol{A})$*. If* $Q \sim \chi^2(r)$*, then* $E\left(e^{tQ}\right) = (1-2t)^{-\frac{r}{2}}$*,* $E\left(Qe^{tQ}\right) = r(1-2t)^{-\frac{r}{2}-1}$ *and* $E\left(Q^2e^{tQ}\right) = (r+2)\,r(1-2t)^{-\frac{r}{2}-2}$*.*

Proof of Lemma 3.2: The matrix $\boldsymbol{A}$ is symmetric and of rank r, so it can be written as $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$, where $\boldsymbol{U}$: $n \times r$ is semi-orthogonal and $\boldsymbol{D}$: $r \times r$ is diagonal. Note that $\boldsymbol{A}$ is idempotent, then $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A} \Rightarrow \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}' = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}' \Rightarrow \boldsymbol{U}'\boldsymbol{U} = \boldsymbol{D}^{-1}$. Let $\boldsymbol{z} = \frac{\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}'\boldsymbol{\varepsilon}}{\sigma}$, then the distribution of $\boldsymbol{z}$ is $\boldsymbol{z} \sim N_r(\mathbf{0}, \boldsymbol{I}_r)$ and $Q = \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}\boldsymbol{\varepsilon}}{\sigma^2} = \boldsymbol{z}'\boldsymbol{z} \sim \chi^2(r)$, where $r = \text{rank}(\boldsymbol{A})$. Then

$$\begin{aligned}
\mathrm{E}\left(e^{tQ}\right) &= \int_0^{\infty} e^{tQ} f(Q)\, dQ \\
&= \int_0^{\infty} \frac{1}{\Gamma\left(\frac{r}{2}\right)2^{\frac{r}{2}}} Q^{\frac{r}{2}-1} e^{-\left(\frac{1}{2}-t\right)Q} dQ \\
&= 2^{-\frac{r}{2}}\left(\frac{1-2t}{2}\right)^{-\frac{r}{2}} \int_0^{\infty} \frac{1}{\Gamma\left(\frac{r}{2}\right)\left(\frac{2}{1-2t}\right)^{\frac{r}{2}}} Q^{\frac{r}{2}-1} e^{-\left(\frac{1}{2}-t\right)Q} dQ \\
&= (1-2t)^{-\frac{r}{2}},
\end{aligned}$$

because $\frac{1}{\Gamma\left(\frac{r}{2}\right)\left(\frac{2}{1-2t}\right)^{\frac{r}{2}}}Q^{\frac{r}{2}-1}e^{-\left(\frac{1}{2}-t\right)Q}$ is the pdf of $\Gamma\left(\frac{r}{2},\frac{2}{1-2t}\right)$.

$$\begin{aligned}
\mathrm{E}\left(Qe^{tQ}\right) &= \int_0^\infty Qe^{tQ}f\left(Q\right)dQ \\
&= \int_0^\infty \frac{1}{\Gamma\left(\frac{r}{2}\right)2^{\frac{r}{2}}}Q^{\frac{r}{2}}e^{-\left(\frac{1}{2}-t\right)Q}dQ \\
&= \frac{\Gamma\left(\frac{r}{2}+1\right)\left(\frac{2}{1-2t}\right)^{\frac{r}{2}+1}}{\Gamma\left(\frac{r}{2}\right)2^{\frac{r}{2}}}\int_0^\infty \frac{1}{\Gamma\left(\frac{r}{2}+1\right)\left(\frac{2}{1-2t}\right)^{\frac{r}{2}+1}}Q^{\frac{r}{2}}e^{-\left(\frac{1}{2}-t\right)Q}dQ \\
&= r\left(1-2t\right)^{-\frac{r}{2}-1},
\end{aligned}$$

and similarly,

$$\mathrm{E}\left(Q^2e^{tQ}\right) = \left(r+2\right)r\left(1-2t\right)^{-\frac{r}{2}-2}.$$

□

Theorem 3.4. *Assume (A3.1)-(A3.4). The variance estimator* $\hat{V}\left(\hat{Y}_l^{SHT}\right)$ *is asymptotically unbiased for Var* $\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{Y}_N\right)$ *with respect to the sampling design and the weight model.*

Proof of Theorem 3.4: First, consider the expectation of $\hat{\mathrm{V}}_1$,

$$\begin{aligned}
\mathrm{E}\left(\hat{\mathrm{V}}_1\middle|\boldsymbol{Y}_N\right) &= \mathrm{E}\left\{\mathrm{E}\left(\hat{\mathrm{V}}_1\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} \\
&= \mathrm{E}\left\{\mathrm{E}\left(\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}w_iw_jy_iy_jI_iI_j\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} \\
&= \mathrm{E}\left\{\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}\left(\pi_{ij}-\pi_i\pi_j\right)w_iw_jy_iy_j\middle|\boldsymbol{Y}_N\right\} \\
&= \mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}.
\end{aligned}$$

Therefore, $\hat{\mathrm{V}}_1$ is asymptotically unbiased for $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$.

Second, show that $\hat{\mathrm{V}}_2$ is asymptotically unbiased for $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$. Conditional on the sample, we consider $\hat{V}_2$ as a function of the estimated model parameters. Asymptotically, we write $n\hat{V}_2 = g_n(\hat{\boldsymbol{\nu}},\hat{\sigma}^2)$, a sequence of functions mapping $\mathbb{R}^{r+1}$ into $\mathbb{R}$,

$$\begin{aligned}
&g_n\left(\hat{\boldsymbol{\nu}},\hat{\sigma}^2\right) \\
&= \frac{n}{N^2}\sum_{i\in S}\sum_{j\in S}y_iy_j\exp\left\{\left(\boldsymbol{B}_i+\boldsymbol{B}_j\right)\hat{\boldsymbol{\nu}}+\hat{\sigma}^2\right\}\left\{\boldsymbol{B}_i\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}_j^T\hat{\sigma}^2+\frac{1}{2\left(n-r\right)}\left(\hat{\sigma}^2\right)^2\right\}.
\end{aligned}$$

Considering the expectation conditional on the sample, from Lemma 3.1 and Lemma 3.2, we obtain

$$
\begin{aligned}
&\mathrm{E}\left\{ g_n(\hat{\boldsymbol{\nu}}, \hat{\sigma}^2) - n\mathrm{Var}\left(\hat{Y}_l^{SHT} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right) \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right\} \\
&= \frac{n}{N^2} \sum_{i \in S} \sum_{j \in S} y_i y_j \boldsymbol{B}_i \left(\boldsymbol{B}^T \boldsymbol{B} \right)^{-1} \boldsymbol{B}_j^T \times \mathrm{E}\left\{ e^{(\boldsymbol{B}_i + \boldsymbol{B}_j)\hat{\boldsymbol{\nu}}} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right\} \times \mathrm{E}\left\{ \left(\hat{\sigma}^2\right) e^{\hat{\sigma}^2} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right\} \\
&\quad + \frac{n}{N^2} \sum_{i \in S} \sum_{j \in S} y_i y_j \frac{1}{2(n-r)} \times \mathrm{E}\left\{ e^{(\boldsymbol{B}_i + \boldsymbol{B}_j)\hat{\boldsymbol{\nu}}} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right\} \times \mathrm{E}\left\{ \left(\hat{\sigma}^2\right)^2 e^{\hat{\sigma}^2} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right\} \\
&\quad - \frac{n}{N^2} \sum_{i \in S} \sum_{j \in S} y_i y_j e^{(\boldsymbol{B}_i + \boldsymbol{B}_j)\boldsymbol{\nu} + \sigma^2} \times \left\{ \boldsymbol{B}_i \left(\boldsymbol{B}^T \boldsymbol{B} \right)^{-1} \boldsymbol{B}_j^T \sigma^2 + \frac{1}{2(n-r)} \left(\sigma^2\right)^2 \right\} \\
&= \frac{n}{N^2} \sum_{i \in S} \sum_{j \in S} y_i y_j \boldsymbol{B}_i \left(\boldsymbol{B}^T \boldsymbol{B} \right)^{-1} \boldsymbol{B}_j^T \times e^{(\boldsymbol{B}_i + \boldsymbol{B}_j)\boldsymbol{\nu}} \times \sigma^2 \\
&\quad \times \left\{ e^{\frac{1}{2}(\boldsymbol{B}_i + \boldsymbol{B}_j)(\boldsymbol{B}^T \boldsymbol{B})^{-1}(\boldsymbol{B}_i + \boldsymbol{B}_j)^T \sigma^2} \times \left(1 - \frac{2\sigma^2}{n-r} \right)^{-\frac{n-r}{2}-1} - e^{\sigma^2} \right\} \\
&\quad + \frac{n}{N^2} \sum_{i \in S} \sum_{j \in S} y_i y_j \frac{1}{2(n-r)} \times e^{(\boldsymbol{B}_i + \boldsymbol{B}_j)^T \boldsymbol{\nu}} \times \left(\sigma^2\right)^2 \\
&\quad \times \left\{ e^{\frac{1}{2}(\boldsymbol{B}_i + \boldsymbol{B}_j)(\boldsymbol{B}^T \boldsymbol{B})^{-1}(\boldsymbol{B}_i + \boldsymbol{B}_j)^T \sigma^2} \times \frac{n-r+2}{n-r} \times \left(1 - \frac{2\sigma^2}{n-r} \right)^{-\frac{n-r}{2}-2} - e^{\sigma^2} \right\}.
\end{aligned}
$$

Since

$$
\lim_{n \to \infty} e^{\frac{1}{2}(\boldsymbol{B}_i + \boldsymbol{B}_j)(\boldsymbol{B}^T \boldsymbol{B})^{-1}(\boldsymbol{B}_i + \boldsymbol{B}_j)^T \sigma^2} \times \left(1 - \frac{2\sigma^2}{n-r} \right)^{-\frac{n-r}{2}-1} = e^{\sigma^2}
$$

and

$$
\lim_{n \to \infty} e^{\frac{1}{2}(\boldsymbol{B}_i + \boldsymbol{B}_j)(\boldsymbol{B}^T \boldsymbol{B})^{-1}(\boldsymbol{B}_i + \boldsymbol{B}_j)^T \sigma^2} \times \frac{n-r+2}{n-r} \times \left(1 - \frac{2\sigma^2}{n-r} \right)^{-\frac{n-r}{2}-2} = e^{\sigma^2},
$$

we can write

$$
e^{\frac{1}{2}(\boldsymbol{B}_i + \boldsymbol{B}_j)(\boldsymbol{B}^T \boldsymbol{B})^{-1}(\boldsymbol{B}_i + \boldsymbol{B}_j)^T \sigma^2} \times \left(1 - \frac{2\sigma^2}{n-r} \right)^{-\frac{n-r}{2}-1} - e^{\sigma^2} = o(1)
$$

and

$$
e^{\frac{1}{2}(\boldsymbol{B}_i + \boldsymbol{B}_j)(\boldsymbol{B}^T \boldsymbol{B})^{-1}(\boldsymbol{B}_i + \boldsymbol{B}_j)^T \sigma^2} \times \frac{n-r+2}{n-r} \times \left(1 - \frac{2\sigma^2}{n-r} \right)^{-\frac{n-r}{2}-2} - e^{\sigma^2} = o(1).
$$

As assumptions (A3.1) - (A3.4) ensure that all the terms inside the sample summations remain bounded,

$$
\mathrm{E}\left(g_n\left(\hat{\boldsymbol{\nu}}, \hat{\sigma}^2\right) \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right) = n\mathrm{Var}\left(\hat{Y}_l^{SHT} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N \right) + o(1),
$$

that is

$$\mathrm{E}\left(\hat{\mathrm{V}}_2\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)=\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)+o\left(n^{-1}\right). \tag{3.15}$$

Because of the boundedness assumptions, this result continues to hold for all samples, and hence

$$\mathrm{E}\left(\hat{\mathrm{V}}_2\middle|\boldsymbol{Y}_N\right)=\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}+o\left(n^{-1}\right).$$

Thus, $\hat{\mathrm{V}}_2$ is asymptotically unbiased for $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$.

Third, we need to show that $\hat{\mathrm{V}}_3$ is asymptotically unbiased for $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$. As before, we let $f_n(\hat{\boldsymbol{\nu}},\hat{\sigma}^2)$ be a sequence of functions mapping $\mathbb{R}^{r+1}$ into $\mathbb{R}$, such that

$$f_n(\hat{\boldsymbol{\nu}},\hat{\sigma}^2)=n\hat{\mathrm{V}}_3=\frac{n}{N^2}\sum_{i\in S}y_i^2\left(e^{\hat{\sigma}^2}-1\right)e^{2\boldsymbol{B}_i\hat{\boldsymbol{\nu}}+\hat{\sigma}^2}.$$

Consider the expectation conditional on the sample, from Lemma 3.1 and Lemma 3.2, we obtain

$$\begin{aligned}
&\mathrm{E}\left\{f_n(\hat{\boldsymbol{v}},\hat{\sigma}^2)-n\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right\}\\
&= \frac{n}{N^2}\sum_{i\in S}y_i^2\left\{\mathrm{E}\left(e^{2\boldsymbol{B}_i\hat{\boldsymbol{\nu}}+2\hat{\sigma}^2}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)-\mathrm{E}\left(e^{2\boldsymbol{B}_i\hat{\boldsymbol{\nu}}+\hat{\sigma}^2}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\right\}\\
&\quad -\frac{n}{N^2}\sum_{i\in S}y_i^2(e^{\sigma^2}-1)e^{2\boldsymbol{B}_i\boldsymbol{\nu}+\sigma^2}\\
&= \frac{n}{N^2}\sum_{i\in S}y_i^2e^{2\boldsymbol{B}_i\boldsymbol{\nu}}e^{2\boldsymbol{B}_i(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}_i^T\sigma^2}\times\left\{\left(1-\frac{4\sigma^2}{n-r}\right)^{-\frac{n-r}{2}}-\left(1-\frac{2\sigma^2}{n-r}\right)^{-\frac{n-r}{2}}\right\}\\
&\quad -\frac{n}{N^2}\sum_{i\in S}y_i^2(e^{\sigma^2}-1)e^{2\boldsymbol{B}_i\boldsymbol{\nu}+\sigma^2}\\
&= \frac{n}{N^2}\sum_{i\in S}y_i^2e^{2\boldsymbol{B}_i\boldsymbol{\nu}}\left\{\left[e^{2\boldsymbol{B}_i(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}_i^T\sigma^2}\times\left(1-\frac{4\sigma^2}{n-r}\right)^{-\frac{n-r}{2}}-e^{2\sigma^2}\right]\right.\\
&\quad \left.-\left[e^{2\boldsymbol{B}_i(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}_i^T\sigma^2}\times\left(1-\frac{2\sigma^2}{n-r}\right)^{-\frac{n-r}{2}}-e^{\sigma^2}\right]\right\}.
\end{aligned}$$

Since

$$\lim_{n\to\infty}e^{2\boldsymbol{B}_i(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}_i^T\sigma^2}\times\left(1-\frac{4\sigma^2}{n-r}\right)^{-\frac{n-r}{2}}=e^{2\sigma^2}$$

and

$$\lim_{n\to\infty}e^{2\boldsymbol{B}_i(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}_i^T\sigma^2}\times\left(1-\frac{2\sigma^2}{n-r}\right)^{-\frac{n-r}{2}}=e^{\sigma^2},$$

we can write

$$e^{2\boldsymbol{B}_i(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}_i^T\sigma^2} \times \left(1 - \frac{4\sigma^2}{n-r}\right)^{-\frac{n-r}{2}} - e^{2\sigma^2} = o(1)$$

and

$$e^{2\boldsymbol{B}_i(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}_i^T\sigma^2} \times \left(1 - \frac{2\sigma^2}{n-r}\right)^{-\frac{n-r}{2}} - e^{\sigma^2} = o(1).$$

As assumption (A3.1) - (A3.4) ensure that all the terms inside the sample summations remain bounded and hence that

$$\mathrm{E}\left\{f_n(\hat{\boldsymbol{\nu}}, \hat{\sigma}^2)\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right\} = n\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right) + o(1),$$

that is,

$$\mathrm{E}\left(\hat{\mathrm{V}}_3\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right) = \mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right) + o\left(n^{-1}\right). \tag{3.16}$$

Because of the boundedness assumptions, this again continues to hold for all the samples from the population, so that

$$\mathrm{E}\left(\hat{\mathrm{V}}_3\middle|\boldsymbol{Y}_N\right) = \mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} + o\left(n^{-1}\right).$$

Thus, $\hat{\mathrm{V}}_3$ is asymptotically unbiased for $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$.

We have shown that $\hat{\mathrm{V}}_1$, $\hat{\mathrm{V}}_2$, $\hat{\mathrm{V}}_3$ are asymptotically unbiased for $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{Z}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$, $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$ and $\mathrm{E}\left\{\mathrm{Var}\left(\hat{Y}^{HT}\middle|\boldsymbol{I}_N, \boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}$, respectively. That is, the total variance estimator $\hat{V}\left(\hat{Y}_l^{SHT}\right)$ is asymptotically unbiased for $\mathrm{Var}\left(\hat{Y}_l^{SHT}\middle|\boldsymbol{Y}_N\right)$. □

3.6 Smoothed Hájek estimator

In this section, we consider the Hájek version of the weight smoothing estimator, which we will refer to here as the "smoothed Hájek estimator." The Hájek estimator is often more efficient than the Horvitz-Thompson estimator in practice, and as will be shown further below, the smoothed Hájek estimator tends to be more robust to weight model misspecification than the smoothed Horvitz-Thompson estimator. However, we do not develop the detailed asymptotic theory for this estimator here and instead only sketch the expected results.

