USE OF RANDOM PERMUTATION MODEL
IN RATE ESTIMATION AND STANDARDIZATION

A Dissertation Presented
by
WENJUN LI

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

February 2003

Public Health
Biostatistics and Epidemiology
© Copyright by Wenjun Li 2003

All Rights Reserved
USE OF RANDOM PERMUTATION MODEL
IN RATE ESTIMATION AND STANDARDIZATION

A Dissertation Presented
by
WENJUN LI

Approved as to style and content by:

____________________________________
Edward J. Stanek III, Chair

____________________________________
Carol Biglow, Member

____________________________________
John P. Buonaccorsi, Member

Philip C. Nasca, Department Chair
Biostatistics and Epidemiology
DEDICATION

To my parents Li Beihua and Wu Dehua,

and my wife Lu Lu
ACKNOWLEDGMENTS

I would like to express appreciation on Professors Edward J. Stanek III, John P. Buonaccorsi and Carol Biglow. I am deeply grateful to Professor Stanek, without his encouragement and guidance, this dissertation research would not be completed successfully. I thank Drs David Hosmer and John Griffith for their helpful comments.
ABSTRACT

USE OF RANDOM PERMUTATION MODEL IN RATE ESTIMATION AND STANDARDIZATION

FEBRUARY 2003

WENJUN LI, B.S., FUDAN UNIVERSITY
M.S., UNIVERSITY OF MASSACHUSETTS AMHERST
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Edward J. Stanek, III

Through integrating techniques from several areas in survey sampling, we develop an alternative method of deriving estimators using random permutation models under (stratified) simple random sampling without replacement. The finite random permutation model links the samples to the population. The joint permutation of response and auxiliary variables is modeled using seemingly unrelated regression. We use prediction theory from the super-population sampling literature to derive the linear unbiased minimum variance predictors of population means under the design-based framework using the finite estimating equation approach of Binder and Patak (1994). The predictors have functional forms similar to those derived using design-based, model-assisted and calibration approaches, but depend on neither superpopulation nor regression model assumptions. We applied the results to standardization of multiple rates, and illustrate how our methods account for the covariance of the standardized rates, unlike conventional standardization methods.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Motivations</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Rate estimation and adjustment in epidemiology</td>
<td>2</td>
</tr>
<tr>
<td>1.3. Three major approaches in survey sampling</td>
<td>7</td>
</tr>
<tr>
<td>1.4. Calibration estimation</td>
<td>11</td>
</tr>
<tr>
<td>1.5. Direct adjustment and poststratification</td>
<td>12</td>
</tr>
<tr>
<td>1.6. Sampling framework</td>
<td>13</td>
</tr>
<tr>
<td>1.7. Seemingly unrelated regression models</td>
<td>15</td>
</tr>
<tr>
<td>1.8. Functional form approach in survey sampling</td>
<td>16</td>
</tr>
<tr>
<td>1.9. Optimization methods</td>
<td>18</td>
</tr>
<tr>
<td>2. NOTATIONS, DEFINITIONS AND BASICS OF RANDOM PERMUTATION MODELS</td>
<td>20</td>
</tr>
<tr>
<td>2.1. Notation</td>
<td>20</td>
</tr>
<tr>
<td>2.2. Definition of population and population parameters</td>
<td>22</td>
</tr>
<tr>
<td>2.3. Single stage random permutation</td>
<td>23</td>
</tr>
<tr>
<td>2.4. Simultaneous permutation of multiple variables</td>
<td>24</td>
</tr>
<tr>
<td>2.5. Sampling and partition of random matrices</td>
<td>26</td>
</tr>
<tr>
<td>2.6. Derivation of variance-covariance matrix of $\text{vec}(U)$</td>
<td>27</td>
</tr>
<tr>
<td>2.7. Derivation of variance-covariance matrix in Section 2.4</td>
<td>28</td>
</tr>
<tr>
<td>2.8. Derivation of variance-covariance matrix of the partitioned matrix</td>
<td>32</td>
</tr>
<tr>
<td>3. GENERAL ESTIMATION STRATEGY</td>
<td>35</td>
</tr>
<tr>
<td>3.1. Parameters of interest</td>
<td>37</td>
</tr>
<tr>
<td>3.2. Random permutation seemingly unrelated regression model</td>
<td>37</td>
</tr>
<tr>
<td>3.3. Simultaneous estimation of $p + 1$ population totals</td>
<td>39</td>
</tr>
<tr>
<td>3.4. Example: best linear unbiased estimator of a population total</td>
<td>41</td>
</tr>
<tr>
<td>3.5. Estimation of a population total using auxiliary information</td>
<td>42</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>3.6</td>
<td>Enumeration and simulation</td>
</tr>
<tr>
<td>3.7</td>
<td>Research objectives</td>
</tr>
<tr>
<td>4</td>
<td>ESTIMATION OF POPULATION TOTALS</td>
</tr>
<tr>
<td>4.1</td>
<td>Motivations</td>
</tr>
<tr>
<td>4.2</td>
<td>Parameters of interest</td>
</tr>
<tr>
<td>4.3</td>
<td>Formulation of a RPSUR model</td>
</tr>
<tr>
<td>4.4</td>
<td>Two classes of estimators for population total vector</td>
</tr>
<tr>
<td>4.5</td>
<td>Estimating multiple population totals</td>
</tr>
<tr>
<td>4.6</td>
<td>Remarks</td>
</tr>
<tr>
<td>4.7</td>
<td>Derivations of results in Section 4.5.1</td>
</tr>
<tr>
<td>4.8</td>
<td>Derivations of results in Section 4.5.2</td>
</tr>
<tr>
<td>4.9</td>
<td>Derivations of results in Section 4.5.3</td>
</tr>
<tr>
<td>4.10</td>
<td>Derivations of results in Section 4.5.4</td>
</tr>
<tr>
<td>5</td>
<td>ESTIMATING POPULATION TOTALS BY REPARAMETERIZATION</td>
</tr>
<tr>
<td>5.1</td>
<td>Motivation</td>
</tr>
<tr>
<td>5.2</td>
<td>Formulation of the model through reparameterization</td>
</tr>
<tr>
<td>5.3</td>
<td>Estimating $T^{(0)}$ and $\mu^{(0)}$ with a known auxiliary population total</td>
</tr>
<tr>
<td>5.4</td>
<td>Estimating $T^{(0)}$ and $\mu^{(0)}$ with several known auxiliary totals</td>
</tr>
<tr>
<td>5.5</td>
<td>Estimation of variance and confidence intervals</td>
</tr>
<tr>
<td>5.6</td>
<td>Remarks</td>
</tr>
<tr>
<td>5.7</td>
<td>Derivations of results in Section 5.3</td>
</tr>
<tr>
<td>5.8</td>
<td>Derivations of results in Section 5.4</td>
</tr>
<tr>
<td>6</td>
<td>ESTIMATORS BASED ON RANDOM PERMUTATION MODELS UNDER STRATIFIED SIMPLE RANDOM SAMPLING</td>
</tr>
<tr>
<td>6.1</td>
<td>Motivation</td>
</tr>
<tr>
<td>6.2</td>
<td>Definition and notations</td>
</tr>
<tr>
<td>6.3</td>
<td>Independent permutation of one response variable across $H$ strata</td>
</tr>
<tr>
<td>6.4</td>
<td>Simultaneous permutations of one response variable and $p$ auxiliary variables across $H$ strata</td>
</tr>
<tr>
<td>6.5</td>
<td>Stratified random sampling and partition of random matrices</td>
</tr>
<tr>
<td>6.6</td>
<td>Parameter of interest</td>
</tr>
<tr>
<td>6.7</td>
<td>Formulation of the model when auxiliary variables are present</td>
</tr>
<tr>
<td>6.8</td>
<td>Estimating $T^{(0)}$ when one auxiliary variable is present</td>
</tr>
<tr>
<td>6.9</td>
<td>Derivations of results in Section 6.8.4</td>
</tr>
<tr>
<td>6.10</td>
<td>Derivations of results in Section 6.8.5</td>
</tr>
</tbody>
</table>
7. APPLICATIONS TO RATE ESTIMATION .......................................................... 111

7.1. Motivations ........................................................................................................ 111
7.2. Cases with single categorical auxiliary variable ........................................ 111
7.3. Cases with two categorical variables .............................................................. 125
7.4. Cases with single continuous auxiliary variable ........................................ 128
7.5. Derivations of results (7.1) and (7.2) in Section 7.2 ....................... 146
7.6. Derivations of results (7.3) and (7.4) in Section 7.2 .......................... 150

8. DIRECT RATE STANDARDIZATION USING RANDOM PERMUTATION
   MODELS .......................................................................................................... 154