Let $\hat{Y}^{HA}$ be the Hájek estimator,

$$\hat{Y}^{HA} = \frac{\sum_{i\in S} w_i y_i}{\sum_{i\in S} w_i}.$$

Define the smoothed random variable $\tilde{Y}^{SHA}$ as following

$$\tilde{Y}^{SHA} = \frac{\mathrm{E}\left(\sum_{i\in S} w_i y_i \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N\right)}{\mathrm{E}\left(\sum_{i\in S} w_i \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N\right)} = \frac{\sum_{i\in S} \tilde{w}_i y_i}{\sum_{i\in S} \tilde{w}_i}.$$

Then the smoothed Hájek estimator $\hat{Y}^{SHA}$ is of the form

$$\hat{Y}^{SHA} = \frac{\sum_{i\in S} \hat{w}_i y_i}{\sum_{i\in S} \hat{w}_i}. \tag{3.17}$$

In order to obtain an approximate variance for inference for $\hat{Y}^{SHA}$, we first derive its linearized form. Leting $\tilde{N} = \sum_{i\in S} \tilde{w}_i$ and applying a Taylor expansion, we find that

$$\begin{aligned}
\hat{Y}^{SHA} &= \tilde{Y}^{SHA} + \left\{ \left. \frac{\partial \hat{Y}^{SHA}}{\partial \hat{\boldsymbol{\nu}}} \right|_{\left(\begin{smallmatrix} \hat{\boldsymbol{\nu}} \\ \hat{\sigma}^2 \end{smallmatrix}\right)=\left(\begin{smallmatrix} \boldsymbol{\nu} \\ \sigma^2 \end{smallmatrix}\right)} \right\}^T (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}) \\
&\quad + \left\{ \left. \frac{\partial \hat{Y}^{SHA}}{\partial \hat{\sigma}^2} \right|_{\left(\begin{smallmatrix} \hat{\boldsymbol{\nu}} \\ \hat{\sigma}^2 \end{smallmatrix}\right)=\left(\begin{smallmatrix} \boldsymbol{\nu} \\ \sigma^2 \end{smallmatrix}\right)} \right\}^T (\hat{\sigma}^2 - \sigma^2) + o_p\left(\left|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}, \hat{\sigma}^2 - \sigma^2\right|\right) \\
&= \tilde{Y}^{SHA} + \frac{1}{\tilde{N}} \sum_{i\in S} y_i \exp\left(\boldsymbol{B}_i \boldsymbol{\nu} + \frac{\sigma^2}{2}\right) \left\{ \boldsymbol{B}_i (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}) + \frac{1}{2}(\hat{\sigma}^2 - \sigma^2) \right\} \\
&\quad - \frac{1}{\tilde{N}^2} \tilde{T}_y^{SHT} \times \sum_{i\in S} \exp\left(\boldsymbol{B}_i \boldsymbol{\nu} + \frac{\sigma^2}{2}\right) \left\{ \boldsymbol{B}_i (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}) + \frac{1}{2}(\hat{\sigma}^2 - \sigma^2) \right\} + o_p(n^{-1/2}).
\end{aligned}$$

Letting $A_i = \exp\left(\boldsymbol{B}_i \boldsymbol{\nu} + \frac{\sigma^2}{2}\right) \left\{ \boldsymbol{B}_i (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}) + \frac{1}{2}(\hat{\sigma}^2 - \sigma^2) \right\}$, we rewrite this as

$$\hat{Y}^{SHA} = \hat{Y}_l^{SHA} + o_p(n^{-1/2}),$$

where

$$\hat{Y}_l^{SHA} = \tilde{Y}^{SHA} + \frac{1}{\tilde{N}} \sum_{i\in S} A_i \left(y_i - \tilde{Y}^{SHA}\right).$$

As $\hat{Y}_l^{SHA}$ has the same asymptotic distribution as $\hat{Y}^{SHA}$, we consider the variance of $\hat{Y}_l^{SHA}$ with respect to the sampling design and the weight model. The variance of $\hat{Y}_l^{SHA}$ is written

$$\mathrm{Var}\left(\hat{Y}_l^{SHA} \middle| \boldsymbol{Y}_N\right) = \mathrm{E}\left\{ \mathrm{Var}\left(\hat{Y}_l^{SHA} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N\right) \middle| \boldsymbol{Y}_N \right\} + \mathrm{Var}\left\{ \mathrm{E}\left(\hat{Y}_l^{SHA} \middle| \boldsymbol{I}_N, \boldsymbol{Y}_N\right) \middle| \boldsymbol{Y}_N \right\}. \tag{3.18}$$

As in the previous section, consider first the fact that

$$\begin{aligned}\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{Y}_N\right) &= \text{Var}\left\{\text{E}\left(\hat{Y}^{HA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}+\text{E}\left\{\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}\\ &= \text{E}\left\{\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\},\end{aligned}$$

and hence,

$$\text{Var}\left\{\text{E}\left(\hat{Y}^{HA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}=\text{E}\left\{\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}-\text{E}\left\{\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}.$$

Appling Taylor linearization again around $\tilde{Y}^{HA}$, $\hat{Y}^{HA}$ is approximated as follows

$$\hat{Y}^{HA}\approx\tilde{Y}^{SHA}+\frac{1}{\tilde{N}}\sum_{i\in S}\left(w_i-\tilde{w}_i\right)\left(y_i-\tilde{Y}^{SHA}\right).$$

Then the second component of (3.18) is

$$\begin{aligned}\text{Var}\left\{\text{E}\left(\hat{Y}_l^{SHA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\} &= \text{Var}\left(\tilde{Y}^{SHA}\middle|\boldsymbol{Y}_N\right)\\ &\approx \text{Var}\left\{\text{E}\left(\hat{Y}^{HA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}.\end{aligned}$$

the first equation follows because $\text{E}\left(A_i\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)=0$. Therefore, the total variance of $\hat{Y}_l^{SHA}$ is approximated as

$$\begin{aligned}\text{Var}\left(\hat{Y}_l^{SHA}\middle|\boldsymbol{Y}_N\right) &\approx \text{E}\left\{\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\middle|\boldsymbol{Y}_N\right\}\\ &\quad+\text{E}\left[\left\{\text{Var}\left(\hat{Y}_l^{SHA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)-\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\right\}\middle|\boldsymbol{Y}_N\right]. \quad (3.19)\end{aligned}$$

The variance of Hájek estimator conditional on the design in (3.19) is approximated by

$$\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{Z}_N,\boldsymbol{Y}_N\right)\approx\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}(\pi_{ij}-\pi_i\pi_j)w_iw_j\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right).$$

Let $y_{ic}=y_i-\tilde{Y}^{SHA}$, the conditional variance $\text{Var}\left(\hat{Y}_l^{SHA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)$ in (3.19) is

$$\begin{aligned}\text{Var}\left(\hat{Y}_l^{SHA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right) &= \frac{1}{\tilde{N}^2}\text{Var}\left(\sum_{i\in S}A_iy_{ic}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\\ &= \frac{1}{\tilde{N}^2}\sum_{i\in S}\sum_{j\in S}y_{ic}y_{jc}\exp\left\{\left(\boldsymbol{B}_i+\boldsymbol{B}_j\right)\boldsymbol{\nu}+\sigma^2\right\}\left\{\boldsymbol{B}_i\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}_j^T\sigma^2\right.\\ &\qquad\left.+\frac{1}{2\left(n-r\right)}\left(\sigma^2\right)^2\right\}.\end{aligned}$$

The variance of Hájek estimator conditional on the model in (3.19) is

$$\begin{aligned}
&\text{Var}\left(\hat{Y}^{HA}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\\
&\approx \text{Var}\left(\frac{1}{\tilde{N}}\sum_{i\in S}w_i y_{ic}\middle|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)\\
&= \frac{1}{\tilde{N}^2}\sum_{i\in S}y_{ic}^2\left(e^{\sigma^2}-1\right)e^{2\boldsymbol{B}_i\boldsymbol{\nu}+\sigma^2}.
\end{aligned}$$

Let $\hat{N}=\sum_{i\in S}\hat{w}_i$. Using the above approximations, a plug-in estimator for the total variance is as follows:

$$\hat{\text{V}}\left(\hat{Y}_l^{SHA}\right)=\hat{\text{V}}_1+\left\{\hat{\text{V}}_2-\hat{\text{V}}_3\right\}, \tag{3.20}$$

where

$$\hat{\text{V}}_1 = \frac{1}{N^2}\sum_{i\in S}\sum_{j\in S}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_{ij}}w_i w_j\left(y_i-\hat{Y}^{HA}\right)\left(y_j-\hat{Y}^{HA}\right), \tag{3.21}$$

$$\begin{aligned}
\hat{\text{V}}_2 = &\frac{1}{\hat{N}^2}\sum_{i\in S}\sum_{j\in S}\left(y_i-\hat{Y}^{SHA}\right)\left(y_j-\hat{Y}^{SHA}\right)\exp\left\{\left(\boldsymbol{B}_i+\boldsymbol{B}_j\right)\hat{\boldsymbol{\nu}}+\hat{\sigma}^2\right\}\\
&\times\left\{\boldsymbol{B}_i\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}_j^T\hat{\sigma}^2+\frac{1}{2\left(n-r\right)}\left(\hat{\sigma}^2\right)^2\right\},
\end{aligned} \tag{3.22}$$

and

$$\hat{\text{V}}_3=\frac{1}{\hat{N}^2}\sum_{i\in S}\left(y_i-\hat{Y}^{SHA}\right)^2\left(e^{\hat{\sigma}^2}-1\right)e^{2\boldsymbol{B}_i\hat{\boldsymbol{\nu}}+\hat{\sigma}^2}. \tag{3.23}$$

3.7 Simulation Study

3.7.1 Beaumont's population

We begin by evaluating the performance of the smoothed Horvitz-Thompson and smoothed Hájek estimators under the same simulation setup as in Beaumont [2008]. In this set-up, the weight model is not correctly specified, because the relationship between the survey variables and the weights is implicitly determined. First, a population U of 50,000 units is generated. A design

variable z_i for each population unit i is drawn from an exponential distribution with mean 30, then increased by 0.5. The three variables of interest follow model

$$y_{ki} = \beta_0 + \beta_k z_i + \varepsilon_{ki} \ (k = 1, 2, 3),$$

where $\beta_0 = 30$, and the ε_{ki} are independent normal random variables with mean zero and variance 2000. The constants β_1, β_2 and β_3 are chosen to yield correlation coefficients $\rho_1 = 0$, $\rho_2 = 0.01^{1/2}$ and $\rho_3 = 0.8^{1/2}$, respectively. From the finite population, a set of independent samples of size $n = 100$ and $n = 500$ are selected under pps sampling with replacement. Hence, because of this population construction and design, y_1 is unrelated to the weights, y_2 is weakly related and y_3 is strongly related, but the functional form of the relationship between the survey variables and the weights is unknown.

We start with the smoothed Horvitz-Thompson estimators. Following Beaumont [2008], eight estimators are computed, denoted HT, SHT-U, SHT-1, SHT-2, SHT-3, SHT-1(5), SHT-2(5), SHT-3(5). The smoothed Horvitz-Thompson estimators are obtained using different versions of (3.1). For SHT-U estimator, we use $\boldsymbol{B}_i = 1$. The SHT-k estimators use $\boldsymbol{B}_i = (1, y_{ki})$, and the SHT-$k$(5) are polynomial models of order 5 with $\boldsymbol{B}_i = (1, y_{ki}, y_{ki}^2, y_{ki}^3, y_{ki}^4, y_{ki}^5)$. The Monte Carlo sample sizes are all 10,000 in the simulation.

Table 3.1 presents the relative biases (RB) and relative efficiencies (RE) of the estimators, both expressed as percentages and computed as

$$\text{RB}\left(\hat{Y}\right) = 100\frac{\text{E}\left\{(\hat{Y} - \bar{Y})\middle|\boldsymbol{Z}_N, \boldsymbol{Y}_N\right\}}{\bar{Y}}, \tag{3.24}$$

$$\text{RE}\left(\hat{Y}\right) = 100\frac{\text{E}\left\{(\hat{Y} - \bar{Y})^2\middle|\boldsymbol{Z}_N, \boldsymbol{Y}_N\right\}}{\text{E}\left\{(\hat{Y}^{HT} - \bar{Y})^2\middle|\boldsymbol{Z}_N, \boldsymbol{Y}_N\right\}}, \tag{3.25}$$

with $\hat{Y}$ denoting one of the eight considered estimators. The smoothed estimators exhibit a modest amount of bias for y_1 and y_2, but are still more efficient than the HT estimator for those variables, because of the very large reduction in variance. The results are similar for the linear and the polynomial model specification and regardless of which variables are used in the smoothing model.

Table 3.1: Relative biases and relative efficiency of the smoothed Horvitz-Thompson estimators under Beaumont's population.

n	Estimators	Variable y_1		Variable y_2		Variable y_3	
		RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)
100	HT	0.16	100.00	-0.21	100.00	0.03	100.00
	SHT-U	-7.64	33.66	3.24	34.58	59.01	5616.54
	SHT-1	-7.62	45.08	3.46	35.18	58.84	5587.91
	SHT-2	-7.19	33.66	-7.74	47.92	57.66	5371.16
	SHT-3	-11.76	41.66	-8.00	38.86	4.45	90.12
	SHT-1(5)	-6.32	51.02	4.45	38.63	58.04	5452.77
	SHT-2(5)	-5.87	34.54	-6.87	52.95	56.93	5252.78
	SHT-3(5)	-7.28	46.49	-4.59	46.61	-1.40	57.60
500	HT	0.07	100.00	0.00	100.00	0.03	100.00
	SHT-U	-7.98	56.48	3.15	38.66	59.02	25699.46
	SHT-1	-8.24	70.03	3.20	38.93	58.98	25669.22
	SHT-2	-7.75	55.17	-8.11	79.41	57.79	24642.40
	SHT-3	-12.21	96.21	-8.19	71.46	4.35	195.99
	SHT-1(5)	-8.01	69.67	3.38	40.08	58.84	25554.16
	SHT-2(5)	-7.51	53.92	-8.00	79.66	57.66	24535.94
	SHT-3(5)	-8.15	68.40	-5.16	56.08	-1.69	71.78

These results are somewhat surprising for y_1, since this variable is not related to the weights at all. However, most likely because of the misspecification of the lognormal weight model itself, the weight smoothing estimator ends up biased even in this case. For y_3, the variable most strongly related to the weights, the bias is very large when the smoothing is done with respect to other variables and leads to highly inefficient estimation, but the bias is significantly reduced when the weight model uses the correct variable. The efficiency is improved when the polynomial model is used, relative to a linear model.

We next consider variance estimation and inference. Table 3.2 shows the simulation expectations of $\hat{\mathrm{V}}_1$, $\hat{\mathrm{V}}_2$, $\hat{\mathrm{V}}_3$ and $\hat{\mathrm{V}}$ as defined in (3.12), (3.13), (3.14) and (3.11), respectively, together with the true variances for comparison. The results clearly illustrate the fact that an incorrect model specification leads to very poor performance of the model-based variance estimator for the smoothed Horvitz-Thompson estimators, including substantial overestimation for y_1 and y_2 and even negative estimates in the case of y_3. This is further confirmed in Table 3.3, which presents the relative biases of the variance estimators as percentages, average mean lengths and coverages

Table 3.2: Expectations of estimated variance components for smoothed Horvitz-Thompson estimators, under Beaumont's population.

n	Estimators	Variable y_1					Variable y_2				
		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.
100	HT	99.52	-	-	99.52	96.86	99.78	-	-	99.78	97.72
	SHT-U	99.52	7.08	23.99	82.61	27.38	99.78	11.74	29.32	82.20	32.54
	SHT-1	99.52	18.80	24.65	93.68	38.47	99.78	11.99	29.67	82.10	32.94
	SHT-2	99.52	7.27	24.15	82.65	27.98	99.78	19.47	24.95	94.30	39.66
	SHT-3	99.52	2.05	6.80	94.77	27.99	99.78	2.79	7.57	95.00	30.31
	SHT-1(5)	99.52	24.98	33.56	90.94	45.85	99.78	13.79	32.09	81.48	35.38
	SHT-2(5)	99.52	8.94	26.39	82.07	30.37	99.78	24.63	32.02	92.38	46.10
	SHT-3(5)	99.52	3.08	7.78	94.83	40.30	99.78	3.70	8.38	95.09	43.03
500	HT	19.67	-	-	19.67	19.71	19.91	-	-	19.91	19.33
	SHT-U	19.67	1.33	4.57	16.43	5.43	19.91	2.24	5.62	16.53	6.28
	SHT-1	19.67	3.54	4.58	18.63	7.73	19.91	2.25	5.63	16.53	6.30
	SHT-2	19.67	1.34	4.55	16.46	5.51	19.91	3.67	4.66	18.92	7.47
	SHT-3	19.67	0.37	1.27	18.77	5.64	19.91	0.52	1.42	19.00	5.77
	SHT-1(5)	19.67	3.76	4.90	18.53	7.99	19.91	2.30	5.70	16.51	6.38
	SHT-2(5)	19.67	1.39	4.62	16.43	5.59	19.91	3.89	4.91	18.88	7.73
	SHT-3(5)	19.67	0.42	1.33	18.76	7.54	19.91	0.54	1.46	18.99	7.65

n	Estimators	Variable y_3				
		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.
100	HT	97.27	-	-	97.27	96.14
	SHT-U	97.27	332.53	506.28	-76.47	337.96
	SHT-1	97.27	333.37	505.57	-74.92	339.23
	SHT-2	97.27	322.85	485.80	-65.68	331.39
	SHT-3	97.27	35.72	38.94	94.05	57.84
	SHT-1(5)	97.27	336.30	501.60	-68.03	345.46
	SHT-2(5)	97.27	326.36	483.24	-59.60	337.78
	SHT-3(5)	97.27	26.39	26.66	97.00	52.53
500	HT	19.38	-	-	19.38	19.96
	SHT-U	19.38	64.46	97.43	-13.59	66.87
	SHT-1	19.38	64.49	97.39	-13.53	66.90
	SHT-2	19.38	62.43	93.53	-11.72	64.70
	SHT-3	19.38	6.85	7.47	18.76	11.62
	SHT-1(5)	19.38	64.58	97.27	-13.31	67.25
	SHT-2(5)	19.38	62.56	93.41	-11.48	65.00
	SHT-3(5)	19.38	4.96	5.02	19.32	10.19

of nominal normal 95% confidence interval estimators. The mean length is $2 \times 1.96\sqrt{\hat{\mathrm{V}}}$ and the confidence intervals are $\left(\hat{Y}^{SHT} - 1.96\sqrt{\hat{\mathrm{V}}}, \hat{Y}^{SHT} - 1.96\sqrt{\hat{\mathrm{V}}}\right)$.

We now turn to the Hájek-type estimators. We hypothesize that a Hájek-type estimator will correct for model misspecification effects that result in a "level" mismatch between the original and the smoothed weights, but might still result in improved efficiency if the weight variability not related to the target survey variable(s) is removed. Therefore, we also consider a Hájek type smoothed mean estimator $\hat{Y}^{SHA}$ as defined in (3.17). The results for the estimator are shown in Table 3.4. For variable y_1, we observed that all the smoothed Hájek estimators have small relative biases, with the absolute values are less than 2%, and they are more efficient than the original Hájek estimator. For variable y_2, both SHA-2 and SHA-2(5) estimators now have very small relative biases, and both of them performed well in terms of efficiency. For y_3, however, the estimator continues to be strongly biased, with only SHA-3(5) exhibiting less than 10% bias. We note that the relationship between the weights and y_3 is quite strong, with a high correlation between the design variable and the target variable, but that the log-linear model is still misspecified, so that even the Hájek-type estimator is not able to lead to a well-behaved weight smoothing estimator. In comparison, for a significant but less strong model such as that between the weights and y_2, the model misspecification biasing effect was effectively removed by the Hájek-type estimator.

Table 3.5 presents the expectations of estimated variance components for the smoothed Hájek estimators under Beaumont's population. The variances of the smoothed Hájek estimator $\hat{\mathrm{V}}_1$, $\hat{\mathrm{V}}_2$, $\hat{\mathrm{V}}_3$ and the total expected variance estimator $\hat{\mathrm{V}}$ are defined in (3.21), (3.22), (3.23) and (3.20), respectively. For variable y_1 and variable y_2, the expected variance estimators $\hat{\mathrm{V}}$ for the smoothed Hájek estimators overestimate the true variance. For variable y_3, the expected variance estimators $\hat{\mathrm{V}}$ either overestimate or underestimate the true variance, but they are never close to the true variances. Finally, Table 3.6 presents the relative biases of the variance estimators as percentages, average mean lengths and coverages of 95% confidence interval estimators. The relative biases of the variance estimators are all large, and the coverages of 95% confidence intervals are far away from 95%. This is not surprising since the population distribution of $w_i - 1$ do not follow a mixture

Table 3.3: Relative biases of the variance estimators, average mean lengths and coverages of 95% confidence interval for the smoothed Horvitz-Thompson estimators under Beaumont's population.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	2.75	39.11	91.84	2.10	39.16	90.48
	SHT-U	201.67	35.63	99.89	152.65	35.54	99.38
	SHT-1	143.47	37.94	99.62	149.20	35.52	99.39
	SHT-2	195.38	35.64	99.88	137.76	38.07	99.65
	SHT-3	238.58	38.16	99.96	213.41	38.21	99.89
	SHT-1(5)	98.36	37.38	99.06	130.34	35.39	99.02
	SHT-2(5)	170.23	35.51	99.77	100.40	37.68	99.06
	SHT-3(5)	135.32	38.17	99.38	120.97	38.23	99.41
500	HT	-0.19	17.39	93.11	2.96	17.49	93.54
	SHT-U	202.39	15.89	99.48	163.11	15.94	99.50
	SHT-1	141.04	16.92	98.89	162.51	15.94	99.49
	SHT-2	198.80	15.90	99.49	153.46	17.05	98.64
	SHT-3	232.93	16.98	98.50	229.14	17.09	99.44
	SHT-1(5)	131.88	16.87	98.80	158.89	15.93	99.44
	SHT-2(5)	194.15	15.89	99.48	144.33	17.03	98.43
	SHT-3(5)	148.75	16.98	99.01	148.37	17.08	99.51

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HT	1.18	38.66	91.09
	SHT-U	-122.63	NaN	NA
	SHT-1	-122.09	NaN	NA
	SHT-2	-119.82	NaN	NA
	SHT-3	62.59	38.02	95.42
	SHT-1(5)	-119.69	NaN	NA
	SHT-2(5)	-117.64	NaN	NA
	SHT-3(5)	84.66	38.61	98.80
500	HT	-2.93	17.26	93.12
	SHT-U	-120.33	NaN	NA
	SHT-1	-120.22	NaN	NA
	SHT-2	-118.12	NaN	NA
	SHT-3	61.49	16.98	82.83
	SHT-1(5)	-119.79	NaN	NA
	SHT-2(5)	-117.66	NaN	NA
	SHT-3(5)	89.56	17.23	98.39

Table 3.4: Relative biases and relative efficiency of the smoothed Hájek estimators under Beaumont's population.

n	Estimators	Variable y_1		Variable y_2		Variable y_3	
		RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)
100	HA	-0.10	100.00	0.64	100.00	1.89	100.00
	SHA-U	-1.75	34.31	13.08	66.57	72.84	2964.04
	SHA-1	0.23	53.66	12.94	66.06	72.08	2903.95
	SHA-2	-1.63	34.83	-0.21	52.31	70.55	2783.52
	SHA-3	-0.54	43.07	4.58	45.78	18.88	252.18
	SHA-1(5)	0.31	62.43	12.35	64.79	68.82	2656.05
	SHA-2(5)	-1.47	36.81	-0.61	61.81	67.46	2553.96
	SHA-3(5)	-0.59	55.56	3.26	54.76	7.44	107.55
500	HA	-0.12	100.00	0.30	100.00	0.44	100.00
	SHA-U	-1.73	30.97	12.84	174.13	72.93	13535.29
	SHA-1	0.20	47.45	12.80	173.25	72.76	13471.15
	SHA-2	-1.61	31.08	-0.75	47.47	71.15	12882.95
	SHA-3	-0.57	37.51	4.29	53.37	18.75	942.74
	SHA-1(5)	0.23	49.39	12.67	170.74	72.12	13237.24
	SHA-2(5)	-1.57	31.21	-0.74	49.11	70.58	12678.30
	SHA-3(5)	-0.54	45.96	3.02	53.71	7.39	205.48

log-normal distribution, therefore, the model-dependent variance component estimators as shown in (3.22) and (3.23) do not hold for an arbitrary population. Hence, variance estimators based on the weight smoothing model continue to be inappropriate under model misspecification, despite the improvement in the estimators themselves.

3.7.2 Simulation under correct model

We now consider a simulation in which the relationship between the weight and the target variable are correctly specified in the sample. Consider weight model (3.1) again, and letting $x_i = w_i - 1$, then $x_i|y_i$ has a Log-normal distribution of $lN(\boldsymbol{B}_i\boldsymbol{\nu}, \sigma^2)$ with $\mathrm{E}\left(x_i|y_i\right) = \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right)$ under the sample. However, we are faced with an additional difficulty: in order for x_i to have a Log-normal distribution in the sample, we first need the population distribution of x_i, so that we can generate the population-level weights and hence inclusion probabilities. We apply a result of Pfeffermann and Sverchkov [1999], who showed the following relationship between the sample

Table 3.5: Expectations of estimated variance components for the smoothed Hájek estimators under Beaumont's population.

n	Estimators	Variable y_1					Variable y_2				
		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	V	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	V
100	HA	58.83	-	-	58.83	61.14	59.94	-	-	59.94	61.82
	SHA-U	58.83	0.00	18.24	40.59	20.71	59.94	0.00	18.51	41.43	20.30
	SHA-1	58.83	12.80	18.45	53.18	32.80	59.94	0.13	18.63	41.45	20.43
	SHA-2	58.83	0.13	18.13	40.83	21.06	59.94	12.72	18.60	54.06	32.36
	SHA-3	58.83	0.02	5.53	53.33	26.31	59.94	0.04	5.55	54.44	25.76
	SHA-1(5)	58.83	16.85	24.34	51.33	38.16	59.94	1.13	19.54	41.53	21.47
	SHA-2(5)	58.83	1.10	18.98	40.95	22.32	59.94	16.92	24.29	52.57	38.20
	SHA-3(5)	58.83	0.57	5.40	54.00	33.94	59.94	0.54	5.36	55.12	32.58
500	HA	13.81	-	-	13.81	13.68	13.85	-	-	13.85	13.85
	SHA-U	13.81	0.00	3.70	10.11	3.97	13.85	0.00	3.75	10.11	4.02
	SHA-1	13.81	2.60	3.71	12.70	6.49	13.85	0.01	3.75	10.11	4.02
	SHA-2	13.81	0.01	3.66	10.16	4.02	13.85	2.58	3.74	12.69	6.51
	SHA-3	13.81	0.00	1.12	12.70	5.10	13.85	0.00	1.12	12.74	5.15
	SHA-1(5)	13.81	2.82	4.04	12.59	6.75	13.85	0.04	3.79	10.10	4.06
	SHA-2(5)	13.81	0.04	3.69	10.16	4.05	13.85	2.79	3.96	12.68	6.74
	SHA-3(5)	13.81	0.02	1.06	12.77	6.26	13.85	0.03	1.06	12.82	6.33

n	Estimators	Variable y_3				
		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	V
100	HA	237.19	-	-	237.19	263.01
	SHA-U	237.19	0.00	163.50	73.70	179.63
	SHA-1	237.19	1.15	163.19	75.15	179.94
	SHA-2	237.19	3.16	155.41	84.94	176.62
	SHA-3	237.19	10.30	22.28	225.21	154.09
	SHA-1(5)	237.19	11.32	169.97	78.54	187.96
	SHA-2(5)	237.19	13.04	162.54	87.70	184.81
	SHA-3(5)	237.19	22.66	25.05	234.81	207.34
500	HA	56.59	-	-	56.59	57.56
	SHA-U	56.59	0.00	32.74	23.85	35.49
	SHA-1	56.59	0.05	32.69	23.95	35.52
	SHA-2	56.59	0.49	31.10	25.98	34.88
	SHA-3	56.59	2.05	4.52	54.12	30.14
	SHA-1(5)	56.59	0.43	32.94	24.07	35.85
	SHA-2(5)	56.59	0.81	31.34	26.06	35.16
	SHA-3(5)	56.59	4.43	4.85	56.17	38.76

Table 3.6: Relative biases of the variance estimators, average mean lengths and coverages of 95% confidence interval for the smoothed Hájek estimators under Beaumont's population.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	-3.78	30.07	92.00	-3.04	30.35	92.03
	SHA-U	96.02	24.97	99.41	104.12	25.23	96.35
	SHA-1	62.11	28.59	98.69	102.93	25.24	96.38
	SHA-2	93.86	25.05	99.31	67.07	28.82	98.87
	SHA-3	102.70	28.63	99.46	111.34	28.92	99.40
	SHA-1(5)	34.52	28.09	97.34	93.46	25.26	96.43
	SHA-2(5)	83.49	25.08	99.11	37.63	28.42	97.89
	SHA-3(5)	59.10	28.81	98.45	69.18	29.10	98.71
500	HA	0.96	14.57	94.33	0.04	14.59	94.63
	SHA-U	154.48	12.46	99.76	151.52	12.46	80.88
	SHA-1	95.80	13.97	99.38	151.18	12.46	81.13
	SHA-2	152.38	12.49	99.77	95.00	13.97	99.48
	SHA-3	148.80	13.97	99.85	147.38	13.99	99.26
	SHA-1(5)	86.48	13.91	99.12	148.62	12.46	81.56
	SHA-2(5)	150.74	12.49	99.79	88.14	13.96	99.35
	SHA-3(5)	103.98	14.01	99.45	102.51	14.04	99.07

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HA	-9.82	60.37	86.13
	SHA-U	-58.97	33.65	0.00
	SHA-1	-58.23	33.98	0.00
	SHA-2	-51.91	36.13	0.00
	SHA-3	46.16	58.83	70.66
	SHA-1(5)	-58.21	34.74	0.00
	SHA-2(5)	-52.55	36.71	0.00
	SHA-3(5)	13.25	60.07	92.58
500	HA	-1.68	29.49	91.18
	SHA-U	-32.81	19.14	0.00
	SHA-1	-32.58	19.18	0.00
	SHA-2	-25.52	19.98	0.00
	SHA-3	79.59	28.84	6.39
	SHA-1(5)	-32.85	19.23	0.00
	SHA-2(5)	-25.90	20.01	0.00
	SHA-3(5)	44.92	29.38	82.19

and population distribution functions for pairs of vector random variables $(\boldsymbol{u}_i, \boldsymbol{v}_i)$:

$$f_p(\boldsymbol{u}_i|\boldsymbol{v}_i) = \frac{\mathrm{E}_s(w_i|\boldsymbol{u}_i, \boldsymbol{v}_i) f_s(\boldsymbol{u}_i|\boldsymbol{v}_i)}{\mathrm{E}_s(w_i|\boldsymbol{v}_i)},$$

where f_p, f_s are the distribution functions in the population and sample, respectively, and $\mathrm{E}_p, \mathrm{E}_s$ the corresponding expectations. Thus, applying this to $(\boldsymbol{u}_i, \boldsymbol{v}_i) = (x_i, y_i)$, we have

$$f_p(x_i|y_i) = \frac{\mathrm{E}_s(w_i|x_i, y_i) f_s(x_i|y_i)}{\mathrm{E}_s(w_i|y_i)},$$

where

$$\begin{aligned} f_s(x_i|y_i) &= \frac{1}{x_i\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x_i - \boldsymbol{B}_i\boldsymbol{\nu})^2}{2\sigma^2}\right\}, \\ \mathrm{E}_s(w_i|x_i, y_i) &= w_i = x_i + 1, \\ \mathrm{E}_s(w_i|y_i) &= 1 + \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right). \end{aligned}$$

From this, we immediately obtain

$$\begin{aligned} & f_p(x_i|y_i) \\ &= \left\{1 + \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right)\right\}^{-1} \left(1 + \frac{1}{x_i}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x_i - \boldsymbol{B}_i\boldsymbol{\nu})^2}{2\sigma^2}\right\} \\ &= \left\{1 + \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right)\right\}^{-1} \frac{1}{x_i\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x_i - \boldsymbol{B}_i\boldsymbol{\nu})^2}{2\sigma^2}\right\} \\ &\quad + \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right) \left\{1 + \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right)\right\}^{-1} \frac{1}{x_i\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x_i - \boldsymbol{B}_i\boldsymbol{\nu} - \sigma^2)^2}{2\sigma^2}\right\}, \end{aligned}$$

which is the probability density function of a mixture of two Log-Normal distributions $lN(\boldsymbol{B}_i\boldsymbol{\nu}, \sigma^2)$ and $lN(\boldsymbol{B}_i\boldsymbol{\nu} + \sigma^2, \sigma^2)$ with proportions $p = \left\{1 + \exp\left(\boldsymbol{B}_i\boldsymbol{\nu} + \frac{\sigma^2}{2}\right)\right\}^{-1}$ and $1 - p$, respectively.

We are interested in evaluating the effect of different model specifications on the behavior of the weight smoothing estimator. In order to create a population with multiple variables, some of which are in the weight smoothing model and others which are not, we generate four variables of interest from a multivariate normal distribution,

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N\left[\begin{pmatrix} 6 \\ 6 \\ 6 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0.2 & -0.5 \\ 0 & 1 & 0 & 0 \\ 0.2 & 0 & 1 & 0 \\ -0.5 & 0 & 0 & 1 \end{pmatrix}\right], \; i = 1, \cdots, N.$$

Then, we consider the following simple linear regression model under the sample

$$\log(w_i - 1) = \beta_0 + \beta_1 y_{1i} + \beta_2 y_{2i} + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2)$. Both variables y_1 and y_2 are part of the weight model, while y_3 and y_4 are correlated, weakly positively and strongly negatively, respectively, with y_1. Then, $x_i = w_i - 1 | y_{1i}, y_{2i} \sim lN(\beta_0 + \beta_1 y_{1i} + \beta_2 y_{2i}, \sigma^2)$ under the sample. Thus, the population x_i of 50,000 units is generated from a mixture of two Log-Normal distributions of $lN(\beta_0 + \beta_1 y_{1i} + \beta_2 y_{2i}, \sigma^2)$ and $lN(\beta_0 + \beta_1 y_{1i} + \beta_2 y_{2i} + \sigma^2, \sigma^2)$ with the proportion $p = \left\{1 + \exp\left(\beta_0 + \beta_1 y_{1i} + \beta_2 y_{2i} + \frac{\sigma^2}{2}\right)\right\}^{-1}$ and $1 - p$, respectively.

Following construction of these population variables and weights, independent samples of size $n = 100$ and $n = 500$ are generated under pps sampling with replacement, with the selection probabilities proportional to $1/w_i$. This results in samples in which the sample weight model is lognormal, as desired. For $n = 100$, the constants β_0, β_1, β_2, σ^2 are chosen to yield the population correlation coefficients $\rho_{w,y_1} = 0.136$, $\rho_{w,y_2} = -0.267$, $\rho_{w,y_3} = 0.026$ and $\rho_{w,y_4} = -0.066$. For $n = 500$, the constants β_0, β_1, β_2, σ^2 are chosen to yield the population correlation coefficients $\rho_{w,y_1} = 0.127$, $\rho_{w,y_2} = -0.221$, $\rho_{w,y_3} = 0.025$ and $\rho_{w,y_4} = -0.063$. The Monte Carlo sample sizes are all 10,000 in the simulation.

Five estimators are computed: Horvitz-Thompson (HT), and four different weight smoothing estimators, denoted SHT-U, SHT-1, SHT-2, SHT-12. The parameter estimates $\hat{\boldsymbol{\nu}}$ and $\hat{\sigma}^2$ in these last four estimators are obtained by fitting model (3.1) for different sets of covariates. The SHT-U estimator contains $\boldsymbol{B}_i = 1$ only, while the SHT-1 estimator adds $\boldsymbol{B}_i = y_{1i}$, the SHT-2 estimator adds $\boldsymbol{B}_i = y_{2i}$ and the SHT-12 estimator adds $\boldsymbol{B}_i = (y_{1i}, y_{2i})$. We compute the means and variances of five estimators for all y_1, y_2, y_3 and y_4 variables separately.

Table 3.7 presents the relative biases (RB) and relative efficiencies (RE) as in (3.24) and (3.25). The SHT-12 estimator is unbiased and efficient for all four variables, since the SHT-12 estimator is from the smoothed weight with $\boldsymbol{B}_i = (1, y_{1i}, y_{2i})$, so that SHT-12 holds under the true model. For variable y_1, the estimators SHT-1 and SHT-12 are both unbiased and more efficient than the

Table 3.7: Relative biases and relative efficiency of the smoothed Horvitz-Thompson estimators under correct model.

n	Estimators	Variable y_1		Variable y_2		Variable y_3		Variable y_4	
		RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)
100	HT	0.13	100.00	0.09	100.00	0.17	100.00	0.10	100.00
	SHT-U	-3.18	74.25	6.36	122.93	-0.56	77.36	1.39	84.92
	SHT-1	0.47	80.38	6.84	127.43	0.60	80.40	0.37	81.92
	SHT-2	-2.68	74.67	0.31	85.54	-0.13	78.53	1.82	86.70
	SHT-12	0.76	81.15	0.76	87.02	0.89	81.46	0.75	83.18
500	HT	-0.20	100.00	-0.17	100.00	-0.20	100.00	-0.14	100.00
	SHT-U	-4.71	75.70	8.69	183.83	-0.97	66.80	2.67	81.14
	SHT-1	0.15	65.84	8.87	187.75	0.17	68.30	0.40	69.43
	SHT-2	-4.61	75.66	-0.02	72.29	-0.75	67.21	2.78	81.86
	SHT-12	0.35	66.52	0.13	72.63	0.41	69.23	0.54	70.04

HT estimator. Since the smoothed weights from model depend on y_1, so that SHT-1 estimator is an unbiased and efficient estimator for the mean of y_1. As the sample size increases to $n = 500$, the estimators SHT-1 and SHT-12 are more efficient than the HT estimator, the SHT-U estimator and the SHT-2 estimator. Similarly, for variable y_2, the estimators SHT-2 and SHT-12 are both unbiased and more efficient than the HT estimator, the SHT-U estimator and the SHT-2 estimator. Hence, as long as the variable being estimated is included in the weight model, the resulting weight smoothing estimator remains unbiased and is efficient. However, if a variable is part of the true weight model but is not included in the model fitting, the resulting estimator is biased.

For variable y_3, the estimators SHT-U, SHT-1, SHT-2 and SHT-12 are approximately unbiased and more efficient than HT, due to the moderate correlation between w and y_3. For variable y_4, both estimators SHT-1 and SHT-12 are unbiased and efficient since y_4 is strongly correlated with y_1, so their relative biases and relative efficiencies have similar performance to y_1 as expected. Unlike for y_3, SHT-U is biased for y_4, because the stronger correlation with y_1 requires that covariate to be part of the weight smoothing model. We conclude from these results that the weight smoothing estimator appears to perform well when the correct variables are included as model predictors, both for these target variables and for other variables that are correlated with these model predictors.