8.1. Motivations ........................................................................................................ 154
8.2. Rate estimation with internal direct adjustment ........................................ 154
8.3. Rate comparison with internal direct adjustment ...................................... 160
8.4. External direct adjustment ........................................................................... 162
8.5. Discussion ....................................................................................................... 163
8.6. Derivations of the results in Section 8.2 ................................................ 164

9. CONCLUDING REMARKS .............................................................................. 172

BIBLIOGRAPHY ..................................................................................................... 177
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1. Two classes of unbiased estimators of population totals under two optimization criteria</td>
<td>53</td>
</tr>
<tr>
<td>7.1. Mean squared error (x 10,000) of four estimators by population size and sample size</td>
<td>116</td>
</tr>
<tr>
<td>7.2. Coverage rates and width of 95% confidence intervals</td>
<td>117</td>
</tr>
<tr>
<td>7.3. Percentage of left- and right-misses of the 95% confidence intervals, by population size, sample size and estimation methods</td>
<td>118</td>
</tr>
<tr>
<td>7.4. Two categorical auxiliary variable examples</td>
<td>125</td>
</tr>
<tr>
<td>7.5. Parameters of hypothetical population Series A and B</td>
<td>133</td>
</tr>
<tr>
<td>7.6. MSE of design-based estimators adjusted to known continuous or categorical auxiliary variables, by population size and sample size (Series A)</td>
<td>134</td>
</tr>
<tr>
<td>7.7. MSE of design-based estimators adjusted to known continuous or categorical auxiliary variables, by population size and sample size (Series B)</td>
<td>135</td>
</tr>
<tr>
<td>7.8. Probability of encountering negative variance when sample size is small and age is used as a continuous or categorical auxiliary variable</td>
<td>142</td>
</tr>
<tr>
<td>7.9. Coverage rates and width of 95% confidence interval using age as a continuous or categorical variable, by population size and sample size (Series A)</td>
<td>144</td>
</tr>
<tr>
<td>7.10. Coverage rates and width of 95% confidence interval using age as a continuous or categorical variable, by population and sample sizes (Series B)</td>
<td>145</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1. General Estimation Strategy</td>
<td>36</td>
</tr>
<tr>
<td>7.1. Mean squared error by population size and sample size</td>
<td>119</td>
</tr>
<tr>
<td>7.2. MSE reduction (%) of a design-based predictor (DPS or DPP) by population and sample sizes</td>
<td>120</td>
</tr>
<tr>
<td>7.3. MSE Ratio of design-based predictors to SRS estimator by population and sample sizes</td>
<td>121</td>
</tr>
<tr>
<td>7.4. Hypothetical relationship between smoking and age of population series A and B</td>
<td>132</td>
</tr>
<tr>
<td>7.5. MSE and variance of design-based predictors adjusted for continuous or categorical age (Series A)</td>
<td>137</td>
</tr>
<tr>
<td>7.6. MSE and variance ratios of estimators adjusted for age median, quartiles, octiles and sedeciles to estimators adjusted for age mean (Series A)</td>
<td>138</td>
</tr>
<tr>
<td>7.7. MSE and variance of design-based predictors adjusted for continuous or categorical age (Series B)</td>
<td>140</td>
</tr>
<tr>
<td>7.8. MSE and variance ratios of estimators adjusted for age median, quartiles, octiles and sedeciles to estimators adjusted for age mean (Series B)</td>
<td>141</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Estimation and comparison of disease prevalence based on survey samples is one of the most common procedures in epidemiological investigation. Analysis of survey data often involves “adjustment” or “control” for covariates. A methodological issue that arises is how to adjust for the imbalance of covariates in a sample or to remove “confounding”. For example, a sample may have a very different gender ratio than that of the parent population. This issue is closely related to the theory of survey sampling, and application of recently advanced methods for addressing this issue may improve methodologies in Epidemiological surveys.

1.1. Motivations

Rate estimation and standardization in epidemiology and vital statistics, in particular, is closely related to calibration, poststratification and generalized regression estimation (GREG) in survey sampling. Under certain assumptions, directly standardized rates can be viewed as estimators “calibrated” to known population or subpopulation totals, or estimators based on poststratification, or GREG estimators based on group mean models. These methods of estimation have experienced significant theoretical and practical advances in the past twenty years. Application of these techniques may lead to a new class of estimators that will improve rate estimation and standardization procedures commonly used in epidemiology and vital statistics.

This dissertation research develops alternative methods of defining and estimating population proportions (means) based on a functional form approach (Binder
and Patak 1994; Thompson 1997) and calibration methods (Deville and Särndal 1992; Särndal, Swensson and Wretman 1992; Brewer 1995; Estevao and Särndal 2000). Seemingly unrelated regression and finite random permutation model frameworks (Stanek, Singer and Lencina 2003) are used as vehicles for mathematical deviation. Research focuses on developing estimators for population totals and proportions in the setting of simple random sampling without replacement (SRSWOR) and stratified simple random sampling without replacement (STSRS).

In the following sections, I will briefly review current methods of rate standardization and advances in several areas of survey sampling that are directly relevant to the estimation methods proposed in this dissertation. This will lay a ground for theoretical development of later Chapters.

1.2. Rate estimation and adjustment in epidemiology

Various methods of adjusting for the imbalance of covariates resulting from complex sampling have been developed in the epidemiologic literature (Rothman 1986; Breslow and Day 1987; Rosenbaum 1987; Esteve, Benhamou and Raymond 1994). When applied to rate estimation, these methods are used to “control” or “adjust” for confounding, and produce a single index that is “comparable” between populations or subpopulations. For example, in epidemiological research, disease morbidity or mortality experience (hereafter, referred to more generally as prevalence) is often compared across geographical areas that have different distribution with respect to the control variables (e.g., age). Direct comparison of crude prevalence rates may be severely confounded by the difference in the population structure of the areas being compared when the difference is related to the outcome of interest (e.g., mortality).
“Standardization” or “adjustment” needs to be made to remove such effects of
“confounding”. A typical example is the gender- and age-adjusted mortality rates in
vital statistics.

Adjustment for confounding can be made with or without a multiple regression
model (Kahn and Sempos 1989). Methods not based on regression models are
commonly referred to as “standardization” in literature. Briefly, a standardized rate is a
weighted average of the stratum specific prevalence rates in the study populations,
where weights are determined by the relative frequencies of the control variable in the
referred population. Standardization methods can be grossly classified into “direct” or
“indirect” standardization according to whether a same weighting system is applied to
different populations (Kahn and Sempos 1989). We will discuss these methods in more
detail in the subsequent sections.

Adjustment can be also made based on regression models (Kahn and Sempos
1989). For example, a categorical auxiliary variable can be used for poststratification; a
continuous variable strongly correlated with the response variable can be used to
improve the estimation using ratio or regression estimation. When all the covariates in
the regression model are categorical, they are analogous to directly standardized rates
but include the possibility of adjusting for many variables at one time (Feldstein 1966;
can be included in the model in such cases, and it leads to generalized regression
estimators (GREG) and calibration estimators.
1.2.1. Rate standardization and finite sampling

Direct standardization of rates can be viewed as special cases of finite sampling. A fundamental assumption is that the populations under comparison arise from the same parent population or are similar with respect to factors associated with the event under study (Fleiss 1981), which is often implied by the “null hypothesis” (Neison 1844). The choice of a “standard population” is, in statistical term, to define the “parent population” under study. In addition, the populations under comparison are considered as realizations of random samples from that parent population. Under such assumptions, directly standardized rates may be considered as special cases of calibration estimators or generalized regression (GREG) estimators. The weight systems for standardization are “calibrated” to the known population or group quantities of the covariates, and the standardized rates are estimated (predicted) rates of the “parent population” based on these samples.

Direct standardization of rates, under certain assumptions, is also related to the class of estimators commonly referred to as “poststratified estimators” (Thomsen 1978; Oh and Scheuren 1983; Särndal, Swensson and Wretman 1992). A presumption in this methodology is that random sampling may result in an imbalance of auxiliary variables but not the probability distribution of the study variable given the auxiliary values (Mukhopadhyay 2000).

1.2.2. Direct standardization

The purpose of direct adjustment is to present an estimate of event prevalence that characterizes a population if the distribution of the confounding variables were the same as that in the standard population. The adjusted totals or rates are then
“comparable” between populations. There are two major approaches, i.e., internal and external comparisons. The former involves only the populations under comparison, the later involves an outside reference population with a known population structure. The choice of a “standard population” can make a substantial difference in the standardized rates and may have significant impact on the interpretation of study results. This weakness has renewed interest in methodological research on the appropriateness and validity of rate standardization to describe the discrepancy in event prevalence among different populations (CDC 1999; Choi, de Guia and Walsh 1999; Goldman and Brender 2000; Julious, Nicholl and George 2001).