Table 3.8 presents the simulated expected values of $\hat{\mathrm{V}}_1$, $\hat{\mathrm{V}}_2$, $\hat{\mathrm{V}}_3$ and $\hat{\mathrm{V}}$ as defined in (3.12), (3.13), (3.14) and (3.11), respectively. The true variances are also provided for comparison. The

Table 3.8: Expectations of estimated variance components for smoothed Horvitz-Thompson estimators, under correct model.

n	Estimators	Variable y_1					Variable y_2				
		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.
100	HT	0.948	-	-	0.948	0.950	0.710	-	-	0.710	0.713
	SHT-U	0.948	0.646	0.805	0.789	0.669	0.710	0.781	0.969	0.522	0.731
	SHT-1	0.948	0.728	0.917	0.758	0.763	0.710	0.780	0.970	0.519	0.740
	SHT-2	0.948	0.598	0.758	0.788	0.683	0.710	0.595	0.702	0.603	0.610
	SHT-12	0.948	0.663	0.850	0.761	0.769	0.710	0.591	0.700	0.601	0.618
500	HT	0.407	-	-	0.407	0.411	0.297	-	-	0.297	0.299
	SHT-U	0.407	0.225	0.348	0.285	0.231	0.297	0.293	0.450	0.140	0.278
	SHT-1	0.407	0.259	0.414	0.252	0.270	0.297	0.288	0.445	0.139	0.278
	SHT-2	0.407	0.205	0.327	0.286	0.234	0.297	0.209	0.298	0.207	0.216
	SHT-12	0.407	0.233	0.388	0.252	0.273	0.297	0.203	0.293	0.207	0.217
n	Estimators	Variable y_3					Variable y_4				
		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.
100	HT	0.882	-	-	0.882	0.892	0.827	-	-	0.827	0.834
	SHT-U	0.882	0.681	0.848	0.715	0.689	0.827	0.709	0.882	0.654	0.701
	SHT-1	0.882	0.696	0.873	0.705	0.716	0.827	0.678	0.833	0.672	0.683
	SHT-2	0.882	0.629	0.795	0.715	0.700	0.827	0.655	0.828	0.654	0.711
	SHT-12	0.882	0.636	0.810	0.707	0.724	0.827	0.621	0.777	0.671	0.692
500	HT	0.367	-	-	0.367	0.370	0.343	-	-	0.343	0.348
	SHT-U	0.367	0.244	0.375	0.236	0.244	0.343	0.262	0.403	0.203	0.257
	SHT-1	0.367	0.247	0.386	0.228	0.252	0.343	0.240	0.362	0.221	0.241
	SHT-2	0.367	0.223	0.355	0.236	0.247	0.343	0.239	0.378	0.204	0.257
	SHT-12	0.367	0.223	0.362	0.229	0.255	0.343	0.217	0.338	0.221	0.243

Table 3.9: Relative biases of the variance estimators, average mean lengths and coverages of 95% confidence interval for the smoothed Horvitz-Thompson estimators, under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-0.23	3.82	90.57	-0.48	3.30	91.45
	SHT-U	17.93	3.48	96.77	-28.56	2.83	87.71
	SHT-1	-0.59	3.41	95.23	-29.86	2.82	87.02
	SHT-2	15.28	3.48	96.68	-1.02	3.05	95.24
	SHT-12	-1.07	3.42	95.04	-2.77	3.04	94.92
500	HT	-0.80	2.50	91.14	-0.78	2.14	91.83
	SHT-U	23.40	2.09	94.82	-49.54	1.47	66.84
	SHT-1	-6.78	1.97	94.02	-49.93	1.46	66.15
	SHT-2	21.98	2.09	94.85	-4.14	1.78	94.63
	SHT-12	-7.48	1.97	94.03	-4.85	1.78	94.44
n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-1.15	3.68	90.77	-0.89	3.56	90.71
	SHT-U	3.83	3.32	95.77	-6.71	3.17	94.43
	SHT-1	-1.54	3.29	95.05	-1.66	3.21	95.14
	SHT-2	2.09	3.31	95.56	-8.10	3.17	94.04
	SHT-12	-2.28	3.30	94.94	-2.96	3.21	94.87
500	HT	-0.65	2.38	91.76	-1.39	2.30	91.97
	SHT-U	-3.18	1.90	94.78	-21.11	1.76	90.37
	SHT-1	-9.71	1.87	93.76	-8.42	1.84	93.87
	SHT-2	-4.38	1.90	94.66	-20.79	1.77	90.30
	SHT-12	-10.52	1.87	93.62	-8.77	1.84	93.76

expected estimated variances $\hat{\mathrm{V}}$ are close to the true variances when the correct variables are included as model predictors. As shown in Table 3.7, the variance component $\hat{\mathrm{V}}_2$ is always smaller on average than $\hat{\mathrm{V}}_3$, leading to a reduction in the estimated overall variance.

Table 3.9 presents the relative biases of the variance estimators as percentages, and the average mean lengths and coverages of nominal 95% confidence interval estimators. All four smoothed estimators have slight advantages in the mean length compare to the mean length of the Horvitz-Thompson estimator. The SHT-12 estimators show good coverages, which is close to 95%. This is what we expected since the smoothed weights of SHT-12 are from the underlying model, which should fit the data well for all the variables. Similar to the results shown in Table 3.7, SHT-1 variance estimators have small relative biases and the coverages are all close to 95% for the

variables y_1, y_3 and y_4 since both y_3 and y_4 correlate with y_1. Instead, SHT-2 variance estimators have small relative biases and the coverages are all close to 95% for the variable y_2.

Table 3.10: Relative biases and relative efficiency of the smoothed Hájek estimators, under correct model.

n	Estimators	Variable y_1		Variable y_2		Variable y_3		Variable y_4	
		RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)
100	HA	-0.10	100.00	0.17	100.00	0.01	100.00	0.01	100.00
	SHA-U	-3.14	147.71	6.51	482.96	-0.49	35.74	1.47	57.46
	SHA-1	-0.06	69.84	6.41	471.96	0.12	35.69	-0.07	43.21
	SHA-2	-3.06	147.40	0.10	70.03	-0.49	40.86	1.48	63.60
	SHA-12	-0.09	70.37	0.12	70.22	0.10	40.58	0.00	47.58
500	HA	-0.08	100.00	0.09	100.00	-0.03	100.00	0.06	100.00
	SHA-U	-4.90	789.36	8.50	2005.53	-1.16	63.70	2.48	220.48
	SHA-1	-0.21	51.78	8.52	2012.99	-0.18	22.79	0.07	27.93
	SHA-2	-4.93	804.75	-0.29	50.34	-1.07	63.55	2.45	221.51
	SHA-12	-0.17	53.36	-0.32	51.82	-0.09	28.68	0.06	32.66

The Hájek type smoothed estimators are also considered under correct model. Table 3.10 presents the relative biases and relative efficiency of the smoothed Hájek estimators under correct model. For all variables, the resulting weight smoothing estimators are unbiased and efficient as long as the variable being estimated is included in the weight model. Table 3.11 presents the simulated expected variance components for smoothed Hájek estimators under correct model. The variance components for smoothed Hájek estimators are all small compared to smoothed Horvitz-Thompson estimators. Table 3.12 represents the relative biases of the variance estimators as percentages, and the average mean lengths and coverages of 95% confidence interval for the smoothed Hájek estimators under correct model. For $n = 100$, the SHA-12 estimators show good coverages, which is close to 95%. Similar to the results shown in Table 3.10, the SHA-1 estimators show good coverages for the variable y_1, y_3 and y_4, while the SHA-2 estimators show good coverages for the variable y_2 as expected. However, as n increases to 500, the coverages for all smoothed estimators degrade and are below 95%, even for the SHA-12 estimators. This suggests that model-based variance estimator might not be appropriate for the Hájek-type estimators, but further study is warranted.

Table 3.11: Expectations of estimated variance components for smoothed Hájek estimators, under correct model.

n	Estimators	Variable y_1					Variable y_2				
		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.
100	HA	0.030	0.000	0.000	0.030	0.031	0.032	0.000	0.000	0.032	0.034
	SHA-U	0.030	0.000	0.021	0.008	0.010	0.032	0.000	0.022	0.010	0.010
	SHA-1	0.030	0.011	0.022	0.019	0.021	0.032	0.000	0.022	0.010	0.011
	SHA-2	0.030	0.000	0.020	0.010	0.012	0.032	0.010	0.023	0.019	0.024
	SHA-12	0.030	0.009	0.020	0.019	0.022	0.032	0.010	0.023	0.019	0.024
500	HA	0.011	0.000	0.000	0.011	0.011	0.013	0.000	0.000	0.013	0.013
	SHA-U	0.011	0.000	0.010	0.001	0.002	0.013	0.000	0.010	0.002	0.002
	SHA-1	0.011	0.003	0.011	0.004	0.006	0.013	0.000	0.010	0.002	0.002
	SHA-2	0.011	0.000	0.009	0.002	0.003	0.013	0.003	0.012	0.004	0.006
	SHA-12	0.011	0.003	0.010	0.004	0.006	0.013	0.003	0.012	0.004	0.006

n	Estimators	Variable y_3					Variable y_4				
		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}$	Var.
100	HA	0.029	0.000	0.000	0.029	0.030	0.029	0.000	0.000	0.029	0.031
	SHA-U	0.029	0.000	0.021	0.008	0.010	0.029	0.000	0.022	0.008	0.010
	SHA-1	0.029	0.001	0.021	0.009	0.011	0.029	0.003	0.022	0.010	0.013
	SHA-2	0.029	0.000	0.020	0.010	0.012	0.029	0.000	0.020	0.009	0.012
	SHA-12	0.029	0.001	0.019	0.011	0.012	0.029	0.003	0.020	0.012	0.015
500	HA	0.011	0.000	0.000	0.011	0.011	0.011	0.000	0.000	0.011	0.011
	SHA-U	0.011	0.000	0.010	0.001	0.002	0.011	0.000	0.010	0.001	0.002
	SHA-1	0.011	0.000	0.010	0.001	0.002	0.011	0.001	0.010	0.002	0.003
	SHA-2	0.011	0.000	0.009	0.001	0.003	0.011	0.000	0.009	0.002	0.003
	SHA-12	0.011	0.000	0.009	0.001	0.003	0.011	0.001	0.009	0.002	0.004

Table 3.12: Relative biases of the variance estimators, average mean lengths and coverages of 95% confidence interval for the smoothed Hájek estimators, under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	-2.84	0.676	91.32	-5.00	0.701	89.17
	SHA-U	-14.49	0.359	47.15	-0.02	0.395	2.81
	SHA-1	-13.43	0.533	93.38	-1.83	0.401	3.71
	SHA-2	-11.26	0.397	56.13	-20.52	0.537	92.01
	SHA-12	-11.35	0.542	93.69	-19.75	0.540	92.02
500	HA	-0.78	0.413	92.19	-2.18	0.443	90.95
	SHA-U	-47.97	0.126	0.00	18.71	0.192	0.00
	SHA-1	-31.58	0.244	89.07	12.39	0.196	0.00
	SHA-2	-36.67	0.161	0.01	-42.52	0.235	85.61
	SHA-12	-32.06	0.248	89.31	-41.03	0.241	85.94
n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	-3.38	0.670	91.68	-5.63	0.669	91.57
	SHA-U	-21.66	0.346	90.53	-24.10	0.340	79.09
	SHA-1	-20.17	0.363	91.87	-22.53	0.398	91.54
	SHA-2	-16.56	0.384	91.58	-19.86	0.380	81.97
	SHA-12	-14.21	0.402	92.98	-18.51	0.429	92.35
500	HA	-1.28	0.403	93.75	-1.49	0.408	93.19
	SHA-U	-68.46	0.098	31.96	-67.19	0.102	1.55
	SHA-1	-60.83	0.118	76.68	-47.85	0.156	84.08
	SHA-2	-57.14	0.133	51.23	-40.07	0.155	8.84
	SHA-12	-51.76	0.150	82.46	-35.88	0.188	88.48

3.8 Replicate Variance Estimation

3.8.1 Proposed estimator

As discussed in the previous section, the weight smoothing estimator appears to be reasonably robust to model misspecification when it is implemented as a Hájek-type estimator. However, even that modification was ineffective in producing reliable model-based variance estimators. Here, we investigate jackknife variance estimation as a more robust and practical alternative. We consider the widely used delete-a-group jackknife (DAGJK), as described in Kott [2001].

Let R denote the number of variance replication groups. The sample is divided into R random groups of size m, and we assume $n = mR$ for simplicity. We investigate two procedures to perform DAGJK variance estimation, depending on whether the weights are smoothed before or after the creation of the replicates. The first procedure, which we will denote by $JK-A$, creates weights as follows: for $r = 1, \cdots, R$, delete the rth group, fit weight model (3.1) on the remaining observations to obtain replicate model predictions $\hat{w}_{r,i}$, and create replicate weights $\hat{w}_i^{(r)} = \hat{w}_{r,i}/(1-R^{-1})$. Then, the rth replicate of the smoothed Horvitz-Thompson estimator is

$$\hat{Y}^{HT(r)} = \frac{1}{N} \sum_{i \in S^{(r)}} \hat{w}_i^{(r)} y_i, \tag{3.26}$$

with $S^{(r)}$ denoting the sample remaining after remove the rth group. The jackknife variance estimator is then defined as either

$$\hat{V}_{JK1-A} = \frac{R-1}{R} \times \sum_{r=1}^{R} \left(\hat{Y}^{HT(r)} - \hat{Y}^{HT} \right)^2, \tag{3.27}$$

or

$$\hat{V}_{JK2-A} = \frac{R-1}{R} \times \sum_{r=1}^{R} \left(\hat{Y}^{HT(r)} - \bar{Y}^{R} \right)^2, \tag{3.28}$$

with $\bar{Y}^R = \sum_{r=1}^{R} \hat{Y}^{HT(r)}/R$. For the second procedure, denoted $JK-B$, the smoothed weights $\hat{w}_i$ are not recomputed in each replicate. Instead, group r is removed and the replicate weights are $\hat{w}_i^{(r)} = \hat{w}_i/(1-R^{-1})$. Two jackknife variance estimators, $JK1-B$ and $JK2-B$, are defined completely analogously as above. All these replication variants can readily be applied to the Hájek-type estimators, by using the suitably modified replication weights in the Hájek estimators.

3.8.2 Beamont's population

We evaluate the performance of JKDAG variance estimators for the smoothed Horvitz-Thompson estimators under the setup from Beaumont [2008] as described in Section 7.1, and we set $R = 20$. Table 3.13 through Table 3.16 represent the relative biases of the jackknife variance estimators, average mean lengths and coverages of 95% confidence interval estimators for $JK1 - A$, $JK1 - B$, $JK1 - B$ and $JK2 - B$, respectively. For $JK - A$, the jackknife variance estimators for SHT-1, SHT-2 and SHT-3 have moderate relative biases, with the absolute values less than 5%. However, the smoothed estimators have poor coverages, due to the bias of the estimators themselves shown in Table 3.1. For $JK - B$, all the smoothed estimators perform badly.

We implement the JKDAG variance estimator for the Hájek-type weight smoothing estimators. Table 3.17 through Table 3.20 present the relative biases of the $JK1 - A$, $JK2 - A$, $JK1 - B$ and $JK2 - B$ variance estimators, average mean lengths and coverages of 95% confidence interval for the Hájek estimators, respectively. The $JK1$ and $JK2$ continue to have similar results, and $JK - B$ fails to result in reasonable variance estimation across all cases. In contrast, the $JK - A$ variance estimators have modest bias and coverage for the weight smoothing estimators that is at least as good as for the unsmoothed estimator, when the estimator itself is appropriate. Specifically, for y_1, which is uncorrelated with the weights, the weight smoothing estimator is approximately unbiased regardless of which survey variable is used as a covariate in the weight model, and the $JK - A$ variance estimators result in confidence intervals that are narrower than for the Hájek estimator and have close to nominal coverage. For y_2, only the weight smoothing estimators that use y_2 as covariate are unbiased, and $JK - A$ leads to narrow intervals with good coverage. When weight smoothing is done without y_2 as a covariate, the bias in the estimator leads to significant undercoverage. For y_3, poor performance of the estimator continues to lead to unacceptable inference.

3.8.3 Simulation under correct model

We evaluate the JKDAG variance estimators under the correct weight model distribution. The set-up is identical to that in Section 7.2. Table 3.21 through 3.24 present the relative biases of

Table 3.13: Relative biases of the $JK1 - A$ variance estimation for the smoothed Horvitz-Thompson estimators, average mean lengths and coverages of 95% confidence interval estimators under Beaumont's model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	3.17	39.19	90.73	1.75	39.09	89.53
	SHT-U	2.19	20.74	87.21	0.51	22.42	93.75
	SHT-1	4.40	24.84	88.62	0.90	22.60	93.86
	SHT-2	2.67	21.01	87.95	0.04	24.69	86.47
	SHT-3	4.07	21.16	83.35	0.27	21.61	85.66
	SHT-1(5)	10.60	27.91	89.71	5.50	23.95	94.52
	SHT-2(5)	7.76	22.43	89.59	7.86	27.64	88.24
	SHT-3(5)	10.37	26.14	88.45	4.84	26.33	89.31
500	HT	-0.45	17.36	91.93	2.91	17.49	92.19
	SHT-U	0.18	9.15	77.15	1.65	9.91	92.82
	SHT-1	-0.18	10.89	79.89	1.81	9.92	92.81
	SHT-2	0.00	9.20	78.10	2.91	10.87	77.66
	SHT-3	-0.10	9.30	61.55	2.36	9.53	73.29
	SHT-1(5)	0.99	11.14	81.01	2.36	10.01	92.76
	SHT-2(5)	0.75	9.30	79.24	3.83	11.10	78.57
	SHT-3(5)	0.74	10.81	80.12	3.36	11.02	86.05

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HT	1.05	38.64	90.08
	SHT-U	2.31	72.89	0.01
	SHT-1	2.85	73.22	0.01
	SHT-2	2.10	72.11	0.01
	SHT-3	1.85	30.09	93.14
	SHT-1(5)	5.40	74.80	0.02
	SHT-2(5)	4.47	73.64	0.06
	SHT-3(5)	2.24	28.73	89.35
500	HT	-3.34	17.22	91.60
	SHT-U	0.46	32.13	0.00
	SHT-1	0.60	32.16	0.00
	SHT-2	0.80	31.66	0.00
	SHT-3	-2.28	13.21	67.26
	SHT-1(5)	0.88	32.29	0.00
	SHT-2(5)	1.10	31.78	0.00
	SHT-3(5)	-3.98	12.26	84.90

Table 3.14: Relative biases of the $JK2 - A$ variance estimation for the smoothed Horvitz-Thompson estimators, average mean lengths and coverages of 95% confidence interval estimators under Beaumont's model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	3.17	39.19	90.73	1.75	39.09	89.53
	SHT-U	2.18	20.74	87.21	0.50	22.42	93.75
	SHT-1	4.38	24.84	88.62	0.88	22.60	93.86
	SHT-2	2.64	21.01	87.95	0.02	24.69	86.47
	SHT-3	4.06	21.16	83.34	0.26	21.61	85.66
	SHT-1(5)	10.21	27.86	89.69	5.25	23.92	94.52
	SHT-2(5)	7.44	22.39	89.55	7.39	27.58	88.21
	SHT-3(5)	10.16	26.12	88.43	4.63	26.30	89.31
500	HT	-0.45	17.36	91.93	2.91	17.49	92.19
	SHT-U	0.18	9.15	77.15	1.65	9.91	92.82
	SHT-1	-0.19	10.89	79.89	1.80	9.92	92.80
	SHT-2	0.00	9.20	78.10	2.91	10.87	77.66
	SHT-3	-0.10	9.30	61.55	2.36	9.53	73.29
	SHT-1(5)	0.94	11.13	81.00	2.33	10.01	92.75
	SHT-2(5)	0.71	9.30	79.24	3.78	11.10	78.57
	SHT-3(5)	0.72	10.80	80.12	3.34	11.02	86.05

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HT	1.05	38.64	90.08
	SHT-U	2.30	72.89	0.01
	SHT-1	2.84	73.22	0.01
	SHT-2	2.09	72.10	0.01
	SHT-3	1.84	30.09	93.13
	SHT-1(5)	5.29	74.76	0.02
	SHT-2(5)	4.37	73.60	0.05
	SHT-3(5)	2.17	28.72	89.34
500	HT	-3.34	17.22	91.60
	SHT-U	0.46	32.13	0.00
	SHT-1	0.60	32.16	0.00
	SHT-2	0.80	31.66	0.00
	SHT-3	-2.28	13.21	67.26
	SHT-1(5)	0.87	32.29	0.00
	SHT-2(5)	1.09	31.78	0.00
	SHT-3(5)	-4.00	12.26	84.90