Once the standard population (SP) is chosen, direct standardization methods utilize the distribution of the confounders in this population to define weights that are used in the weighted average of prevalence estimates for all populations under comparison. To illustrate this method, we consider the following simple case. Suppose that two populations, indexed as $i = 1, 2$, are under comparison. Assume each population has $J$ age strata that are indexed as $j = 1, 2, \ldots, J$ and that age is a confounder of the study outcome. The Directly Standardized Rate (DSR) for population $i$ is defined as

$$DSR_i = \sum_{j=1}^{J} \frac{N_j}{N_{sp}} r_{ij},$$

where $r_{ij}$ is the rate in the $j$-th stratum of the $i$-th population, $N_j/N_{sp}$ is the relative frequency of stratum $j$ in the standard population. Using DSRs, a Directly Standardized Rate Ratio (DSRR) can be defined as
Predicted overall rate of SP with stratum-specific rates of population $i$ over Actual overall rate of SP

\[
\text{DSRR}_i = \frac{\sum_{j=1}^J N_{ij} r_j}{\sum_{j=1}^J N_{ij} R_j} = \sum_{j=1}^J w_j r_j,
\]

where $w_j = N_j \left( \sum_{j=1}^J N_j R_j \right)^{-1}$. Notice that the weights ($w_j$) are fixed and independent of the populations under comparison (or not dependent on samples); therefore, the weight system is “identical” for all populations under comparison. DSRR is commonly referred as the Comparative Mortality Figure (CMF) or the Standardized Rate Ratio (SRR) in the literature. Since we are interested in not only mortality rates but also other prevalence rates, we use a more general term DSRR.

1.2.3. Indirect standardization, ISR and ISRR

Indirect standardization utilizes as its weighting system the distribution of the confounder in the comparison (e.g., “exposed”) population. Thus, a separate “Indirectly Standardized Rate Ratio” (ISRR) is produced for each population being compared. ISRR is commonly referred as “Standardized Mortality Ratio” (SMR) in vital statistics. The indirectly standardized rate for population $i$ is defined as

\[
\text{ISR}_i = \frac{\text{Observed overall rate in population } i}{\text{Predicted overall rate in population } i \times \text{[SP overall rate]}}
\]

using stratum-specific rates of SP

\[
= \frac{\sum_{j=1}^J N_{ij} r_j / \sum_{j=1}^J N_{ij}}{\sum_{j=1}^J N_j R_j / \sum_{j=1}^J N_{ij}} \times R
\]

\[
= R \sum_{j=1}^J v_j r_j
\]

where $R = \sum_{j=1}^J N_j R_j / \sum_{j=1}^J N_j$ is the overall rate of the standard population, a known constant; $v_j = N_{ij} / \sum_{j=1}^J N_j R_j$, is dependent on the size ($N_{ij}$) of group $j$ in population $i$, therefore each population under comparison will have different weighting factor
unless all corresponding group relative frequency distribution of the control variable cross populations are equal.

The ISRR of population \( i \) is a rate ratio defined as

\[
\text{ISRR}_i = \frac{\text{ISR}_i}{R} = \frac{\sum_{j=1}^{J} N_{y,j}r_{y,j}}{\sum_{j=1}^{J} N_{y,j}R_j} = \sum_{j=1}^{J} v_{y,j}r_{y,j}.
\]

ISRR is the most commonly used technique and has been used to inform allocation of resources and health-related police-making (Julious, Nicholl and George 2001).

1.2.4. Other standardization methods

There are many other standardization methods in literature (Inskip et al. 1983; Julious, Nicholl and George 2001). For example, Lee and Liaw proposed a new weighting system to compute directly standardized rates, where the set of weights are determined by the data rather than specified \textit{a priori}, and optimized for stability while ensuring the comparability of the rates between populations (Lee and Liaw 1999). This method requires an assumption that the event counts follow a Poisson distribution, and achieves its optimality by minimizing the total variance of the age-adjustment indices in the logarithmic scale. Based on such optimized weights, (Lee 2002) proposed a new standardized summary measure - “harmonically weighted ratio” that can be applied to external and internal comparisons if the rate-ratio homogeneity assumption holds, i.e., rate-ratios of different groups are equal.

1.3. Three major approaches in survey sampling

Current sample survey methods include three major approaches: 1) design-based, 2) model-based and 3) model-assisted approaches. What distinguishes them is the choice of probability model that is used for statistical inference.
The design-based approach uses probability sampling (randomization) for both sample selection and inference from sample data. The probability distribution associated with the randomized sample selection provides the basis for probabilistic inferences about the target population. The bias, variance and mean squared error (MSE) are defined in terms of the expectation over all possible samples under the sampling design. This approach leads to valid repeated sampling inference regardless of the population structure, and is free from “model misspecification”. The design-based framework established by (Horvitz and Thompson 1952) and (Godambe 1955) is a typical example. For example, the population total may be written as \( T_y = \mathbf{1}_N' \mathbf{y}_{N \times 1} \), and its Horvitz-Thompson estimator is \( \hat{T}_{HT} = \mathbf{1}_n' \mathbf{\Pi}_s^{-1} \mathbf{y}_s \), where \( \mathbf{y}_s \) is an \( n \)-vector of the sample values of \( y \); \( \mathbf{\Pi}_s \) and \( \mathbf{\Pi}_s \) are \( N \times N \) and \( n \times n \) diagonal matrices of inclusion probabilities, respectively, and \( \mathbf{1}_n' \) and \( \mathbf{1}_N' \) indicate row vectors with all elements equal to one. When auxiliary variables are available, the Horvitz-Thompson (HT) estimator can be written as

\[
\hat{T}_{HTR} (\mathbf{y}) = \frac{\hat{T}_{HT} (\mathbf{y})}{\hat{T}_{HT} (\mathbf{x})} T (\mathbf{x})
\]

where \( T (\mathbf{x}) \) and \( \hat{T}_{HT} (\mathbf{x}) \) are population total of \( x \) and its HT estimator, respectively (Brewer 1963). This estimator is asymptotically design-unbiased and design-consistent, i.e., \( E[\hat{T}_{HTR} (y) - T(y)] \to 0 \) and \( Pr[|T_{HTR} (y) - T(y)| > \varepsilon] \to 0 \) when both sample and population sizes become large \( (n \to \infty, N \to \infty) \) (Brewer 1979; Särndal and Wright 1984).
The *Model- or prediction-based approach* assumes that the target population follows a specified model (distribution), and the model distribution yields valid inferences conditioning on the particular sample that has been drawn (Rao 1997; Brewer 1999). The bias, variance and mean squared error (MSE) are defined in terms of the expectation over all possible realizations of a stochastic model that connects the variable of interest to a set of auxiliary variables (Brewer 1995). Model-dependent methods can perform poorly as evaluated by the expected value of the model based MSE over the design in large samples if the model is misspecified (Hansen, Madow and Tepping 1983; Rao 1997). Model-based approaches have application when handling nonsampling errors, such as measurement errors and non-response (Brewer 1963; Royall 1970; Brewer 1995). In small-area estimation, model-based estimators can have smaller MSE than design based estimators. The best linear unbiased predictor (BLUP) is a typical model-based estimator (Ghosh and Rao 1994; Rao 1997).

The *model-assisted approach* is a hybrid of design-based and model-based approaches. Plausible population models are often used to choose designs and efficient design-consistent estimators, but the inference remains design-based. This means that probability statements refer to the set of all possible samples that can be drawn with the given probability sampling design. In the model-assisted approach, models may be used to construct estimators, but randomization must be used to select the sample, and statistical properties of estimators are computed with respect to the probability sampling distribution (Särndal, Swensson and Wretman 1992; Rao 1999). The resulting estimator is model-unbiased under the assumed model as well as design consistent (Särndal, Swensson and Wretman 1992; Rao 1999). The model-assisted approach provides a
formal framework for using auxiliary information at the estimation stage. The popular
generalized regression estimator (GREG) is an example of this type (Cassel, Särndal
and Wretman 1977; Särndal, Swensson and Wretman 1992). For example, the general
form of GREG estimator of the population total is

\[
\hat{\beta}_{GREG} = \hat{\beta}_{HT} (y) + \left[ T(\mathbf{X}) - T_{HT} (\mathbf{X}) \right] \hat{\beta}_{GREG}
\] (1.1)

where \( \mathbf{X} \) is an \( N \times p \) matrix of auxiliary variables, \( T(\mathbf{X}) = \mathbf{1}'_n \mathbf{X} \) is the \( 1 \times p \) row vector
of its column sums, \( \hat{T}_{HT} (\mathbf{X}) = \mathbf{1}'_s \mathbf{\Pi}^{-1}_s \mathbf{X}_s \) is the Horvitz-Thompson estimators of \( T(\mathbf{X}) \),
\( \mathbf{X}_s \) is the \( n \times p \) matrix of known auxiliary variables of sampled subjects in a
hypothesized linear model linking \( y \) and \( \mathbf{X} \), e.g., \( y = \mathbf{X}\beta + \varepsilon \), where \( E(\varepsilon) = 0 \),
\( E(\varepsilon \varepsilon') = \sigma^2 \mathbf{A} \), \( \mathbf{A} \) is a \( N \times N \) diagonal matrix. Under this model, the estimator of the
\( p \)-vector regression parameter \( \beta \) is

\[
\hat{\beta}_{GREG} = \left( \mathbf{X}_s^{-1} \mathbf{A}_s^{-1} \mathbf{\Pi}^{-1}_s \mathbf{X}_s \right)^{-1} \mathbf{X}_s^{-1} \mathbf{A}_s^{-1} \mathbf{\Pi}^{-1}_s \mathbf{y}_s.
\] (1.2)