Table 3.15: Relative biases of the $JK1 - B$ variance estimation for the smoothed Horvitz-Thompson variance estimators, average mean lengths and coverages of 95% confidence interval estimators under Beaumont's model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	3.17	39.19	90.73	1.75	39.09	89.53
	SHT-U	-36.86	16.30	80.07	-45.88	16.45	83.35
	SHT-1	-53.63	16.56	74.98	-45.69	16.58	83.51
	SHT-2	-36.27	16.55	80.88	-59.59	15.69	70.60
	SHT-3	-19.68	18.59	79.92	-25.69	18.60	82.07
	SHT-1(5)	-40.81	20.42	78.63	-42.94	17.61	84.09
	SHT-2(5)	-34.17	17.53	82.90	-49.16	18.98	74.44
	SHT-3(5)	-10.63	23.52	86.46	-16.11	23.55	87.07
500	HT	-0.45	17.36	91.93	2.91	17.49	92.19
	SHT-U	-37.81	7.21	66.05	-44.93	7.29	81.13
	SHT-1	-56.09	7.22	61.71	-44.83	7.31	81.33
	SHT-2	-37.48	7.27	67.24	-59.03	6.86	55.28
	SHT-3	-22.45	8.20	54.28	-23.73	8.23	66.27
	SHT-1(5)	-53.53	7.55	63.85	-44.53	7.37	81.07
	SHT-2(5)	-37.11	7.35	68.61	-57.01	7.15	57.04
	SHT-3(5)	-16.47	9.84	76.56	-16.37	9.91	82.35

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HT	1.05	38.64	90.08
	SHT-U	-53.88	48.94	0.00
	SHT-1	-54.17	48.88	0.00
	SHT-2	-54.69	48.03	0.00
	SHT-3	-77.09	14.27	55.87
	SHT-1(5)	-55.36	48.68	0.01
	SHT-2(5)	-55.77	47.91	0.01
	SHT-3(5)	-75.59	14.04	53.40
500	HT	-3.34	17.22	91.60
	SHT-U	-54.20	21.69	0.00
	SHT-1	-54.24	21.69	0.00
	SHT-2	-54.30	21.32	0.00
	SHT-3	-78.05	6.26	26.50
	SHT-1(5)	-54.53	21.68	0.00
	SHT-2(5)	-54.59	21.30	0.00
	SHT-3(5)	-79.44	5.67	48.61

Table 3.16: Relative biases of the $JK2 - B$ variance estimation for the smoothed Horvitz-Thompson esitmators, average mean lengths and coverages of 95% confidence interval estimators under Beaumont's model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	3.17	39.19	90.73	1.75	39.09	89.53
	SHT-U	-36.86	16.30	80.07	-45.88	16.45	83.35
	SHT-1	-53.63	16.56	74.98	-45.69	16.58	83.51
	SHT-2	-36.27	16.55	80.88	-59.59	15.69	70.60
	SHT-3	-19.68	18.59	79.92	-25.69	18.60	82.07
	SHT-1(5)	-40.81	20.42	78.63	-42.94	17.61	84.09
	SHT-2(5)	-34.17	17.53	82.90	-49.16	18.98	74.44
	SHT-3(5)	-10.63	23.52	86.46	-16.11	23.55	87.07
500	HT	-0.45	17.36	91.93	2.91	17.49	92.19
	SHT-U	-37.81	7.21	66.05	-44.93	7.29	81.13
	SHT-1	-56.09	7.22	61.71	-44.83	7.31	81.33
	SHT-2	-37.48	7.27	67.24	-59.03	6.86	55.28
	SHT-3	-22.45	8.20	54.28	-23.73	8.23	66.27
	SHT-1(5)	-53.53	7.55	63.85	-44.53	7.37	81.07
	SHT-2(5)	-37.11	7.35	68.61	-57.01	7.15	57.04
	SHT-3(5)	-16.47	9.84	76.56	-16.37	9.91	82.35

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HT	1.05	38.64	90.08
	SHT-U	-53.88	48.94	0.00
	SHT-1	-54.17	48.88	0.00
	SHT-2	-54.69	48.03	0.00
	SHT-3	-77.09	14.27	55.87
	SHT-1(5)	-55.36	48.68	0.01
	SHT-2(5)	-55.77	47.91	0.01
	SHT-3(5)	-75.59	14.04	53.40
500	HT	-3.34	17.22	91.60
	SHT-U	-54.20	21.69	0.00
	SHT-1	-54.24	21.69	0.00
	SHT-2	-54.30	21.32	0.00
	SHT-3	-78.05	6.26	26.50
	SHT-1(5)	-54.53	21.68	0.00
	SHT-2(5)	-54.59	21.30	0.00
	SHT-3(5)	-79.44	5.67	48.61

Table 3.17: Relative biases of the $JK1 - A$ variance estimation for the smoothed Hájek estimators, average mean lengths and coverages of 95% confidence interval estimators under Beaumont's model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.22	30.99	92.67	3.56	31.36	92.79
	SHA-U	-2.83	17.58	92.98	0.79	17.73	81.21
	SHA-1	0.56	22.51	93.34	1.62	17.86	81.49
	SHA-2	-1.97	17.81	93.10	3.05	22.64	93.85
	SHA-3	-0.35	20.07	93.17	2.75	20.17	92.35
	SHA-1(5)	5.29	24.85	93.47	7.07	18.79	83.77
	SHA-2(5)	2.39	18.74	93.71	7.89	25.16	94.00
	SHA-3(5)	4.37	23.33	93.32	8.31	23.29	93.04
500	HA	2.91	14.71	93.50	1.26	14.68	93.81
	SHA-U	0.98	7.85	92.65	1.77	7.93	39.58
	SHA-1	1.43	10.06	93.58	1.99	7.94	39.82
	SHA-2	1.06	7.91	92.83	2.10	10.11	93.76
	SHA-3	1.79	8.93	93.59	1.99	8.98	88.45
	SHA-1(5)	3.04	10.34	93.66	3.07	8.02	41.27
	SHA-2(5)	2.00	7.97	92.92	3.46	10.35	93.94
	SHA-3(5)	3.15	9.96	93.72	2.45	9.98	91.44

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HA	1.16	63.94	88.89
	SHA-U	-1.21	52.22	0.00
	SHA-1	-0.16	52.54	0.00
	SHA-2	0.18	52.14	0.00
	SHA-3	-0.44	48.55	53.43
	SHA-1(5)	8.24	55.91	0.36
	SHA-2(5)	8.86	55.60	0.45
	SHA-3(5)	1.21	56.79	83.41
500	HA	1.11	29.90	91.43
	SHA-U	0.18	23.38	0.00
	SHA-1	0.24	23.39	0.00
	SHA-2	0.45	23.20	0.00
	SHA-3	1.49	21.68	1.77
	SHA-1(5)	1.78	23.68	0.00
	SHA-2(5)	1.65	23.44	0.00
	SHA-3(5)	1.87	24.63	65.20

Table 3.18: Relative biases of the $JK2 - A$ variance estimation for the smoothed Hájek estimators, average mean lengths and coverages of 95% confidence interval estimators under Beaumont's model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.14	30.98	92.67	3.46	31.35	92.79
	SHA-U	-2.83	17.58	92.98	0.79	17.73	81.21
	SHA-1	0.56	22.51	93.34	1.62	17.86	81.49
	SHA-2	-1.97	17.81	93.10	3.05	22.64	93.85
	SHA-3	-0.36	20.07	93.17	2.75	20.17	92.35
	SHA-1(5)	5.09	24.83	93.46	6.86	18.78	83.70
	SHA-2(5)	2.20	18.72	93.71	7.66	25.14	94.00
	SHA-3(5)	4.23	23.32	93.32	8.16	23.27	93.02
500	HA	2.90	14.71	93.50	1.25	14.68	93.81
	SHA-U	0.98	7.85	92.65	1.77	7.93	39.58
	SHA-1	1.43	10.06	93.58	1.99	7.94	39.82
	SHA-2	1.06	7.91	92.83	2.10	10.11	93.76
	SHA-3	1.79	8.93	93.59	1.99	8.98	88.45
	SHA-1(5)	2.98	10.34	93.66	3.03	8.02	41.26
	SHA-2(5)	1.98	7.97	92.92	3.41	10.35	93.94
	SHA-3(5)	3.13	9.96	93.71	2.43	9.98	91.42

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HA	0.98	63.89	88.88
	SHA-U	-1.21	52.22	0.00
	SHA-1	-0.18	52.54	0.00
	SHA-2	0.16	52.14	0.00
	SHA-3	-0.45	48.55	53.43
	SHA-1(5)	7.31	55.67	0.36
	SHA-2(5)	7.95	55.37	0.43
	SHA-3(5)	1.16	56.77	83.41
500	HA	1.08	29.90	91.43
	SHA-U	0.18	23.38	0.00
	SHA-1	0.23	23.39	0.00
	SHA-2	0.44	23.20	0.00
	SHA-3	1.49	21.68	1.77
	SHA-1(5)	1.60	23.66	0.00
	SHA-2(5)	1.50	23.42	0.00
	SHA-3(5)	1.86	24.63	65.20

Table 3.19: Relative biases of the $JK1 - B$ variance estimation for the smoothed Hájek estimators, average mean lengths and coverages of 95% confidence interval estimators under Beaumont's model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.22	30.99	92.67	3.56	31.36	92.79
	SHA-U	-2.83	17.58	92.98	0.79	17.73	81.21
	SHA-1	37.92	17.69	86.46	0.86	17.79	81.30
	SHA-2	-2.68	17.75	93.07	-34.79	18.01	87.42
	SHA-3	-0.48	20.06	93.16	2.19	20.11	92.34
	SHA-1(5)	26.52	20.76	88.06	1.23	18.27	82.80
	SHA-2(5)	-3.22	18.22	93.11	-23.82	21.15	88.16
	SHA-3(5)	0.40	22.88	93.11	3.96	22.82	92.67
500	HA	2.91	14.71	93.50	1.26	14.68	93.81
	SHA-U	0.98	7.85	92.65	1.77	7.93	39.58
	SHA-1	-37.98	7.86	85.75	1.83	7.93	39.79
	SHA-2	0.91	7.90	92.78	-35.88	8.01	86.83
	SHA-3	1.76	8.93	93.60	1.57	8.96	88.41
	SHA-1(5)	-35.27	8.20	86.12	1.97	7.98	41.04
	SHA-2(5)	0.98	7.93	92.72	-34.30	8.25	87.05
	SHA-3(5)	2.13	9.91	93.61	1.13	9.92	91.27

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HA	1.16	63.94	88.89
	SHA-U	-1.21	52.22	0.00
	SHA-1	-1.62	52.16	0.00
	SHA-2	-2.23	51.51	0.00
	SHA-3	-30.63	40.53	41.85
	SHA-1(5)	-0.90	53.50	0.31
	SHA-2(5)	-0.77	53.08	0.36
	SHA-3(5)	-19.14	50.76	79.72
500	HA	1.11	29.90	91.43
	SHA-U	0.18	23.38	0.00
	SHA-1	-0.02	23.36	0.00
	SHA-2	-0.89	23.05	0.00
	SHA-3	-29.62	18.05	0.70
	SHA-1(5)	-0.19	23.45	0.00
	SHA-2(5)	-0.96	23.13	0.00
	SHA-3(5)	-23.72	21.31	57.41

Table 3.20: Relative biases of the $JK2 - B$ variance estimation for the smoothed Hájek estimators, average mean lengths and coverages of 95% confidence interval estimators under Beaumont's model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.14	30.98	92.67	3.46	31.35	92.79
	SHA-U	-2.83	17.58	92.98	0.79	17.73	81.21
	SHA-1	-37.92	17.69	86.46	0.86	17.79	81.30
	SHA-2	-2.68	17.75	93.07	-34.79	18.01	87.42
	SHA-3	-0.48	20.06	93.15	2.19	20.11	92.34
	SHA-1(5)	-26.53	20.76	88.06	1.22	18.27	82.80
	SHA-2(5)	-3.22	18.22	93.11	-23.84	21.14	88.16
	SHA-3(5)	0.39	22.88	93.11	3.95	22.81	92.67
500	HA	2.90	14.71	93.50	1.25	14.68	93.81
	SHA-U	0.98	7.85	92.65	1.77	7.93	39.58
	SHA-1	-37.98	7.86	85.75	1.83	7.93	39.79
	SHA-2	0.91	7.90	92.78	-35.88	8.01	86.83
	SHA-3	1.76	8.93	93.60	1.57	8.96	88.41
	SHA-1(5)	-35.27	8.20	86.12	1.97	7.98	41.04
	SHA-2(5)	0.98	7.93	92.72	-34.30	8.25	87.05
	SHA-3(5)	2.13	9.91	93.61	1.13	9.92	91.27

n	Estimators	Variable y_3		
		RB (%)	AL	CR (%)
100	HA	0.98	63.89	88.88
	SHA-U	-1.21	52.22	0.00
	SHA-1	-1.62	52.16	0.00
	SHA-2	-2.23	51.51	0.00
	SHA-3	-30.64	40.52	41.85
	SHA-1(5)	-0.91	53.50	0.31
	SHA-2(5)	-0.78	53.08	0.36
	SHA-3(5)	-19.19	50.74	79.72
500	HA	1.08	29.90	91.43
	SHA-U	0.18	23.38	0.00
	SHA-1	-0.02	23.36	0.00
	SHA-2	-0.89	23.05	0.00
	SHA-3	-29.63	18.05	0.70
	SHA-1(5)	-0.19	23.45	0.00
	SHA-2(5)	-0.96	23.13	0.00
	SHA-3(5)	-23.72	21.31	57.41

the jackknife variance estimators, average mean lengths and coverages of 95% confidence interval estimators for $JK1-A, JK2-A, JK1-B$ and $JK2-B$, respectively. As expected, we see that the $JK1$ and $JK2$ have very similar results, so that the choice between these two versions can be left to the discretion of the user.

For $JK-A$, the results in Tables 3.21 and 3.22 exhibit a modest negative bias for the Horvitz-Thompson estimators for all sample sizes and survey variables, and results in undercoverage. Interestingly, both the bias and the undercoverage are improved for the weight smoothing estimators when the weight models include the relevant survey variables, while still resulting in shorter confidence intervals relative to the Horvitz-Thompson estimator. More specifically, SHT-12 has the best coverages for all four variables, SHT-1 has good coverages for variable y_1 and SHT-2 has good coverages for variable y_2. Also, the SHT-U, SHT-1 and SHT-2 estimators have good coverages for variable y_3 and variable y_4 since both y_3 and y_4 have weak relationship with the weight although they are correlate with y_1. In contrast, as shown in Tables 3.23 and 3.24, $JK-B$ performed very poorly, so that this method cannot be recommended even for correctly specified models.

Finally, we implement the JKDAG variance estimator for the Hájek-type weight smoothing estimators under correct model. Table 3.25 through Table 3.28 present the relative biases of the $JK1-A$, $JK2-A$, $JK1-B$ and $JK2-B$ variance estimators, average mean lengths and coverages of 95% confidence interval for the Hájek estimators, respectively. Again, the $JK1$ and $JK2$ have similar results, and $JK-B$ performed poorly in general, although there are several smoothed Hájek scenarios in which they happen to have coverages close to 95% for variable y_3.

For $JK-A$ variance estimators, the biases of the variance estimation are modest for all variables. The coverages are close to 95% as long as the weight smoothing estimator itself is appropriate. Specifically, SHA-1 and SHA-12 show good coverages for y_1, y_3 and y_4, while SHA-2 and SHA-12 show good coverages for y_2.

3.9 Conclusions

We have extended the theoretical investigations of the weight smoothing estimator proposed in Beaumont [2008], focusing on a lognormal linear model for the weights that was also suggested by

Table 3.21: Relative biases of the $JK1 - A$ variance estimation for the smoothed Horvitz-Thompson estimators, average mean lengths and coverages of 95% confidence interval estimators under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-0.59	3.81	89.67	-0.72	3.30	90.48
	SHT-U	1.20	3.23	89.08	1.30	3.37	94.00
	SHT-1	1.48	3.45	92.34	1.51	3.40	94.03
	SHT-2	1.42	3.26	89.58	1.16	3.08	92.34
	SHT-12	1.59	3.46	92.53	1.23	3.10	92.72
500	HT	-1.18	2.50	90.41	-1.20	2.13	90.89
	SHT-U	-0.18	1.88	85.45	-0.47	2.06	85.33
	SHT-1	-0.07	2.04	93.15	-0.43	2.06	84.97
	SHT-2	-0.16	1.89	85.82	-0.45	1.82	92.84
	SHT-12	0.08	2.05	93.25	-0.39	1.82	92.89
n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-1.48	3.67	89.80	-1.25	3.56	89.80
	SHT-U	0.85	3.27	91.52	1.00	3.30	92.77
	SHT-1	0.99	3.33	92.36	1.01	3.26	92.05
	SHT-2	0.94	3.30	91.89	1.02	3.32	92.98
	SHT-12	1.02	3.35	92.59	1.01	3.28	92.43
500	HT	-0.95	2.37	90.53	-1.54	2.29	90.77
	SHT-U	-0.21	1.93	92.13	-0.31	1.98	93.47
	SHT-1	-0.20	1.97	93.05	-0.35	1.92	92.95
	SHT-2	-0.25	1.94	92.36	-0.14	1.99	93.58
	SHT-12	-0.18	1.98	93.34	-0.23	1.93	93.19

Table 3.22: Relative biases of the $JK2 - A$ variance estimation for the smoothed Horvitz-Thompson estimators, average mean lengths and coverages of 95% confidence interval estimators under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-0.59	3.81	89.67	-0.72	3.30	90.48
	SHT-U	1.17	3.22	89.08	1.27	3.37	94.00
	SHT-1	1.42	3.45	92.33	1.45	3.40	94.02
	SHT-2	1.36	3.26	89.58	1.10	3.08	92.33
	SHT-12	1.49	3.46	92.52	1.13	3.10	92.68
500	HT	-1.18	2.50	90.41	-1.20	2.13	90.89
	SHT-U	-0.19	1.88	85.45	-0.48	2.06	85.31
	SHT-1	-0.09	2.04	93.14	-0.45	2.06	84.97
	SHT-2	-0.18	1.89	85.82	-0.47	1.82	92.84
	SHT-12	0.06	2.05	93.24	-0.42	1.82	92.89
n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-1.48	3.67	89.80	-1.25	3.56	89.80
	SHT-U	0.82	3.27	91.51	0.98	3.30	92.76
	SHT-1	0.92	3.33	92.36	0.94	3.25	92.05
	SHT-2	0.88	3.30	91.89	0.96	3.32	92.96
	SHT-12	0.92	3.35	92.59	0.91	3.28	92.41
500	HT	-0.95	2.37	90.53	-1.54	2.29	90.77
	SHT-U	-0.22	1.93	92.13	-0.32	1.98	93.47
	SHT-1	-0.22	1.97	93.04	-0.37	1.92	92.95
	SHT-2	-0.27	1.94	92.36	-0.15	1.99	93.58
	SHT-12	-0.21	1.98	93.34	-0.26	1.93	93.19