This estimator is design-consistent and asymptotically design-unbiased. It is also an
estimator “calibrated” against all the auxiliary variables in \( \mathbf{X} \) (Deville and Särndal

However, model-assisted approaches have also their limitations. When the
model is correctly specified, model-assisted inferences are inferior to those from model-
based approach (Rao 1997); when the model is misspecified, they are inferior to the
unconditional designed based approach (Rao 1997).
1.4. **Calibration estimation**

Calibration, or cosmetic calibration (Särndal and Wright 1984; Brewer 1999), is another popular method of incorporating known auxiliary information to improve estimation. This approach does not necessarily need a model at all (Thompson 1997). The strategy is to identify a set of sample-dependent (or sometimes called adjusted) weights, such that when they are applied to auxiliary variables, the “estimator” of the auxiliary quantities will equate the known corresponding auxiliary quantities (Deville and Särndal 1992; Särndal, Swensson and Wretman 1992). The rational behind calibration methods is that a weight system should be good for a response variable if it is good for a strongly correlated auxiliary variable (Deville and Särndal 1992; Särndal, Swensson and Wretman 1992; Brewer 1999).

The idea can be illustrated using the following simple example. Suppose we are interested in estimating the prevalence of teenager smoking in a population of size $N$, and the total number of male and female teenagers are known. It is believed that male teenagers are much more likely to be smokers than their female peers. A SRSWOR sample of teenagers of a predetermined size $(n)$ is drawn from the population. Without using any auxiliary variable, one would use a simple expansion weight $a = \frac{N}{n} \mathbf{1}_n$ to estimate the prevalence. Using the calibration method, one can find a set of sample-dependent weights ($w$) by minimizing the squared distance between $w$ and $a$ subject to a constraint such that when $w$ is applied to “estimating” the total number of males or females, the result must be equal to the known numbers of males and females. In numerous publications, calibration estimators have been shown to be more efficient.
than the sample expansion estimators (Deville and Särndal 1992; Särndal, Swensson and Wretman 1992).

1.5. Direct adjustment and poststratification

In the statistical literature, direct adjustment and poststratification are often referred to as the same estimation procedure (Cochran 1977; Holt and Smith 1979; Little 1982; Oh and Scheuren 1983; Rosenbaum 1987). While “direct adjustment” in epidemiology does not necessarily mean the same as in statistical literature, “direct adjustment” in epidemiology can be viewed as a special case of poststratification in survey sampling if the “standard population” is viewed as the parent population from which the populations under comparison arise.

Poststratification has been discussed intensively in both model-based and model-assisted approaches (Royall 1971; Holt and Smith 1979; Rosenbaum 1987; Smith 1990; Särndal, Swensson and Wretman 1992). Using a model based approach and assuming a distribution function of the response variable $y$ and the auxiliary variable $x$, i.e., under a superpopulation model, the post-stratified estimator of mean (proportion or rate) is shown to be a maximum likelihood estimator of the population mean (Jagers 1986; Mukhopadhyay 2000). Rosenbaum (1987) showed that sample mean and poststratified or direct adjusted estimators are special cases of model-based direct adjustment under two different extreme models for the subclass-specific selection probabilities. The model-based poststratification approach is also used to improve small-area estimations, by forming “synthetic estimators” (Holt, Smith and Tomberlin 1979).
From the perspective of a model-assisted approach, “poststratified estimators” are a subclass of GREG estimators related to population groups (Särndal, Swensson and Wretman 1992; Rao 1999). The underlying model is a group mean model, where group membership is identified after sampling for sampled subjects, and the population group sizes may or may not be known (Särndal, Swensson and Wretman 1992; Rao 1999). Poststratification is particularly useful in multipurpose surveys because of its flexibility in forming poststrata as compared to stratified simple random sampling.

Poststratification is a common tool for reducing nonresponse bias for survey data (Oh and Scheuren 1983; Smith 1990; Zhang 1999). When non-response probabilities are independent of the study outcome variable, poststratification can efficiently reduce non-response bias.

1.6. Sampling framework

Making inference from samples drawn from a population is fundamental to epidemiology surveys and statistics. In order to make inference, a probabilistic “link” is needed to relate a sample to its parent population. This dissertation will adopt random permutation models to provide such necessary “links” (Stanek, Singer and Lencina 2003). Permutation models were discussed more comprehensively in the earlier work of Cassel, Särndal and Wretman (1977) and Rao and Bellhouse (1978). This framework, similar to conventional design-based inference, does not require assumptions about parametric distributions. The permuted population consists of random variables corresponding to permutations of subjects in the finite population, where permutations occur with equal probabilities. The population is represented as a non-stochastic vector of fixed elements, whereas the permuted population is a similar vector of random
variables. Parameters in the population consist of the population values, and functions of these parameters. Parameters of the permuted population are defined as the expected value of functions of permuted population of random variables. The randomness is thus attributable to random sampling (permutation). Inference can be made through evaluating the connection of parameters in the population with parameters in the model for the permuted population. Under this framework, the structural relationship between the permuted population and the population can be explicitly represented using matrices. This provides a convenient mathematical vehicle when regression models are involved.

Briefly, random variables and a stochastic model are defined to represent the random permutation of a finite real-valued population. Suppose that a finite population consists of \( s = 1, 2, \ldots, N \) subjects labeled by \( s \), and associated with each subject is an unknown but potentially observable value, \( y_s \). The population values are represented with a \( N \times 1 \) column vector, namely, \( y = \left( y_1 \ y_2 \ \cdots \ y_N \right)' \). Define a set of indicator random variables \( U_i \) for \( i = 1, 2, \ldots, N \) that have a value of 1 if the \( i^{th} \) subject in a permutation is subject \( s \), and 0 otherwise. Let the matrix

\[
U_{N \times N} = \begin{pmatrix} U_1 & U_2 & \cdots & U_N \end{pmatrix}
\]

represent a matrix of indicator random variables, where

\[
U_s = \begin{pmatrix} U_{1s} & U_{2s} & \cdots & U_{Ns} \end{pmatrix}'
\]. Then the vector of random variables representing a permutation is given by

\[
Y_{N \times 1} = Uy
\],

(1.3)
where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)^\prime$, and $i$ indicates the position in a permutation. The random vector $\mathbf{Y}$ is a vector of linear combinations of random variables in the matrix $\mathbf{U}$ because the vector of population values $\mathbf{y}$ is constant. It is obvious that observing the random vector $\mathbf{Y}$ is equivalent to observing the random indicator matrix $\mathbf{U}$. Since any two subjects can not be at the same position in a permutation, it is easy to show that

$$E(U_{is}U_{is'}) = \begin{cases} 
1/N & \text{when } i \neq i' \text{ and } s = s', \\
1/N(N-1) & \text{when } i \neq i' \text{ and } s \neq s', \\
0 & \text{otherwise}.
\end{cases}$$

It immediately follows,

$$E(\mathbf{Y}) = \mu_y \mathbf{1}_N,$$

$$\text{cov}(\mathbf{Y}) = \sigma^2_y \mathbf{P}_N,$$

where $\mu_y = N^{-1}\sum_{s=1}^{N} y_s$, $\sigma^2_y = (N-1)^{-1} \mathbf{y}' \mathbf{P}_N \mathbf{y}$, and $\mathbf{P}_N = \mathbf{I}_N - N^{-1} \mathbf{J}_N$.

1.7. **Seemingly unrelated regression models**

Theories about seemingly unrelated regression models (SUR) are well established. Finite sample properties have been discussed in a few occasions in literature (Zellner 1963; Revankar 1974; 1976; Liu and Wang 1999), but these models generally have a strong assumption about the error structure, typically in a form of that assumes error vectors are *independently and identically distributed*, and multivariate normal $\mathbf{MN}(\mathbf{0}, \Sigma)$. SUR models enable simultaneous estimation or testing of multiple
parameters across equations in the model. Applications of SUR in epidemiological surveys are quite rare.

Theoretical considerations on integrating SUR models, survey sampling theory, and random permutation models, have not been seen in literature. Combining these three aspects may lead to alternative estimation methods for survey samples. The feature of SUR of preserving covariance structures between covariates during the estimation process is particularly useful for evaluating the properties of estimators based on permutation models. With a SUR set-up, permuting response and auxiliary variables jointly can be represented easily in matrix form, and the covariance structures can be preserved during estimation.

1.8. Functional form approach in survey sampling

Many linear or non-linear finite population parameters under complex sampling designs can be expressed as a smooth function of population totals, and their point estimators are similar functions of the estimators of the totals (Thompson 1997). For example, finite population parameters or functions, such as means, ratios and variances, can be defined implicitly by a population equation of the form $\sum_{s=1}^{N} \phi_s (y_s, x_s, \theta_N) = 0$, where $y_s$ and $x_s$ are values of observable variables, $\phi_s$ are known real-valued functions, and $\theta_N$ is a population parameter of interest. For example, the population ratio $R$ of $y$ to $x$ can be defined by $\sum_{s=1}^{N} (y_s - \theta_N x_s) = 0$, that is $\theta_N = g(T) = T_y / T_x$, where $T_y$ and $T_x$ are the total of $y$ and $x$, respectively. An estimators can be defined as a solution of
the sample estimating equation $\phi_s(\theta_N) = \sum_{s=1}^{N_s} \phi_s(y_s, x_s, \theta_N)/\pi_s = 0$. For any $\theta_N$, the sample estimating function $\phi_s$ is design-consistent if $E\{\phi_s(\theta_N)\} = \sum_{s=1}^{N_s} \phi_s(y_s, x_s, \theta_N)$.

Linearization of the error as a function of the component total estimates provides a fundamental basis for this approach. The properties of the point estimate $\hat{\theta} = g(\hat{T})$ are evaluated using a Taylor series approximation to the error. The first-order approximation of this type is a “linearization” of the error, for example, the linearized error of population ratio is

$$\epsilon = R \hat{\phi}_s = R = g\left(T\right) - g\left(T\right) = \frac{\partial g}{\partial T_y} \left(\hat{R}_s - R\right) + \frac{\partial g}{\partial T_x} \left(T_s - T_x\right) + \text{remainder}.$$

Many estimators based on functional approach are definable through systems of estimating equations rather than single estimating equation (Thompson 1997). Some population parameters, such as variances or regression coefficients, can be defined in terms of other population parameters and need a system of estimating functions for their definition. For instance, the population variance $\sigma^2_y = \sum_{s=1}^{N} (y_s - \mu_y)^2 / N$ is defined by

$$\left\{ \sum_{s=1}^{N} \left[ (y_s - \mu_y)^2 - \sigma^2_N \right] = 0, \sum_{s=1}^{N} (y_s - \mu_y) = 0 \right\}.$$

Since SURs allow several variables be modeled simultaneously, their joint distributions can be obtained. A combination of the functional approach and SUR models may provide insight in estimation problems for sample surveys.
1.9. **Optimization methods**

Optimization criteria and methods may vary significantly according to the nature of problems under consideration. The parameter estimation can be viewed as a least square problem, which minimizes a specified distance function, say $Φ(w)$, subject to a set of constraints. Commonly used distance functions include trace, generalized mean squared error (GMSE) or other quadratic function of the variance-covariance matrix of the estimators.

Often a set of linear constraints are present, e.g., $r(w) = 0$, where $r(w)$ is a vector of differentiable functions. Available optimization methods under linear constraints include two major approaches, i.e., 1) reparameterization, and 2) modification of $Φ(w)$ by introducing Lagrangian multipliers (Judge et al. 1985).

When linear equality constraints can be expressed in the form $w^* = g(w)$, the restrictions can be introduced by reducing the dimension of the parameter space to circumvent the singularity problem and thus a unique solution can be obtained. The constrained optimum of the objective function $Φ(w)$ can be found by an unconstrained minimization of $Φ_Δ(w^*) = Φ(g(w))$. For example, if $w = (w_1, w_2, w_3)'$ is subject to $1'w = 0$, then the constraint can be represented as a reparameterization of $w$,

$$g(w) = \begin{pmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix} w = \begin{pmatrix} w_1 - w_3 \\ w_2 - w_3 \end{pmatrix} = w^*$$

Minimization of $Φ_Δ(w)$ subject to $1'w = 0$ is equivalent to minimization of $Φ(w^*)$ without constraint.
When multiple constraints are present, it may be difficult to reparameterize the optimization functions. Instead, a vector \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_j) \)' of Lagrangian multipliers can be introduced and the constraints are incorporated in the minimization procedure by defining an alternative optimization function

\[
\hat{\Phi}_R (\mathbf{w}, \lambda) = \Phi (\mathbf{w}) + \mathbf{r}(\mathbf{w}) \lambda
\]

(1.4)

A constrained minimum of \( \Phi(\mathbf{w}) \) is equivalent to minimizing \( \hat{\Phi}_R (\mathbf{w}, \lambda) \).

Differentiating \( \hat{\Phi}_R (\mathbf{w}, \lambda) \) with respect to elements of both \( \mathbf{w} \) and \( \lambda \), and setting the derivatives to zeros, will lead to unique solutions for \( \hat{\mathbf{w}} \).
CHAPTER 2

NOTATIONS, DEFINITIONS AND BASICS OF
RANDOM PERMUTATION MODELS

This Chapter sets up general notations used in this thesis (Section 2.1), defines parameters of interest (Section 2.2), introduces the random permutation framework (Section 2.3) and methods for simultaneously permuting multiple variables, and establishes their relationship to seemingly unrelated regression models (Section 2.4). It is shown that the sampling process can be represented as a partition of random matrices (Section 2.5).

2.1. Notation

Random variables are represented with upper case Latin letters, for example, $Y$, and random vectors with bolded upper case letter, e.g., $Y$. The value of a random variable is represented with lower case letter, e.g., $y$; and a column vector of the values with bolded lower case letter, e.g., $y$.

Population parameters are represented using Greek letters such as $\mu$ for mean, $\Sigma$ for variance-covariance matrix. Estimators of population parameters are denoted using a hat over the letter of their corresponding Greek letters, for example, $\hat{\mu}$ and $\hat{\Sigma}$. Letter “$T$” is reserved for population totals, such as $T^{(i)}$ for the population total of a response variable $Y^{(i)}$.

Matrices are represented using bolded upper or lower case letters. To simplify notation, we define notation to represent matrices that are frequently used in
mathematical derivation (Harville 1997). A vector or matrix of zeros is represented with 

\textbf{0}. For any finite integer \( a \) and \( b \),

1) \textit{Summing Vector} \( \mathbf{1}_a' \) is a \( 1 \times a \) row vector of ones,

2) \textit{Identity Matrix} \( \mathbf{I}_a \) is an \( a \times a \) identity matrix,

3) \textit{J-Matrix} \( \mathbf{J}_a \) is an \( a \times a \) matrix of ones,

4) \textit{Centering Matrix} \( \mathbf{P}_a = \mathbf{I}_a - \frac{1}{a} \mathbf{J}_a \),

5) \textit{Projection Matrix} \( \mathbf{P}_{XW} = \mathbf{X}(\mathbf{X}'\mathbf{WX})^{-1}\mathbf{X}'\mathbf{W} \), where \( \mathbf{X} \) is any \( a \times b \) matrix,

\( \mathbf{W} \) is any \( a \times a \) matrix and \((\mathbf{X}'\mathbf{WX})^{-1}\) indicates a generalized inverse of 

\( \mathbf{X}'\mathbf{WX} \). When \( \mathbf{W} = \mathbf{I} \), \( \mathbf{P}_{X1} \) is simply denoted as \( \mathbf{P}_X \).

6) Another frequently used matrix is \( \mathbf{P}_{a,N} = \mathbf{I}_n - \frac{1}{N} \mathbf{J}_n \).