Table 3.23: Relative biases of the $JK1 - B$ variance estimation for the smoothed Horvitz-Thompson estimators, average mean lengths and coverages of 95% confidence interval estimators under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-0.59	3.81	89.67	-0.72	3.30	90.48
	SHT-U	-98.50	0.39	17.85	-98.60	0.40	17.12
	SHT-1	-92.76	0.92	37.08	-95.83	0.69	27.11
	SHT-2	-88.78	1.09	44.80	-96.23	0.59	26.98
	SHT-12	-83.02	1.42	55.35	-93.35	0.80	36.35
500	HT	-1.18	2.50	90.41	-1.20	2.13	90.89
	SHT-U	-99.13	0.18	11.84	-99.25	0.18	8.05
	SHT-1	-93.96	0.50	35.85	-96.47	0.39	16.74
	SHT-2	-89.66	0.61	38.68	-95.46	0.39	31.63
	SHT-12	-83.09	0.84	57.02	-91.98	0.52	41.61

n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-1.48	3.67	89.80	-1.25	3.56	89.80
	SHT-U	-98.53	0.39	19.02	-98.54	0.40	18.37
	SHT-1	-95.24	0.72	31.23	-97.56	0.51	22.91
	SHT-2	-88.64	1.11	46.34	-88.36	1.13	47.63
	SHT-12	-85.32	1.28	52.58	-87.40	1.16	49.15
500	HT	-0.95	2.37	90.53	-1.54	2.29	90.77
	SHT-U	-99.18	0.18	14.01	-99.21	0.18	13.60
	SHT-1	-95.95	0.40	30.42	-97.99	0.27	20.99
	SHT-2	-89.37	0.63	46.27	-89.31	0.65	45.51
	SHT-12	-85.13	0.76	53.79	-87.37	0.69	50.20

Table 3.24: Relative biases of the $JK2 - B$ variance estimation for the smoothed Horvitz-Thompson estimators, average mean lengths and coverages of 95% confidence interval estimators under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-0.59	3.81	89.67	-0.72	3.30	90.48
	SHT-U	-98.50	0.39	17.85	-98.60	0.40	17.12
	SHT-1	-92.76	0.92	37.08	-95.83	0.69	27.11
	SHT-2	-88.78	1.09	44.80	-96.23	0.59	26.98
	SHT-12	-83.02	1.42	55.35	-93.35	0.80	36.35
500	HT	-1.18	2.50	90.41	-1.20	2.13	90.89
	SHT-U	-99.13	0.18	11.84	-99.25	0.18	8.05
	SHT-1	-93.96	0.50	35.85	-96.47	0.39	16.74
	SHT-2	-89.66	0.61	38.68	-95.46	0.39	31.63
	SHT-12	-83.09	0.84	57.02	-91.98	0.52	41.61

n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HT	-1.48	3.67	89.80	-1.25	3.56	89.80
	SHT-U	-98.53	0.39	19.02	-98.54	0.40	18.37
	SHT-1	-95.24	0.72	31.23	-97.56	0.51	22.91
	SHT-2	-88.64	1.11	46.34	-88.36	1.13	47.63
	SHT-12	-85.32	1.28	52.58	-87.40	1.16	49.15
500	HT	-0.95	2.37	90.53	-1.54	2.29	90.77
	SHT-U	-99.18	0.18	14.01	-99.21	0.18	13.60
	SHT-1	-95.95	0.40	30.42	-97.99	0.27	20.99
	SHT-2	-89.37	0.63	46.27	-89.31	0.65	45.51
	SHT-12	-85.13	0.76	53.79	-87.37	0.69	50.20

Table 3.25: Relative biases of the $JK1-A$ variance estimation for the smoothed Hájek estimators, average mean lengths and coverages of 95% confidence interval estimators under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.06	0.693	92.11	0.98	0.722	90.69
	SHA-U	0.14	0.388	51.73	-1.25	0.392	3.11
	SHA-1	1.78	0.578	93.68	0.02	0.405	4.43
	SHA-2	0.61	0.423	58.31	-1.65	0.597	92.26
	SHA-12	1.91	0.581	93.65	-1.47	0.598	92.34
500	HA	1.93	0.419	92.30	0.84	0.450	91.32
	SHA-U	-0.09	0.175	0.00	1.77	0.178	0.00
	SHA-1	0.47	0.295	92.95	1.75	0.186	0.00
	SHA-2	-0.47	0.201	0.04	2.12	0.314	93.43
	SHA-12	0.37	0.301	92.89	2.11	0.317	93.52
n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.03	0.689	92.40	-0.60	0.687	92.44
	SHA-U	-0.23	0.391	92.75	1.00	0.393	83.97
	SHA-1	1.64	0.410	93.71	1.26	0.455	93.60
	SHA-2	2.32	0.426	92.83	1.13	0.427	84.78
	SHA-12	3.61	0.442	93.72	1.46	0.478	93.58
500	HA	1.08	0.408	93.41	0.79	0.412	92.58
	SHA-U	-1.07	0.174	63.80	-3.09	0.175	9.54
	SHA-1	-1.32	0.188	92.89	-1.44	0.215	93.27
	SHA-2	-0.96	0.202	73.53	-1.44	0.199	18.53
	SHA-12	-0.72	0.215	93.35	-0.35	0.234	93.60

Table 3.26: Relative biases of the $JK2-A$ variance estimation for the smoothed Hájek estimators, average mean lengths and coverages of 95% confidence interval estimators under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.00	0.693	92.11	0.89	0.722	90.69
	SHA-U	0.14	0.388	51.73	-1.25	0.392	3.11
	SHA-1	1.77	0.578	93.68	0.01	0.405	4.43
	SHA-2	0.61	0.423	58.31	-1.66	0.597	92.26
	SHA-12	1.90	0.581	93.65	-1.48	0.598	92.34
500	HA	1.88	0.419	92.30	0.77	0.450	91.32
	SHA-U	-0.09	0.175	0.00	1.77	0.178	0.00
	SHA-1	0.47	0.295	92.95	1.75	0.186	0.00
	SHA-2	-0.47	0.201	0.04	2.12	0.314	93.43
	SHA-12	0.37	0.301	92.89	2.11	0.317	93.52
n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	1.98	0.689	92.39	-0.65	0.686	92.44
	SHA-U	-0.23	0.391	92.75	1.00	0.393	83.97
	SHA-1	1.64	0.410	93.71	1.25	0.455	93.59
	SHA-2	2.31	0.426	92.83	1.12	0.427	84.78
	SHA-12	3.60	0.442	93.72	1.45	0.478	93.58
500	HA	1.05	0.408	93.41	0.76	0.412	92.58
	SHA-U	-1.07	0.174	63.80	-3.09	0.175	9.54
	SHA-1	-1.32	0.188	92.88	-1.44	0.215	93.27
	SHA-2	-0.96	0.202	73.53	-1.44	0.199	18.53
	SHA-12	-0.73	0.215	93.35	-0.35	0.234	93.60

Table 3.27: Relative biases of the $JK1-B$ variance estimation for the smoothed Hájek estimators, average mean lengths and coverages of 95% confidence interval estimators under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.06	0.693	92.11	0.98	0.722	90.69
	SHA-U	0.14	0.388	51.73	-1.25	0.392	3.11
	SHA-1	-49.93	0.406	82.33	-1.53	0.401	4.19
	SHA-2	-0.75	0.420	58.10	-41.78	0.459	84.13
	SHA-12	-42.28	0.437	84.68	-39.88	0.467	84.57
500	HA	1.93	0.419	92.30	0.84	0.450	91.32
	SHA-U	-0.09	0.175	0.00	1.77	0.178	0.00
	SHA-1	-58.97	0.189	76.50	1.45	0.186	0.00
	SHA-2	-0.78	0.201	0.04	-45.07	0.230	83.55
	SHA-12	-47.48	0.218	81.98	-41.72	0.240	84.10
n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.03	0.689	92.40	-0.60	0.687	92.44
	SHA-U	-0.23	0.391	92.75	1.00	0.393	83.97
	SHA-1	-3.64	0.399	93.31	-20.40	0.404	90.56
	SHA-2	1.07	0.423	92.51	-0.03	0.425	84.69
	SHA-12	-1.61	0.430	93.30	-16.17	0.435	91.36
500	HA	1.08	0.408	93.41	0.79	0.412	92.58
	SHA-U	-1.07	0.174	63.80	-3.09	0.175	9.54
	SHA-1	-7.40	0.182	91.97	-27.78	0.184	88.65
	SHA-2	-1.32	0.201	73.43	-1.84	0.199	18.40
	SHA-12	-4.94	0.211	92.88	-19.72	0.210	90.15

Table 3.28: Relative biases of the $JK2-B$ variance estimation for the smoothed Hájek estimators, average mean lengths and coverages of 95% confidence interval estimators under correct model.

n	Estimators	Variable y_1			Variable y_2		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	2.00	0.693	92.11	0.89	0.722	90.69
	SHA-U	0.14	0.388	51.73	-1.25	0.392	3.11
	SHA-1	-49.93	0.406	82.33	-1.53	0.401	4.19
	SHA-2	-0.75	0.420	58.10	-41.78	0.459	84.13
	SHA-12	-42.29	0.437	84.68	-39.89	0.467	84.57
500	HA	1.88	0.419	92.30	0.77	0.450	91.32
	SHA-U	-0.09	0.175	0.00	1.77	0.178	0.00
	SHA-1	-58.97	0.189	76.50	1.45	0.186	0.00
	SHA-2	-0.78	0.201	0.04	-45.08	0.230	83.55
	SHA-12	-47.48	0.218	81.98	-41.72	0.240	84.10
n	Estimators	Variable y_3			Variable y_4		
		RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
100	HA	1.98	0.689	92.39	-0.65	0.686	92.44
	SHA-U	-0.23	0.391	92.75	1.00	0.393	83.97
	SHA-1	-3.64	0.399	93.31	-20.40	0.404	90.56
	SHA-2	1.07	0.423	92.51	-0.03	0.425	84.69
	SHA-12	-1.61	0.430	93.30	-16.17	0.435	91.36
500	HA	1.05	0.408	93.41	0.76	0.412	92.58
	SHA-U	-1.07	0.174	63.80	-3.09	0.175	9.54
	SHA-1	-7.40	0.182	91.97	-27.78	0.184	88.65
	SHA-2	-1.32	0.201	73.43	-1.84	0.199	18.40
	SHA-12	-4.94	0.211	92.88	-19.72	0.210	90.15

Beaumont [2008]. For this model, asymptotic theory was developed, including the consistency of the weight smoothing estimator with respect to the sampling design and the weight model, and the asymptotic unbiasedness of a model-based variance estimator. This extends the results of Beaumont [2008], who only sketched these theoretical results. Simulation experiments have explored the effect of model choice and inclusion/exclusion of model covariates on the practical behavior of the weight smoothing estimator. Our results show that the estimator and the model-based variance estimator perform well when the model is correctly specified, but can be unacceptably biased in case of model failure.

We therefore considered a Hájek-type version of the weight smoothing estimator. This is a commonly used survey estimation adjustment to improve the efficiency of survey estimators, and we applied it here to improve their robustness to model misspecification. The results were encouraging, in the sense that the adjusted weight smoothing estimator was close to unbiased and more efficient than the unsmoothed estimator under modestly misspecified models, as long as the required covariates are included in the weight model.

With regards to inference, the Hájek-type adjustment was not effective in correcting the model-based variance estimator under model misspecification. However, the delete-a-group jackknife, a widely used variance estimation replication method, resulted in variance estimators with low bias and confidence intervals with close-to-nominal coverage, for the same scenarios in which the Hájek-type weight smoothing estimator worked well. While we did not explore other replication methods, we conjecture that other jackknife or bootstrap methods will perform equally well.

CHAPTER 4

AN INVESTIGATION OF WEIGHT SMOOTHING ESTIMATORS UNDER MIXED MODEL SPECIFICATIONS

4.1 Weight model with random effects

In the previous chapter, we considered a model-based weight smoothing estimator, using a lognormal linear model specification for the survey weights. These results are extended here to the case with a random effect weight model.

As before, we address the estimation of $\bar{Y} = \frac{1}{N}\sum_{i\in U_N} y_i$ for a finite population U_N. A sample S of size n is drawn from U_N with the survey weight w_i, based on the sampling design variables are specified. The Horvitz-Thompson estimator of $\hat{Y}^{HT}$ is

$$\hat{Y}^{HT} = \frac{1}{N}\sum_{i\in S} w_i y_i,$$

which is unbiased estimator for $\bar{Y}$. Following the modeling approach described in the previous chapter, we generalize the model here to the case with random effects, i.e.

$$w_i = 1 + \exp(\boldsymbol{B}_i\boldsymbol{\nu} + \boldsymbol{H}_i\boldsymbol{u} + \varepsilon_i)$$

for $i \in S$, where $\boldsymbol{B}_i = [\boldsymbol{B}_1(y_i), \cdots, \boldsymbol{B}_{r_1}(y_i)]$ and $\boldsymbol{H}_i = [\boldsymbol{H}_1(y_i), \cdots, \boldsymbol{H}_{r_2}(y_i)]$ are known functions depending on y_i, and $\boldsymbol{\nu}_{r_1\times 1}$ and $\boldsymbol{u}_{r_2\times 1}$ are vectors of coefficients. The vector of coefficients $\boldsymbol{u}$ given $\boldsymbol{I}_N$ and $\boldsymbol{Y}_N$ are normally distributed with $\mathrm{E}(\boldsymbol{u}|\boldsymbol{I}_N, \boldsymbol{Y}_N) = 0$ and $\mathrm{Var}(\boldsymbol{u}|\boldsymbol{I}_N, \boldsymbol{Y}_N) = \sigma_u^2\boldsymbol{G}$, where $\boldsymbol{G}$ is a known $r_2 \times r_2$ matrix, which could be the identity matrix. The ε_i given $\boldsymbol{I}_N$ and $\boldsymbol{Y}_N$ are independently and identically normally distributed with $\mathrm{E}(\varepsilon_i|\boldsymbol{I}_N, \boldsymbol{Y}_N) = 0$ and $\mathrm{Var}(\varepsilon_i|\boldsymbol{I}_N, \boldsymbol{Y}_N) = \sigma_\varepsilon^2 > 0$. We define $\lambda^2 = \sigma_\varepsilon^2/\sigma_u^2$, and we will assume here that λ is known. Hence, the only unknown model parameters are $\boldsymbol{\nu}$ and σ_ε^2.

There are two possible approaches to construct a weight smoothing estimator, depending on whether the random effects are predicted or removed from the smoothed weights. For the former approach, the expected weight is taken to be conditional on the random effect $\boldsymbol{u}$, so that the

(unfeasible) weight smoothing estimator is defined as

$$\begin{aligned}\tilde{Y}^{SHT,BLUP} &= \frac{1}{N}\sum_{i\in S}\mathrm{E}\left(w_i|\boldsymbol{I}_N,\boldsymbol{Y}_N,\boldsymbol{u}\right)y_i=\sum_{i\in S}\tilde{w}_i y_i \\ &= \frac{1}{N}\sum_{i\in S}\left\{1+\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\boldsymbol{H}_i\boldsymbol{u}+\sigma_\varepsilon^2/2\right)\right\}y_i.\end{aligned}$$

This contains unknown parameters as well as the unknown values of the random effect. We will apply Best Linear Unbiased Prediction (BLUP) ideas to estimate $\boldsymbol{\nu}$ and σ_ε^2 and predict $\boldsymbol{u}$.

Letting $l_i = \log(w_i - 1)$, we obtain a linear mixed model with normal errors and random effects, for which straightforward application of BLUP leads to

$$\begin{aligned}\hat{\boldsymbol{\nu}} &= \left(\boldsymbol{B}^T\boldsymbol{V}^{*-1}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^T\boldsymbol{V}^{*-1}\boldsymbol{L}, \\ \hat{\boldsymbol{u}} &= \lambda^{-2}\boldsymbol{G}\boldsymbol{H}^T\boldsymbol{V}^{*-1}(\boldsymbol{L}-\boldsymbol{B}\hat{\boldsymbol{\nu}}), \\ \hat{\sigma}_\varepsilon^2 &= \frac{1}{n-r_1}\left(\boldsymbol{L}-\boldsymbol{B}\hat{\boldsymbol{\nu}}-\boldsymbol{H}\hat{\boldsymbol{u}}\right)^T\boldsymbol{V}^*\left(\boldsymbol{L}-\boldsymbol{B}\hat{\boldsymbol{\nu}}-\boldsymbol{H}\hat{\boldsymbol{u}}\right),\end{aligned}$$

with

$$\boldsymbol{V}^* = \boldsymbol{I}_n + \lambda^{-2}\boldsymbol{H}\boldsymbol{G}\boldsymbol{H}^T.$$

Hence, the (feasible) BLUP smoothed estimator is given by

$$\hat{Y}^{SHT,BLUP} = \frac{1}{N}\sum_{i\in S}\left\{1+\exp\left(\boldsymbol{B}_i\hat{\boldsymbol{\nu}}+\boldsymbol{H}_i\hat{\boldsymbol{u}}+\hat{\sigma}_\varepsilon^2/2\right)\right\}y_i.$$

Under the latter approach, the random effect is integrated out, which implies that it becomes part of the variance structure of the model, rather than its (conditional) mean. To differentiate it from the previous approach, we will denote the estimator as NOBLUP. The unfeasible estimator is now defined as

$$\begin{aligned}\tilde{Y}^{SHT,NOBLUP} &= \frac{1}{N}\sum_{i\in S}\mathrm{E}\left(w_i|\boldsymbol{I}_N,\boldsymbol{Y}_N\right)y_i=\sum_{i\in S}\tilde{w}_i y_i \\ &= \frac{1}{N}\sum_{i\in S}\left\{1+\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\boldsymbol{H}_i\boldsymbol{G}\boldsymbol{H}_i^T\sigma_u^2/2+\sigma_\varepsilon^2/2\right)\right\}y_i \\ &= \frac{1}{N}\sum_{i\in S}\left\{1+\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\left(\lambda^{-2}\boldsymbol{H}_i\boldsymbol{G}\boldsymbol{H}_i^T+1\right)\sigma_\varepsilon^2/2\right)\right\}y_i.\end{aligned}$$

To obtain the feasible smoothed weight estimator, we again estimate the model parameters $\boldsymbol{\nu}$ and σ_ε^2 by

$$\begin{aligned}\hat{\boldsymbol{\nu}} &= \left(\boldsymbol{B}^T\boldsymbol{V}^{*-1}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^T\boldsymbol{V}^{*-1}\boldsymbol{L},\\ \hat{\sigma}_\varepsilon^2 &= \frac{1}{n-r_1}\left(\boldsymbol{L}-\boldsymbol{B}\hat{\boldsymbol{\nu}}-\boldsymbol{H}\hat{\boldsymbol{u}}\right)^T\boldsymbol{V}^*\left(\boldsymbol{L}-\boldsymbol{B}\hat{\boldsymbol{\nu}}-\boldsymbol{H}\hat{\boldsymbol{u}}\right)\end{aligned}$$

with

$$\boldsymbol{V}^* = \boldsymbol{I}_n + \lambda^{-2}\boldsymbol{H}\boldsymbol{G}\boldsymbol{H}^T,$$

and write

$$\hat{Y}^{SHT,NOBLUP} = \frac{1}{N}\sum_{i\in S}\left\{1+\exp\left(\boldsymbol{B}_i\hat{\boldsymbol{\nu}}+\left(\lambda^{-2}\boldsymbol{H}_i\boldsymbol{G}\boldsymbol{H}_i^T+1\right)\hat{\sigma}_\varepsilon^2\right)\right\}y_i.$$

It will be of interest to compare the behavior of the two above versions of the weight smoothing estimators, and in particular, to determine whether an estimator that includes prediction might be preferable to one that removes the random effect. This will be done through simulations in the next section.