For matrix operations, the transpose of a matrix is represented as \( \mathbf{A}' \), and an inverse of a matrix \( \mathbf{A} \) is denoted as \( \mathbf{A}^{-1} \) if it exists. The generalized inverse of \( \mathbf{A} \) is denoted as \( \mathbf{A}^{-} \). A block diagonal matrix of \( \mathbf{A}_i, i = 1,2,\ldots,N \), is represented using

\[
\bigoplus_{i=1}^{N} \mathbf{A}_i.
\]

Kronecker products of \( \mathbf{A} \) and \( \mathbf{B} \) are represented using \( \mathbf{A} \otimes \mathbf{B} \),

\[
\mathbf{A}_{M \times N} \otimes \mathbf{B}_{K \times L} = \\
\begin{bmatrix}
    a_{11} \mathbf{B} & a_{12} \mathbf{B} & \cdots & a_{1N} \mathbf{B} \\
    a_{21} \mathbf{B} & a_{22} \mathbf{B} & \cdots & a_{2N} \mathbf{B} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{M1} \mathbf{B} & a_{M2} \mathbf{B} & \cdots & a_{MN} \mathbf{B}
\end{bmatrix}.
\]

The VEC operator is denoted as \( \text{vec} \), for example,
\[
\text{vec} \left[ A_{m \times n} \right] = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix},
\]

where \( a_i, i = 1, 2, \ldots, n \), is the \( i \)-th column of the \( m \times n \) matrix \( A \).

2.2. Definition of population and population parameters

A finite population \( \mathcal{P} \) consists of \( N \) subjects with labels \( s = 1, 2, \ldots, N \), where \( N \) is known. The labels are non-informative and serve only to identify the units (Royall 1970). Associated with each subject (e.g., \( s \)) is a non-stochastic \( (p+1) \times 1 \) column vector \( y_s \), where \( y_s = \left( y_s^{(0)} \mid y_s^{(1)} \mid \cdots \mid y_s^{(k)} \mid \cdots \mid y_s^{(p)} \right)' \). \( y_s^{(0)} \) is the response variable under study and \( y_s^{(k)}, k = 1, 2, \ldots, p \) are auxiliary variables.

The values of the \( k \)-th variable \( y_s^{(k)} \), \( k = 0, 1, \ldots, p \), for all subjects are represented using an \( N \times 1 \) column vector \( y^{(k)} = \left( y_1^{(k)} \mid \cdots \mid y_N^{(k)} \right)' \). The values of all \( p+1 \) variables for all \( N \) subjects can be represented with an \( N \times (p+1) \) matrix \( y \), such that

\[
y_{N \times (p+1)} = \begin{pmatrix} y_0' \\ y_1' \\ \vdots \\ y_N' \end{pmatrix} = \left( y^{(0)} \mid y^{(1)} \mid \cdots \mid y^{(k)} \mid \cdots \mid y^{(p)} \right).
\]

Further, the population values are alternatively represented as a \( (p+1)N \times 1 \) column vector \( z \), such that \( z = \text{vec}(y) \).
For any variable \( y^{(k)} \), \( k = 0,1,2,\ldots,p \), we define population parameters as follows,

Mean:  
\[
\mu^{(k)} = \frac{1}{N} \mathbf{1}_N' \mathbf{y}^{(k)}
\]

Total:  
\[
T^{(k)} = \mathbf{1}_N' \mathbf{y}^{(k)}
\]

Ratio of \( y^{(k)} \) to \( y^{(k')} \):  
\[
R_{k,k'} = \frac{T^{(k)}}{T^{(k')}}
\]

Variance:  
\[
\left( \frac{N-1}{N} \right) \sigma_k^2 = \frac{1}{N} \mathbf{y}^{(k)y} \mathbf{P}_N \mathbf{y}^{(k)}
\]

Covariance between \( y^{(k)} \) and \( y^{(k')} \):  
\[
\left( \frac{N-1}{N} \right) \sigma_{k,k'} = \frac{1}{N} \mathbf{y}^{(k)y} \mathbf{P}_N \mathbf{y}^{(k')}
\]

Variance-covariance matrix:
\[
\Sigma_{(p+1)(p+1)} = \begin{pmatrix}
\sigma_0^2 & \sigma_{01} & \cdots & \sigma_{0p} \\
\sigma_{10} & \sigma_1^2 & \cdots & \sigma_{1p} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{p0} & \sigma_{p1} & \cdots & \sigma_p^2
\end{pmatrix}
\]

2.3. Single stage random permutation

Suppose that a sample of \( n \) subjects is drawn from a population of size \( N \) via sample random sampling without replacement (SRSWOR). A stochastic model can be defined as arising from a single-stage random permutation of the population. In a permutation, the position of a subject is indexed by \( i = 1,2,\ldots,N \); there will be a different subject, \( s \), at each position, \( i \). The subjects at the first \( i = 1,2,\ldots,n \) positions in a permutation constitute a sample from SRSWOR.
We define a set of random variables and a stochastic model to represent the random permutation of the subjects in $\mathcal{P}$. First, we define a set of indicator random variables $U_i$, $i = 1, 2, \ldots, N$, that have a value of 1 if the subject in the $i$-th position in a permutation is subject $s$, and 0 otherwise. Let $U_{N \times N} = (U'_1, U'_2, \ldots, U'_{N, s})'$ represent a matrix of indicator random variables, where $U'_{i, s} = (U_{i1}, U_{i2}, \ldots, U_{iN})'$. Then the vector of random variables representing a permutation of $y^{(k)}$ is given by

$$Y^{(k)} = Uy^{(k)}.$$  

For a random permutation, $E(U) = \frac{1}{N} J_N$. The variance of $Y^{(k)}$ can be expressed in terms of the variance and covariance of the elements in $U$. Through proper arrangement in order, the variance for $U$ can be written as

$$\text{var} \left[ \text{vec}(U) \right] = \frac{1}{N-1} P_N \otimes P_N. \quad (2.1)$$

Derivation of (2.1) is presented in Section 2.6.

Since $Y^{(k)} = Uy^{(k)} = \left(y^{(k)} \otimes I_N \right) \text{vec}(U)$, we have,

$$E(Y^{(k)}) = I_n \mu_k,$$

$$\text{var}(Y^{(k)}) = \sigma_k^2 P_N.$$  

2.4. Simultaneous permutation of multiple variables

We develop expressions for simultaneous permutation of multiple variables, first considering the joint permutation of two variables. The results can be readily
extended to scenarios with multiple variables. The joint permutation of both $y^{(0)}$ and $y^{(l)}$ can be expressed as

$${Y^{(0)} \choose Y^{(l)}} = (I_2 \otimes U) {y^{(0)} \choose y^{(l)}},$$

(2.2)

Since $E(U) = J_N / N$, it is easy to show that

$${E} {Y^{(0)} \choose Y^{(l)}} = (I_2 \otimes 1_N) {\mu_0 \choose \mu_1},$$

(2.3)

and

$${cov} {Y^{(0)} \choose Y^{(l)}} = \Sigma_{0,1} \otimes P_N,$$

(2.4)

where $\Sigma_{0,1} = \begin{pmatrix} \sigma_0^2 & \sigma_{0,1} \\ \sigma_{1,0} & \sigma_1^2 \end{pmatrix}$. Derivation of (2.4) is given in Section 2.7.

When there are $p + 1$ covariates of interest, the matrix of random variables representing a permutation is given by

$$Y_{N \times (p+1)} = U_{N \times N} Y_{N \times (p+1)},$$

(2.5)

where $Y = {Y^{(0)} \choose Y^{(l)} \cdots Y^{(k)} \cdots Y^{(p)}}$, $Y^{(k)}$ is the random variable corresponding to variable $y^{(k)}$, $k = 0,1,\ldots,p$. Denote a column vector $Z_{(p+1)N \times 1}$ as

$$Z = vec(Y) = (I_{p+1} \otimes U) vec(y) = (I_{p+1} \otimes U) z,$$

(2.6)

where $z_{(p+1)N \times 1} = vec(y)$. 

25
As an application of (2.3) and (2.4) when $\rho = 1$, it can be shown that the expected value and variance-covariance matrix of $Z$ are,

$$E(Z) = (I_{p+1} \otimes I_N)\mu,$$  \hfill (2.7)

$$cov(Z) = \Sigma \otimes P_v,$$  \hfill (2.8)

where $\Sigma$ is as defined in Section 2.2.