4.2 Simulation

As explained in the previous chapter, population distributions need to be constructed so that the sample distributions have the desired functional forms. Using the results of Pfeffermann and Sverchkov [1999], we obtain that the population distribution of the weights is a mixture of two lognormal distributions, i.e.

$$w_i - 1|y_i \sim p\times lN\left(\boldsymbol{B}_i\boldsymbol{\nu}+\boldsymbol{H}_i\boldsymbol{u},\sigma_\varepsilon^2\right)+(1-p)\times lN\left(\boldsymbol{B}_i\boldsymbol{\nu}+\boldsymbol{H}_i\boldsymbol{u}+\sigma_\varepsilon^2,\sigma_\varepsilon^2\right)$$

with

$$p = \left\{1+\exp\left(\boldsymbol{B}_i\boldsymbol{\nu}+\boldsymbol{H}\boldsymbol{u}+\sigma_\varepsilon^2/2\right)\right\}^{-1}.$$

For the random effect, we want that $\boldsymbol{u}$ has a normal distribution with mean 0 and variance $\sigma_u^2\boldsymbol{G}$ under the sample. Applying Pfeffermann and Sverchkov [1999] again, we find that the population model for the $\boldsymbol{u}$ is a mixture of normal distributions, or

$$\boldsymbol{u} \sim p\times N\left(0,\sigma_u^2\boldsymbol{G}\right)+(1-p)\times N\left(\sigma_u^2\boldsymbol{H}_i\boldsymbol{G},\sigma_u^2\boldsymbol{G}\right)$$

with

$$p = \left\{ 1 + \exp\left(\boldsymbol{B}_i \boldsymbol{\nu} + \frac{\sigma_u^2 \boldsymbol{H}_i \boldsymbol{G} \boldsymbol{H}_i^T}{2} + \frac{\sigma_\varepsilon^2}{2} \right) \right\}^{-1}. \tag{4.1}$$

However, this result is for random variables that are independently generated across all $i \in U_N$. This is not the case for $\boldsymbol{u}$, which represents a single random vector for the population. We will discuss a number of ways to address this below.

The population variables of interest are generated from a multivariate normal distribution, which is identical to that in the previous chapter. To generate the random effect, we split the population in groups of k people. Assume N/k is an integer. So, there are $r_2 = N/k$ groups. The sample sizes $n = 100$ and $n = 500$ are considered. We use 200 random groups for $n = 100$ and 2000 groups for $n = 500$. For each group, we randomly generate a random intercept, which will follow a mixture of normal distributions. We consider four different scenarios to generate the random intercepts, depending on whether we use a unique or a group-specific mixture coefficient p and whether we average the $\boldsymbol{B}_i, \boldsymbol{H}_i$ prior to computing p or first compute p for each i and then average.

The four random effect population generation scenarios are compared based on whether the sample distribution of the random effects follows the stated normal. Figure 4.1 shows Q-Q plots for representative single sample realizations of the four scenarios. Plots (a) and (b) are for p computed for each i, and plots (c) and (d) are for $\boldsymbol{B}_i, \boldsymbol{H}_i$ averaged prior to computing p. Plots (a) and (c) are based on a single distribution for the population, while plots (b) and (d) are for separate distributions for each group. None of these plots appear to indicate severe departures from normality in the sample. We also used the Shapiro-Wilks test to detect deviations from normality in repeated simulations, and none of the four scenarios led to rejection of normality at more than the 5% level. Hence, we decided to generate the population random effects from a single population mixture distribution with p computed for each i and averaged for the population (corresponding to plot (a) in Figure 4.1).

Next, we generated the weights. A population $w_i - 1$ of $50,000$ units is generated from a mixture of two Log-Normal distribution of $lN(\boldsymbol{B}_i\boldsymbol{\nu} + \boldsymbol{H}_i\boldsymbol{u}, \sigma_\varepsilon^2)$ and $lN(\boldsymbol{B}_i\boldsymbol{\nu} + \boldsymbol{H}_i\boldsymbol{u} + \sigma_\varepsilon^2, \sigma_\varepsilon^2)$ with the proportion $p = \{1 + \exp(\boldsymbol{B}_i\boldsymbol{\nu} + \boldsymbol{H}\boldsymbol{u} + \sigma_\varepsilon^2/2)\}$ and $1 - p$, respectively.

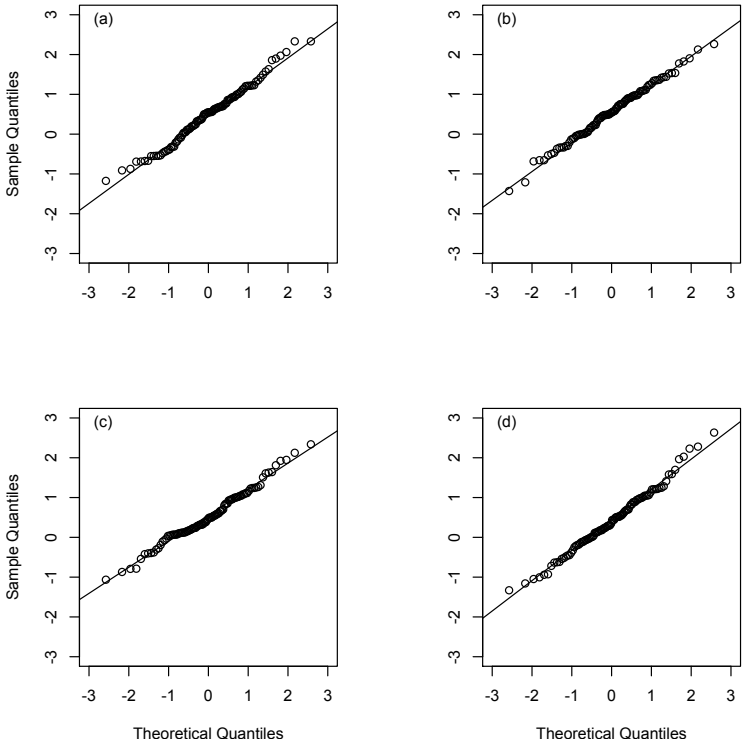

Figure 4.1: Q-Q plots of a sample of a random effect for (a) overall mean of the mixture probability, (b) mean of the mixture probability by group, (c) mixture probability from the overall mean of the population and (d) mixture probability from the mean of the population by group.

Following construction of these population variables and weights, independent samples of size $n = 100$ and $n = 500$ are generated under pps sampling with replacement. We set $\boldsymbol{\nu}_i = (7.06, 0.3, -0.5)'$ when the sample size is $n = 100$ and $\boldsymbol{\nu}_i = (5.41, 0.3, -0.5)'$ when the sample size is $n = 500$. The Monte Carlo sample sizes are all $10,000$ in the simulation.

Seven estimators are computed: Horvitz-Thompson (HT), and six different weight smoothing estimators, denoted SHT-U, SHT-1, SHT-2, SHT-12, HT-GLS-BLUP and HT-GLS-NOBLUP. For the weight smoothing estimators SHT-U, SHT-1, SHT-2 and SHT-12, the parameter estimators $\hat{\boldsymbol{\nu}}$ and $\hat{\sigma}_\varepsilon^2$ are obtained from the ordinary least square estimation with different sets of covariates, as was done in the previous chapter. The SHT-U estimator contains $\boldsymbol{B}_i = 1$ only, while the SHT-1 estimator adds y_{1i}, the SHT-2 estimator adds y_{2i} and the SHT-12 estimator adds (y_{1i}, y_{2i}) to $\boldsymbol{B}_i$, respectively. For the weight smoothing estimators HT-GLS-BLUP and HT-GLS-NOBLUP, the parameter estimators $\hat{\boldsymbol{\nu}}$ and $\hat{\sigma}_\varepsilon^2$ are obtained by generalized least square estimation on the model with $\boldsymbol{B}_i = (1, y_{1i}, y_{2i})$, followed by prediction of $\hat{\boldsymbol{u}}$ in the case of HT-GLS-BLUP.

Table 4.1 and Table 4.2 present the relative biases as percentages (RB) and the relative efficiencies as percentages (RE) of the estimators when the sample size is 100 and the sample size is 500, respectively. For Table 4.1, a sample size of 100 with groups of 200 for the random effects, we see SHT-12 is the best estimator for all variables, which is unbiased and more efficient than the Horvitz-Thompson estimator. The two models with random effect seem to be biased and less efficient than the Horvitz-Thompson estimator. The same results hold for $n = 500$, as shown in Table 2.

We implement delete-a-group Jackknife (JKDAG) variance estimation as described in the previous chapter, with $R = 20$. We consider the four jackknife variance estimators $JK1 - A$, $JK2 - A$, $JK1 - B$ and $JK2 - B$. Table 4.3 through Table 4.10 present the relative biases of the jackknife variance estimators, average mean lengths and coverages of 95% confidence interval estimators for $JK1-A$, $JK2-A$, $JK1-B$ and $JK2-B$ when the sample size is 100 or 500, respectively. The $JK1$ and $JK2$ again exhibit very similar results. For $JK - A$, the results from the smoothed Horvitz-Thompson estimators without random effect have a modest positive bias

Table 4.1: Relative biases and relative efficiency results for a sample size $n = 100$ and 200 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1		Variable y_2		Variable y_3		Variable y_4	
	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)
HT	-0.14	100.00	-0.09	100.00	-0.10	100.00	-0.08	100.00
SHT-U	-4.83	66.88	8.38	125.92	-0.67	71.85	2.79	82.57
SHT-1	1.00	74.54	8.90	130.00	0.94	75.65	0.80	75.79
SHT-2	-4.21	67.38	0.90	81.40	-0.14	73.23	3.15	83.79
SHT-12	1.25	75.30	1.34	82.72	1.23	76.63	1.04	76.37
HT-GLS-BLUP	-9.57	78.26	-9.48	89.91	-9.58	81.37	-9.64	83.10
HT-GLS-noBLUP	10.92	122.13	11.06	147.41	10.92	127.95	10.74	129.78

Table 4.2: Relative biases and relative efficiency results for a sample size $n = 500$ and 2000 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1		Variable y_2		Variable y_3		Variable y_4	
	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)
HT	0.19	100.00	0.20	100.00	0.17	100.00	0.18	100.00
SHT-U	-4.04	84.48	9.30	261.12	0.00	73.22	3.25	99.19
SHT-1	0.94	75.98	9.39	264.20	1.14	77.67	0.96	79.01
SHT-2	-3.88	83.80	1.11	84.24	0.06	73.85	3.33	100.24
SHT-12	0.92	76.17	1.19	84.70	1.09	77.87	0.98	79.45
HT-GLS-BLUP	11.25	217.81	-11.16	282.22	-11.20	232.21	-11.27	250.48
HT-GLS-noBLUP	7.01	148.72	7.18	188.01	7.11	158.37	6.94	163.08

Table 4.3: Relative biases of the $JK1-A$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 100$ and 200 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HT	-1.27	4.737	88.00	-1.27	3.980	89.12	-0.56	4.513	88.05	-0.18	4.362	88.56
SHT-U	1.51	3.757	86.74	1.23	4.064	94.62	1.47	3.861	91.08	1.66	3.945	93.42
SHT-1	1.67	4.143	92.36	1.41	4.086	94.58	1.74	3.964	92.34	1.89	3.832	92.39
SHT-2	1.96	3.823	87.57	1.36	3.632	92.50	1.83	3.908	91.65	2.18	3.969	93.68
SHT-12	2.09	4.169	92.42	1.50	3.656	92.76	2.10	3.992	92.65	2.43	3.854	92.46
HT-GLS-BLUP	4.96	3.653	79.10	4.74	3.145	79.24	5.37	3.493	79.31	5.59	3.361	79.04
HT-GLS-noBLUP	9.52	4.814	95.44	8.87	4.287	94.94	9.59	4.634	95.33	10.00	4.492	95.26

Table 4.4: Relative biases of the $JK2-A$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 100$ and 200 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HT	-1.27	4.737	88.00	-1.27	3.980	89.12	-0.56	4.513	88.05	-0.18	4.362	88.56
SHT-U	1.46	3.756	86.73	1.19	4.063	94.62	1.43	3.860	91.07	1.61	3.944	93.42
SHT-1	1.59	4.142	92.35	1.33	4.085	94.56	1.65	3.962	92.33	1.80	3.830	92.35
SHT-2	1.88	3.821	87.55	1.28	3.630	92.49	1.75	3.906	91.63	2.10	3.968	93.65
SHT-12	1.98	4.167	92.41	1.38	3.654	92.76	1.98	3.990	92.63	2.31	3.851	92.45
HT-GLS-BLUP	4.77	3.650	79.08	4.53	3.142	79.22	5.17	3.490	79.30	5.39	3.357	79.03
HT-GLS-noBLUP	8.79	4.798	95.37	8.00	4.270	94.73	8.81	4.618	95.21	9.19	4.476	95.13

and the coverage for the weight smoothing estimators that is at least as good as for the unsmoothed estimator, when the mean estimator itself is appropriate. For the smoothed Horvitz-Thompson estimators with random effect, both HT-GLS-BLUP estimator and HT-GLS-NOBLUP overestimate the true variance. However, the model with random effect where we do not predict it is the best one for the coverages. Specifically, the HT-GLS-NOBLUP estimator have good coverage, which is close to 95%. In contrast, $JK - B$ performed very poorly, which is not suggested for use.

Table 4.5: Relative biases of the $JK1-B$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 100$ and 200 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HT	-1.27	4.737	88.00	-1.27	3.980	89.12	-0.56	4.513	88.05	-0.18	4.362	88.56
SHT-U	-98.85	0.399	14.74	-99.04	0.396	14.43	-98.96	0.391	14.96	-98.96	0.399	15.64
SHT-1	-90.14	1.291	42.59	-93.91	1.002	31.64	-93.23	1.022	35.52	-96.16	0.744	27.09
SHT-2	-87.65	1.331	44.52	-94.61	0.837	31.61	-87.44	1.372	47.29	-87.18	1.406	48.48
SHT-12	-78.16	1.928	59.69	-88.91	1.208	45.25	-81.07	1.719	56.86	-83.93	1.526	53.28
HT-GLS-BLUP	-54.42	2.408	61.49	-64.26	1.837	55.42	-56.56	2.243	60.22	-59.10	2.092	58.78
HT-GLS-noBLUP	-79.37	2.090	55.43	-89.91	1.305	37.86	-82.26	1.865	51.02	-85.07	1.655	46.90

Table 4.6: Relative biases of the $JK2-B$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 100$ and 200 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HT	-1.27	4.737	88.00	-1.27	3.980	89.12	-0.56	4.513	88.05	-0.18	4.362	88.56
SHT-U	-98.85	0.399	14.74	-99.04	0.396	14.43	-98.96	0.391	14.96	-98.96	0.399	15.64
SHT-1	-90.14	1.291	42.59	-93.91	1.002	31.64	-93.23	1.022	35.52	-96.16	0.744	27.09
SHT-2	-87.65	1.331	44.52	-94.61	0.837	31.61	-87.44	1.372	47.29	-87.18	1.406	48.48
SHT-12	-78.16	1.928	59.69	-88.91	1.208	45.25	-81.07	1.719	56.86	-83.93	1.526	53.28
HT-GLS-BLUP	-54.42	2.408	61.49	-64.26	1.837	55.42	-56.56	2.243	60.22	-59.10	2.092	58.78
HT-GLS-noBLUP	-79.37	2.090	55.43	-89.91	1.305	37.86	-82.26	1.865	51.02	-85.07	1.655	46.90

Table 4.7: Relative biases of the $JK1-A$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 500$ and 2000 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HT	-1.27	4.737	88.00	-1.27	3.980	89.12	-0.56	4.513	88.05	-0.18	4.362	88.56
SHT-U	1.51	3.757	86.74	1.23	4.064	94.62	1.47	3.861	91.08	1.66	3.945	93.42
SHT-1	1.67	4.143	92.36	1.41	4.086	94.58	1.74	3.964	92.34	1.89	3.832	92.39
SHT-2	1.96	3.823	87.57	1.36	3.632	92.50	1.83	3.908	91.65	2.18	3.969	93.68
SHT-12	2.09	4.169	92.42	1.50	3.656	92.76	2.10	3.992	92.65	2.43	3.854	92.46
HT-GLS-BLUP	4.96	3.653	79.10	4.74	3.145	79.24	5.37	3.493	79.31	5.59	3.361	79.04
HT-GLS-noBLUP	9.52	4.814	95.44	8.87	4.287	94.94	9.59	4.634	95.33	10.00	4.492	95.26

Table 4.8: Relative biases of the $JK2-A$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 500$ and 2000 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HT	-1.27	4.737	88.00	-1.27	3.980	89.12	-0.56	4.513	88.05	-0.18	4.362	88.56
SHT-U	1.46	3.756	86.73	1.19	4.063	94.62	1.43	3.860	91.07	1.61	3.944	93.42
SHT-1	1.59	4.142	92.35	1.33	4.085	94.56	1.65	3.962	92.33	1.80	3.830	92.35
SHT-2	1.88	3.821	87.55	1.28	3.630	92.49	1.75	3.906	91.63	2.10	3.968	93.65
SHT-12	1.98	4.167	92.41	1.38	3.654	92.76	1.98	3.990	92.63	2.31	3.851	92.45
HT-GLS-BLUP	4.77	3.650	79.08	4.53	3.142	79.22	5.17	3.490	79.30	5.39	3.357	79.03
HT-GLS-noBLUP	8.79	4.798	95.37	8.00	4.270	94.73	8.81	4.618	95.21	9.19	4.476	95.13

Table 4.9: Relative biases of the $JK1-B$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 500$ and 2000 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HT	-1.27	4.737	88.00	-1.27	3.980	89.12	-0.56	4.513	88.05	-0.18	4.362	88.56
SHT-U	-98.85	0.399	14.74	-99.04	0.396	14.43	-98.96	0.391	14.96	-98.96	0.399	15.64
SHT-1	-90.14	1.291	42.59	-93.91	1.002	31.64	-93.23	1.022	35.52	-96.16	0.744	27.09
SHT-2	-87.65	1.331	44.52	-94.61	0.837	31.61	-87.44	1.372	47.29	-87.18	1.406	48.48
SHT-12	-78.16	1.928	59.69	-88.91	1.208	45.25	-81.07	1.719	56.86	-83.93	1.526	53.28
HT-GLS-BLUP	-54.42	2.408	61.49	-64.26	1.837	55.42	-56.56	2.243	60.22	-59.10	2.092	58.78
HT-GLS-noBLUP	-79.37	2.090	55.43	-89.91	1.305	37.86	-82.26	1.865	51.02	-85.07	1.655	46.90

Table 4.10: Relative biases of the $JK2 - B$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 500$ and 2000 random groups for the smoothed Horvitz-Thompson estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HT	-1.27	4.737	88.00	-1.27	3.980	89.12	-0.56	4.513	88.05	-0.18	4.362	88.56
SHT-U	-98.85	0.399	14.74	-99.04	0.396	14.43	-98.96	0.391	14.96	-98.96	0.399	15.64
SHT-1	-90.14	1.291	42.59	-93.91	1.002	31.64	-93.23	1.022	35.52	-96.16	0.744	27.09
SHT-2	-87.65	1.331	44.52	-94.61	0.837	31.61	-87.44	1.372	47.29	-87.18	1.406	48.48
SHT-12	-78.16	1.928	59.69	-88.91	1.208	45.25	-81.07	1.719	56.86	-83.93	1.526	53.28
HT-GLS-BLUP	-54.42	2.408	61.49	-64.26	1.837	55.42	-56.56	2.243	60.22	-59.10	2.092	58.78
HT-GLS-noBLUP	-79.37	2.090	55.43	-89.91	1.305	37.86	-82.26	1.865	51.02	-85.07	1.655	46.90

As both HT-GLS-BLUP estimator and HT-GLS-NOBLUP estimator from the model with random effect are biased, we consider the Hájek estimator instead. The smoothed Hájek estimator is given by $\hat{Y}^{SHA} = \sum_{i \in S} \hat{w}_i y_i / \sum_{i \in S} \hat{w}_i$.