2.5. Sampling and partition of random matrices

A sample of size $n$, $n \leq N$, drawn by SRSWOR scheme from $\mathcal{P}$ is thus viewed as the first $n$ subjects in a random permutation. Therefore, a random vector $Z$ can be partitioned into a sample (indexed as $I$) and remaining (indexed as $II$) parts. The elements of $Z$ can be rearranged into the sampled and remaining portions through pre-multiplication by a properly specified positioning matrix $K^*$, i.e.,

$$Z^* = K^*Z = \begin{pmatrix} Z_I \\ Z_{II} \end{pmatrix}.$$

For example, if $Z$ involves 2 variables,

$$K^* = \begin{pmatrix} I_2 \otimes \begin{pmatrix} I_n & 0 \end{pmatrix}_{n \times (N-n)} \\ I_2 \otimes \begin{pmatrix} 0 & I_{N-n} \end{pmatrix}_{(N-n) \times n} \end{pmatrix}.$$

Similarly, if $Z$ involves $p + 1$ variables,

$$K^* = \begin{pmatrix} I_{p+1} \otimes \begin{pmatrix} I_n & 0 \end{pmatrix}_{n \times (N-n)} \\ I_{p+1} \otimes \begin{pmatrix} 0 & I_{N-n} \end{pmatrix}_{(N-n) \times n} \end{pmatrix}. \hfill (2.9)$$

The expected values of $Z^*$ are,
Use the notation, we can derive the expression for the partitioned variance-covariance matrix as follows. The partitioned variance-covariance matrix is defined by

\[ \mathbf{V} = \text{cov} \left( \frac{\mathbf{Z}_I}{\mathbf{Z}_{II}} \right) = \begin{pmatrix} \mathbf{V}_I & \mathbf{V}_{I,II} \\ \mathbf{V}_{II,I} & \mathbf{V}_{II} \end{pmatrix}, \]  

(2.10)

where \( \mathbf{V}_I = \mathbf{\Sigma} \otimes \mathbf{P}_{n,N} \), \( \mathbf{V}_{I,II} = \mathbf{\Sigma} \otimes \left( -\frac{1}{N} \mathbf{J}_{n(n-n)} \right) \) and \( \mathbf{V}_{II} = \mathbf{\Sigma} \otimes \mathbf{P}_{(N-n),N} \).

Furthermore, it can be shown that

\[ \mathbf{V}_I^{-1} = \mathbf{\Sigma}^{-1} \otimes \left( \mathbf{I}_n + \frac{1}{N-n} \mathbf{J}_n \right). \]  

(2.11)

Detailed derivations for the above expressions can be found in Section 2.8.

### 2.6. Derivation of variance-covariance matrix of \( \text{vec}(\mathbf{U}) \)

Since \( \mathbf{U}_{N \times N} = (\mathbf{U}_1', \mathbf{U}_2', \ldots, \mathbf{U}_N')' \) represent a matrix of indicator random variables, where \( \mathbf{U}_i = (U_{i1}, U_{i2}, \ldots, U_{iN})' \),

\[ \text{cov} \left[ \text{vec} (\mathbf{U}) \right] = \begin{pmatrix} \text{cov} (\mathbf{U}_1, \mathbf{U}_1) & \text{cov} (\mathbf{U}_1, \mathbf{U}_2) & \cdots & \text{cov} (\mathbf{U}_1, \mathbf{U}_N) \\ \text{cov} (\mathbf{U}_2, \mathbf{U}_1) & \text{cov} (\mathbf{U}_2, \mathbf{U}_2) & \cdots & \text{cov} (\mathbf{U}_2, \mathbf{U}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov} (\mathbf{U}_N, \mathbf{U}_1) & \text{cov} (\mathbf{U}_N, \mathbf{U}_2) & \cdots & \text{cov} (\mathbf{U}_N, \mathbf{U}_N) \end{pmatrix}. \]  

(2.12)

Since \( \text{cov} (\mathbf{U}_i, \mathbf{U}_j') = E(\mathbf{U}_i \mathbf{U}_j') - E(\mathbf{U}_i)E(\mathbf{U}_j'), \) \( E(\mathbf{U}_i) = E(\mathbf{U}_j') = \frac{1}{N}, \) and
\[
E\left(U_i, U'_i\right) = \begin{cases}
\frac{1}{N} I_N, & \text{when } i = i^*, \\
-\frac{1}{N(N-1)}(I_N - J_N), & \text{when } i \neq i^*,
\end{cases}
\]

It is readily shown that

\[
\text{cov}\left(U_i, U'_i\right) = E\left(U_i, U'_i\right) - E\left(U_i\right)E\left(U'_i\right) = \begin{cases}
\frac{1}{N} P_N, & i = i^* \\
-\frac{1}{N(N-1)} P_N & i \neq i^*
\end{cases}, \tag{2.13}
\]

where \(P_N = I_N - \frac{J_N}{N}\). Applying (2.13) to (2.12), we have

\[
\text{cov}\left[\text{vec}(U)\right] = -\frac{1}{N-1} P_N \otimes P_N.
\]

### 2.7. Derivation of variance-covariance matrix in Section 2.4.

Simultaneous permutation of two variables can be expressed as

\[
\begin{pmatrix}
Y^{(0)}_1 \\
Y^{(1)}_1
\end{pmatrix} = (I_2 \otimes U) \begin{pmatrix}
y^{(0)}_1 \\
y^{(1)}_1
\end{pmatrix} \tag{2.14}
\]

In order to get the variance-covariance matrix, we can rewrite (2.14) as (2.15),

\[
\begin{pmatrix}
Y^{(0)}_1 \\
Y^{(1)}_1
\end{pmatrix} = \left(\begin{pmatrix}
y^{(0)}_1 \\
y^{(1)}_1
\end{pmatrix} \otimes I_{2N}\right) \text{vec} \left(I_2 \otimes U\right). \tag{2.15}
\]

Since \(\text{vec} \left(A_{mn} \otimes B_{pq}\right) = (I_p \otimes K_{qm} \otimes I_p) \left[\text{vec} \left(A_{mn}\right) \otimes \text{vec} \left(B_{pq}\right)\right]\), where \(K_{qm}\) is a vec-permutation matrix. The \(\left[(i-1)n + j\right]th\) row of \(K_{qm}\) is the \(\left[(j-1)m + i\right]th\) row of \(I_{mn}\) \((i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)\) (see Harville 1997, page 345, 349). Therefore,
\[
\begin{align*}
\text{vec}(I_2 \otimes U) &= (I_2 \otimes K_{N_2} \otimes I_N) \left[ \text{vec}(I_2) \otimes \text{vec}(U) \right] \\
&= (I_2 \otimes K_{N_2} \otimes I_N) \begin{pmatrix} \text{vec}(U) \\ 0 \\ 0 \end{pmatrix} \\
&= (I_2 \otimes K_{N_2} \otimes I_N) U^*,
\end{align*}
\]

where \( U^* = \left( \text{vec}(U)' \mid 0 \mid 0 \mid \text{vec}(U)' \right)' \).

Apply (2.16) to (2.15), we have

\[
\begin{pmatrix} Y^{(0)} \\ Y^{(i)} \end{pmatrix} = \begin{pmatrix} \left( y^{(ip)} \right) \otimes I_2 \otimes I_N \end{pmatrix} (I_2 \otimes K_{N_2} \otimes I_N) U^* \\
= \begin{pmatrix} \left[ \left( \left( y^{(ip)} \right) \otimes I_2 \right) (I_2 \otimes K_{N_2}) \right] \otimes I_N \end{pmatrix} U^*,
\]

where \( K_{N_2} \) is a \( 2N \times 2N \) matrix, comprising \( N \) rows and 2 columns of \( N \times 2 \) dimensional blocks whose \( ij \)-th element is 1 and whose other \((2N-1)\) elements are 0.

\( K_{N_2} \) can be written explicitly as

\[
K_{N_2} = \begin{pmatrix} \begin{pmatrix} \mathbf{u}_1 \mathbf{e}'_1 \\ \mathbf{u}_2 \mathbf{e}'_1 \\ \vdots \\ \mathbf{u}_N \mathbf{e}'_1 \\ \mathbf{u}_1 \mathbf{e}'_2 \\ \mathbf{u}_2 \mathbf{e}'_2 \\ \vdots \\ \mathbf{u}_N \mathbf{e}'_2 \\ \vdots \\ \mathbf{u}_1 \mathbf{e}'_N \\ \mathbf{u}_2 \mathbf{e}'_N \\ \vdots \\ \mathbf{u}_N \mathbf{e}'_N \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix},
\]

where \( \mathbf{u}_i \) is the \( i \)-th column of \( I_2 \), \( i = 1, 2 \), and \( \mathbf{e}_j \) is the \( j \)-th column of \( I_N \), \( j = 1, 2, \ldots, N \). Observing that
\[
\begin{align*}
(y^{(0)r} \otimes I_2) &= \begin{pmatrix}
y^{(0)}_1 & 0 \\
y^{(0)}_2 & 0 \\
\vdots & \vdots \\
y^{(0)}_N & 0 \\
\end{pmatrix} \\
&= \begin{pmatrix}
y^{(0)r} u_e c_1' \\
y^{(0)r} u_e c_2' \\
\vdots \\
y^{(0)r} u_e c_N' \\
\end{pmatrix} \\
&= \begin{pmatrix}
y^{(0)r} u_e c_1' \\
y^{(0)r} u_e c_2' \\
\vdots \\
y^{(0)r} u_e c_N' \\
\end{pmatrix} \\
&= \begin{pmatrix}
y^{(0)r} 0 \\
0 \\
\vdots \\
0 \\
\end{pmatrix} \\
&= \begin{pmatrix}
e_1 u'_i \\
e_2 u'_i \\
\vdots \\
e_N u'_i \\
\end{pmatrix} \\
&= (I_2 \otimes y^{(0)r}) K'_{N2}
\end{align*}
\]