Table 4.11 and 4.12 present the relative biases as percentages (RB) and relative efficiencies as percentages (RE) of the estimators when the sample size is 100 and the sample size is 500, respectively. For variable y_1, all the estimators from the model with y_1 are unbiased and more efficient than the Hájek estimator. The model without y_1 are biased and less efficient. For the two estimators with random effect, both HA-GLS-BLUP and HA-GLS-NOBLUP estimators are unbiased. The HA-GLS-NOBLUP estimator has slight advantage with respect to the relative efficiency compared to the HA-GLS-BLUP estimator. For variable y_2, all the estimators from the model with y_2 are unbiased and more efficient than the Hájek estimator. Both HA-GLS-BLUP and HA-GLS-NOBLUP estimators are unbiased and more efficient than the Hájek estimator. For variable y_3, all smoothed Hájek estimators are unbiased and more efficient than Hájek estimator. This is because the variable y_3 is weakly correlated with the design variable. For variable y_4, the results are identical to y_1 as expected since the two variables are highly correlated.

These results suggest that the bias of the GLS Horvitz-Thompson estimators is readily removed by switching to a Hájek version. We also see that the NOBLUP is more efficient than the BLUP for all variables and sample sizes. This makes sense here, because the random intercept is not related to the y variables, so that integrating it out of the weights improves their behavior with respect to estimation of y population means. This is different from most applications of mixed

Table 4.11: Relative biases and relative efficiency results for a sample size $n = 100$ and 200 random groups for the Hájek estimators.

Estimators	Variable y_1		Variable y_2		Variable y_3		Variable y_4	
	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)
HA	-0.22	100.00	0.33	100.00	-0.03	100.00	0.11	100.00
SHA-U	-5.12	259.62	8.18	581.59	-0.94	35.16	2.55	91.19
SHA-1	0.00	64.48	8.05	566.01	0.04	31.11	-0.01	40.94
SHA-2	-5.00	255.18	0.34	63.76	-0.91	41.88	2.41	91.72
SHA-12	-0.05	66.47	0.37	64.53	0.03	38.54	-0.07	47.63
HA-GLS-BLUP	-0.12	71.33	0.33	70.41	-0.02	52.23	0.01	57.27
HA-GLS-noBLUP	-0.08	62.63	0.36	60.39	0.02	38.17	-0.05	46.60

Table 4.12: Relative biases and relative efficiency results for a sample size $n = 500$ and 2000 random groups the Hájek estimators.

Estimators	Variable y_1		Variable y_2		Variable y_3		Variable y_4	
	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)	RB (%)	RE (%)
HA	-0.03	100.00	0.08	100.00	-0.03	100.00	0.01	100.00
SHA-U	-4.87	1002.74	8.37	2500.09	-0.86	56.77	2.37	268.79
SHA-1	-0.08	58.75	8.32	2474.09	0.14	28.42	-0.03	34.54
SHA-2	-4.83	990.87	0.16	56.36	-0.92	68.71	2.32	265.21
SHA-12	-0.12	60.56	0.20	57.08	0.05	35.56	-0.03	40.30
HA-GLS-BLUP	-0.05	66.17	0.13	64.14	0.02	49.79	-0.03	52.04
HA-GLS-noBLUP	-0.04	57.33	0.18	54.17	0.07	35.65	-0.07	39.85

Table 4.13: Relative biases of the $JK1-A$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 100$ and 200 random groups the Hájek estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HA	-1.71	0.779	90.39	3.53	0.825	89.08	5.31	0.769	92.67	3.74	0.767	92.42
SHA-U	-0.95	0.392	14.11	0.89	0.389	0.12	-0.52	0.385	89.53	-1.09	0.393	64.53
SHA-1	-1.89	0.626	92.95	3.29	0.416	0.85	1.63	0.421	93.70	-0.37	0.481	93.17
SHA-2	-0.16	0.447	25.29	2.92	0.655	92.32	3.17	0.443	90.56	1.31	0.450	72.81
SHA-12	-1.41	0.637	92.86	3.22	0.659	91.94	3.89	0.474	93.96	1.54	0.524	93.55
HA-GLS-BLUP	0.35	0.665	92.46	5.66	0.698	91.39	9.81	0.568	94.02	6.90	0.590	93.66
HA-GLS-noBLUP	4.93	0.638	93.62	10.76	0.660	92.95	4.41	0.473	93.91	3.58	0.523	93.55

model regression, where prediction is most often an emphasis of the model analysis. We expect this conclusion to change in cases where the random effect variable itself related to the survey variables, for instance if the random effect corresponds to interactions between them. In such cases, the BLUP can be expected to be more efficient than the NOBLUP estimator.

We evaluate the JKDAG variance estimators for the Hájek estimators. Table 4.13 through Table 4.20 present the relative biases of the jackknife estimators, average mean lengths and coverages of 95% confidence interval estimators for $JK1-A$, $JK2-A$, $JK1-B$ and $JK2-B$, respectively. The $JK1-A$ and $JK2-A$ have very similar results. For $JK-A$, all the smoothed Hájek estimators without random effect have modest biases and the coverages of 95% confidence interval estimators are good when the mean estimator itself is appropriate. Both HA-GLS-BLUP estimator and HA-GLS-NOBLUP estimator overestimate the true variance but they have shorter mean length compared to Hájek estimator. They both provide good coverage for all variables, which is close to 95%. The same as $JK-B$ for Horvitz-Thompson estimator, $JK-B$ for the Hájek estimator performed very poorly.

Table 4.14: Relative biases of the $JK2 - A$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size $n = 100$ and 200 random groups the Hájek estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HA	-1.81	0.778	90.39	3.36	0.824	89.08	5.23	0.769	92.67	3.63	0.767	92.42
SHA-U	-0.95	0.392	14.11	0.89	0.389	0.12	-0.52	0.385	89.53	-1.09	0.393	64.53
SHA-1	-1.90	0.626	92.95	3.27	0.416	0.85	1.62	0.421	93.70	-0.38	0.481	93.17
SHA-2	-0.18	0.447	25.29	2.90	0.655	92.32	3.16	0.443	90.55	1.30	0.450	72.81
SHA-12	-1.42	0.637	92.86	3.20	0.659	91.94	3.88	0.474	93.96	1.52	0.524	93.55
HA-GLS-BLUP	0.25	0.665	92.45	5.54	0.698	91.36	9.60	0.567	93.99	6.72	0.589	93.63
HA-GLS-noBLUP	4.70	0.637	93.59	10.53	0.659	92.94	4.37	0.473	93.89	3.47	0.523	93.54

Table 4.15: Relative biases of the $JK1 - B$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size of 100 people and 200 random groups the Hájek estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HA	-1.71	0.779	90.39	3.53	0.825	89.08	5.31	0.769	92.67	3.74	0.767	92.42
SHA-U	-0.95	0.392	14.11	0.89	0.389	0.12	-0.52	0.385	89.53	-1.09	0.393	64.53
SHA-1	-52.66	0.435	80.62	0.91	0.411	0.72	-4.65	0.408	92.79	-23.26	0.422	89.84
SHA-2	-2.01	0.443	24.60	-41.93	0.492	82.92	1.63	0.439	90.33	-0.41	0.446	72.60
SHA-12	-42.89	0.485	83.66	-37.86	0.511	83.65	-2.03	0.461	93.08	-15.60	0.478	91.27
HA-GLS-BLUP	-28.44	0.562	87.13	-22.84	0.596	86.27	1.21	0.545	92.92	-7.80	0.547	91.97
HA-GLS-noBLUP	-39.84	0.483	84.85	-33.95	0.510	85.00	-1.52	0.460	93.16	-14.21	0.476	91.40

Table 4.16: Relative biases of the $JK2 - B$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size of 100 people and 200 random groups the Hájek estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HA	-1.81	0.778	90.39	3.36	0.824	89.08	5.23	0.769	92.67	3.63	0.767	92.42
SHA-U	-0.95	0.392	14.11	0.89	0.389	0.12	-0.52	0.385	89.53	-1.09	0.393	64.53
SHA-1	-52.66	0.435	80.62	0.91	0.411	0.72	-4.65	0.408	92.79	-23.26	0.422	89.84
SHA-2	-2.01	0.443	24.60	-41.94	0.492	82.92	1.63	0.439	90.33	-0.41	0.446	72.60
SHA-12	-42.90	0.485	83.66	-37.88	0.511	83.65	-2.03	0.461	93.08	-15.60	0.478	91.27
HA-GLS-BLUP	-28.45	0.562	87.13	-22.86	0.596	86.25	1.20	0.545	92.92	-7.81	0.547	91.97
HA-GLS-noBLUP	-39.85	0.483	84.85	-33.96	0.510	85.00	-1.53	0.460	93.16	-14.22	0.476	91.40

Table 4.17: Relative biases of the $JK1 - A$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size of 500 people and 2000 random groups the Hájek estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HA	0.24	0.367	92.67	-1.81	0.391	91.50	-0.14	0.356	93.04	1.64	0.359	93.38
SHA-U	1.27	0.176	0.00	3.06	0.175	0.00	-1.32	0.174	76.49	0.44	0.175	11.95
SHA-1	-0.03	0.280	93.22	3.88	0.182	0.00	-0.62	0.187	93.03	0.74	0.210	93.23
SHA-2	1.74	0.198	0.03	1.24	0.296	92.76	-1.13	0.199	78.42	2.43	0.199	21.71
SHA-12	0.20	0.284	93.07	1.66	0.297	92.62	-0.96	0.211	93.25	1.75	0.228	93.84
HA-GLS-BLUP	1.08	0.299	93.17	1.03	0.317	92.44	1.00	0.252	93.83	4.66	0.263	93.94
HA-GLS-noBLUP	5.32	0.284	93.78	7.22	0.298	93.51	-0.62	0.211	93.34	3.78	0.228	94.20

Table 4.18: Relative biases of the $JK2 - A$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size of 500 people and 2000 random groups the Hájek estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HA	0.22	0.367	92.67	-1.84	0.391	91.50	-0.15	0.356	93.04	1.63	0.359	93.38
SHA-U	1.27	0.176	0.00	3.06	0.175	0.00	-1.32	0.174	76.49	0.44	0.175	11.95
SHA-1	-0.03	0.280	93.22	3.88	0.182	0.00	-0.62	0.187	93.03	0.74	0.210	93.23
SHA-2	1.74	0.198	0.03	1.24	0.296	92.76	-1.13	0.199	78.42	2.43	0.199	21.71
SHA-12	0.19	0.284	93.06	1.66	0.297	92.62	-0.96	0.211	93.25	1.75	0.228	93.84
HA-GLS-BLUP	1.03	0.299	93.17	0.98	0.317	92.44	0.90	0.252	93.83	4.56	0.263	93.93
HA-GLS-noBLUP	5.11	0.284	93.76	7.02	0.298	93.49	-0.64	0.211	93.34	3.70	0.228	94.17

Table 4.19: Relative biases of the $JK1 - B$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size of 500 people and 2000 random groups the Hájek estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HA	0.24	0.367	92.67	-1.81	0.391	91.50	-0.14	0.356	93.04	1.64	0.359	93.38
SHA-U	1.27	0.176	0.00	3.06	0.175	0.00	-1.32	0.174	76.49	0.44	0.175	11.95
SHA-1	-53.84	0.190	79.63	3.35	0.182	0.00	-5.54	0.182	92.59	-22.74	0.184	89.67
SHA-2	1.35	0.198	0.03	-43.09	0.222	83.47	-1.47	0.199	78.29	2.11	0.198	21.65
SHA-12	-43.83	0.212	83.65	-39.84	0.229	84.21	-4.45	0.207	92.85	-15.25	0.208	91.40
HA-GLS-BLUP	-26.07	0.256	88.31	-24.33	0.274	88.21	-3.71	0.247	93.07	-6.66	0.248	92.32
HA-GLS-noBLUP	-40.93	0.213	84.79	-36.66	0.229	85.56	-4.12	0.208	92.89	-13.56	0.209	91.68

Table 4.20: Relative biases of the $JK2 - B$ variance estimation, average mean lengths and coverages of 95% confidence interval estimators with a sample size of 500 people and 2000 random groups the Hájek estimators.

Estimators	Variable y_1			Variable y_2			Variable y_3			Variable y_4		
	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)	RB (%)	AL	CR (%)
HA	0.22	0.367	92.67	-1.84	0.391	91.50	-0.15	0.356	93.04	1.63	0.359	93.38
SHA-U	1.27	0.176	0.00	3.06	0.175	0.00	-1.32	0.174	76.49	0.44	0.175	11.95
SHA-1	-53.84	0.190	79.63	3.35	0.182	0.00	-5.54	0.182	92.59	-22.74	0.184	89.67
SHA-2	1.35	0.198	0.03	-43.10	0.222	83.47	-1.47	0.199	78.29	2.11	0.198	21.65
SHA-12	-43.83	0.212	83.65	-39.84	0.229	84.21	-4.45	0.207	92.85	-15.25	0.208	91.40
HA-GLS-BLUP	-26.07	0.256	88.31	-24.34	0.274	88.21	-3.72	0.247	93.07	-6.66	0.248	92.32
HA-GLS-noBLUP	-40.93	0.213	84.79	-36.66	0.229	85.56	-4.12	0.208	92.89	-13.56	0.209	91.68

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this dissertation, we considered weighting adjustment methods in survey sampling. In Chapter 2, we considered a new survey estimator under nonresponse. The estimator was obtained using the estimated response propensity, which the response propensity was estimated through a nonparametric logistic model. Two variance estimations were considered: reverse approach and the two-phase sampling. The two approaches showed similar behaviors. From the simulation study, the new estimator was seen to be consistent. The nonparametric estimator had advantages as it performed significantly better for inference than the parametric estimator when the response propensity function was nonlinear. In Chapter 3, we considered a smoothed Horvitz-Thompson estimator in which the smoothed weights were obtained from the weight model. The asymptotic distribution of the estimator was derived. We found that the smooth estimator and the model-based variance estimator performed well when the model was correctly specified. When the model was misspecified, we found that the smoothed Hájek version of the weight smoothing estimator improved the efficiency of the survey estimators. Combined with jackknife variance estimation, the confidence interval performed well for inference. In Chapter 4, the results from Chapter 3 were extended to the case with a random effect weight model. We found that the GLS Hájek estimator was unbiased and the NOBLUP estimator was more efficient than the BLUP estimator. Together with the jackknife estimator, the confidence intervals provided good coverages in the simulation study.

5.2 Future work

In Chapter 3, the nonparametric logistic model was considered. For the simulation study, we picked three different smoothing parameters: $\lambda = 1$, $\lambda = 10$ and $\lambda = 200$. The choice of the smoothing parameter λ is critical to the performance of a spline estimate. The larger the smoothing

parameter, the less flexibility of the model fitting. Future work would include the choice of the smoothing parameter, which can lead to an unbiased and more efficient survey estimator under nonresponse. For the random effect weight model in Chapter 5, in the coming study, we will explore the asymptotic distribution of the estimator, the variance estimation with respect to the sampling design and the weight model. We compared several different estimators in this chapter. The selection of the model for the estimators could be a potential research topic as well.

BIBLIOGRAPHY

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

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

Beaumont, J.-F. (2008). A new approach to weighting and inference in sample surveys. *Biometrika*, 95(3):539–553.

Beaumont, J.-F. and Bocci, C. (2008). Another look at ridge calibration. *Metron*, LXVI(1):5–20.

Beaumont, J.-F., Haziza, D., and Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, 100(3):555–569.

Beaumont, J.-F. and Rivest, L.-P. (2009). Dealing with outliers in survey data. In Pfeffermann, D. and Rao, C., editors, *Handbook of Statistics, Sample Surveys, Design Methods and Applications*, volume 29A, pages 247–280. Amsterdam:North Holland.

Bethlehem, J. G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4(3):251–260.

Binder, D. and Roberts, G. (2003). *Analysis of Survey Data*, chapter Introduction to Part A, pages 22–25. Wiley, New York.

Brick, J. M. (2013). Unit Nonresponse and Weighting Adjustments: A Critical Review. *Journal of Official Statistics*, 29(3):329–353.

Cassel, C. M., Särndal, C.-E., and Wretman, J. H. (1983). Some uses of statistical models in connection with the nonresponse problem. In Madow, W. and Olkin, I., editors, *Incomplete Data In Sample Surveys*, volume 3, pages 143–160. New York: Academic Press.

Chambers, R. L. (1996). Robust case-weighting for multipurpose establishment surveys. *Journal of Official Statistics*, 12(1):3–32.

Durbin, J. (1959). A note on the application of quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, 46(3/4):477–480.

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

Elliot, M. R. and Little, R. J. (2000). Model-based alternatives to trimming survey weights. *Journal of Official Statistics*, 16(3):191–209.

Fay, R. E. (1991). A Design-Based Perspective on Missing Data Variance. In *Proceedings of Bureau of the Census Annual Research Conference*, pages 429–440.

Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. In *ASA Proceedings of the Social Statistics Section*, pages 197–202. American Statistical Association.

Fuller, W. A. and An, A. B. (1998). Regression adjustment for nonresponse. *Journal of the Indian Society of Agricultural Staitsitcs*, 51:331–342.

Gerda Claeskens, Tatyana Krivobokova, J. D. O. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, 96(3):529–544.

Giommi, A. (1984). A simple method for estimating individual response probabilities in sampling from finite 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.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Iannacchione, V. G. (2003). Sequential weight adjustment 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. In *ASA Proceedings of the Section on Survey Research Method*, pages 637–642. American Statistical Association.

Kim, J. K. (2010). Calibration estimation using exponential tilting in sample surveys. *Survey Methodology*, 36(2):145–155.

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. (2001). The delete-a-group jackknife. *Journal of Official Statistics*, 17(4):521–526.

Lundström, S. and Särndal, C.-E. (1999). Calibration as a Standard Method for Treatment of Nonresponse. *Journal of Official Statistics*, 15(2):305–327.

Martinoz, C. F., Haziza, D., and Beaumont, J.-F. (2015). A method of determining the winsorization threshold, with an application to domain estimation. *Survey Methodology*, 41(1):57–77.

Nargundkar, M. and Joshi, G. B. (1975). Non-response in sample surveys. In *40th session of the ISI, Warsaw 1975, Contributed papers*, pages 626–628.

Pfeffermann, D. and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya: The India Journal of Statistics*, 61(1):166–186.

Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In *JSM proceedings of the Section, American Staitstical Association, AMSTAT*, volume 225230.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society*, 11(1):68–84.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4):353–360.

Särndal, C.-E. and Lundström, S. (2006). *Estimation in Surveys with Nonresponse*. Wiley, New York.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer.

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of American Statistical Association*, 94(445):254–265.

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

Wand, M. P. and Ormerod, J. T. (2008). On semiparametric regression with O'Sullivan penalised splines. *Australian and New Zealand Journal of Statistics*, 50:179–198.