and \(K'_{N2} K_{N2} = I_{N2}\), we can simplify (2.17) as follows (for matrix operation, see page 348 & p349 of Harville 1997),

\[
\begin{align*}
\left(\begin{array}{c}
Y^{(0)} \\
Y^{(1)}
\end{array}\right) &= \left((I_2 \otimes y^{(0)r}) K_{N2} K'_{N2} \otimes I_N \right) U^* \\
&= \left((I_2 \otimes y^{(0)r}) K'_{N2} K_{N2} \otimes I_N \right) U^* \\
&= \left((I_2 \otimes y^{(0)r}) I_{N2} \otimes I_N \right) U^* \\
&= \left((I_2 \otimes y^{(0)r}) \otimes I_N \right) U^* \\
&= \left(y^{(0)r} 0 \quad y^{(1)r} 0 \quad 0 \quad y^{(0)r} \right) U^* \\
&= \left(I_2 \otimes y^{(0)r} \quad I_2 \otimes y^{(1)r} \otimes I_N \right) U^* \\
\text{cov}\left(\begin{array}{c}
Y^{(0)} \\
Y^{(1)}
\end{array}\right) &= \left((I_2 \otimes y^{(0)r}) \otimes I_N \right) \text{cov}(U^*) \left((I_2 \otimes y^{(1)r}) \otimes I_N \right)
\end{align*}
\]

As shown in Section 2.6, the variance matrix for \(U\) is

\[
\text{cov}[\text{vec}(U)] = \frac{1}{N-1} P_N \otimes P_N,
\]

where \(P_N = I_N - \frac{J_N}{N}\). Therefore,
$$\text{cov}(U^*) = \text{cov}
abla \left( \begin{array}{c}
\text{vec}(U) \\
0 \\
0 \\
\text{vec}(U)
\end{array} \right) = \begin{bmatrix}
\text{cov}(\text{vec}(U)) & 0 & 0 & \text{cov}(\text{vec}(U), \text{vec}(U)) \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\text{cov}(\text{vec}(U), \text{vec}(U)) & 0 & 0 & \text{cov}(\text{vec}(U))
\end{bmatrix}$$

$$\text{cov}(U^*) = \begin{bmatrix}
1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1
\end{bmatrix} \otimes \text{cov}(\text{vec}(U))$$

$$= \frac{1}{N-1} (LL') \otimes (P_N \otimes P_N) \quad (2.18)$$

where $L' = (1 \ 0 \ 0 \ 1)$.

This is a patterned matrix, we can further simplify it as follows.

$$\text{cov}
\begin{bmatrix}
Y^{(0)} \\
Y^{(1)}
\end{bmatrix} = \frac{1}{N-1} \begin{bmatrix}
(I_2 \otimes y^{(0)r}) & (I_2 \otimes y^{(1)r}) \otimes I_N
\end{bmatrix}
\begin{bmatrix}
(\text{LL'}) \otimes P_N & \otimes P_N & (I_2 \otimes y^{(0)}) & (I_2 \otimes y^{(1)}) \otimes I_N
\end{bmatrix}$$

$$= \frac{1}{N-1} \begin{bmatrix}
y^{(0)r} & 0 & y^{(1)r} \\
0 & y^{(0)r} & 0 \\
P_N & 0 & P_N
\end{bmatrix}
\begin{bmatrix}
P_N & 0 & 0 & P_N \\
0 & y^{(0)} & 0 & y^{(1)} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix} \otimes P_N$$

$$= \frac{1}{N-1} \begin{bmatrix}
y^{(0)r}P_Ny^{(0)} \\
y^{(1)r}P_Ny^{(1)}
\end{bmatrix} \otimes P_N$$

$$= \begin{bmatrix}
\sigma_0^2 & \sigma_{0,1} \\
\sigma_{0,1} & \sigma_1^2
\end{bmatrix} \otimes P_N$$

$$= \Sigma_{0,1} \otimes P_N$$
where \( \Sigma_{0,1} = \begin{pmatrix} \sigma_0^2 & \sigma_{0,1} \\ \sigma_{0,1} & \sigma_1^2 \end{pmatrix} \).

### 2.8. Derivation of variance-covariance matrix of the partitioned matrix

\( Y = \begin{pmatrix} Y_{i}^{(0)} & Y_{i}^{(1)} & \cdots & Y_{i}^{(r)} \end{pmatrix}' \) can be partitioned into a sample (indexed as \( I \)) and remainder (indexed as \( II \)) parts. After sampling is completed, the values for the sampled portion \( Y_{i}^{(0)} \), \( Y_{i}^{(1)} \), \( \cdots \), \( Y_{i}^{(r)} \) will be known, and can be represented as realized values.

Through proper specification of permutation matrices \( K \), we can rearrange the elements of \( Y_{i}^{(0)} \), \( Y_{i}^{(1)} \), \( \cdots \), \( Y_{i}^{(r)} \) into the sampled and remaining portions by pre-multiplication by \( K^* \). We can rewrite

\[
K^* \begin{pmatrix} Y_{i}^{(0)} \\ Y_{i}^{(1)} \\ \vdots \\ Y_{i}^{(r)} \end{pmatrix} = \begin{pmatrix} I_{p+1} \otimes (I_n | 0_{n(N-n)}) \\ I_{p+1} \otimes (0_{(N-n)N} | I_{N-n}) \end{pmatrix} \begin{pmatrix} Y_{i}^{(0)} \\ Y_{i}^{(1)} \\ \vdots \\ Y_{i}^{(r)} \end{pmatrix} = \begin{pmatrix} Z_I \\ Z_{II} \end{pmatrix}, \tag{2.19}
\]

where \( K^* \) as defined in (2.9), \( Z_I = \begin{pmatrix} Y_{i}^{(0)} & Y_{i}^{(1)} & \cdots & Y_{i}^{(r)} \end{pmatrix}' \) and

\[
Z_{II} = \begin{pmatrix} Y_{ii}^{(0)} & Y_{ii}^{(1)} & \cdots & Y_{ii}^{(r)} \end{pmatrix}'.
\]

Use this notation, we can derive the expression for the partitioned variance-covariance matrix as follows. The partitioned variance-covariance matrix is thus

\[
V = \text{cov} \begin{pmatrix} Z_I \\ Z_{II} \end{pmatrix} = \begin{pmatrix} V_I & V_{I,II} \\ V_{II,I} & V_{II} \end{pmatrix} = K^* \text{cov} (Y) K^{**}
\]

Since \( \text{cov} (Y) = \Sigma \otimes P_N \), the above expression can be simplified as,

\[
V = K^* (\Sigma \otimes P_N) K^{**}. \tag{2.20}
\]
V can be simplified as follows,

\[
V = \left( I_{p+1} \otimes \left( I_n \left| \begin{array}{c} 0_{ns(N-n)} \end{array} \right. \right) \right) \left( \Sigma \otimes P_N \right) \left( I_{p+1} \otimes \left( I_n \left| \begin{array}{c} 0_{(N-n)} \end{array} \right. \right) \right) \left( I_{p+1} \otimes \left( 0_{n(N-n)} \left| I_{N-n} \right. \right) \right)
\]

\[
= \left( \Sigma \otimes \left( I_n \left| \begin{array}{c} 0_{n(N-n)} \end{array} \right. \right) \right) P_N \left( I_n \left| \begin{array}{c} 0_{(N-n)} \end{array} \right. \right) \left( \Sigma \otimes \left( I_n \left| \begin{array}{c} 0_{n(N-n)} \end{array} \right. \right) \right) \left( I_{N-n} \left| 0_{n(N-n)} \right. \right) \right) P_N \left( 0_{n(N-n)} \left| I_{N-n} \right. \right).\]

Therefore,

\[
V = \begin{bmatrix} V_I & V_{II} \\ V_{II} & V_{II} \end{bmatrix} = \begin{bmatrix} \Sigma \otimes P_{n,N} & \Sigma \otimes \left( -\frac{1}{N} J_{n,s(N-n)} \right) \\ \Sigma \otimes \left( -\frac{1}{N} J_{n,s(N-n)} \right) & \Sigma \otimes P_{n-N,N} \end{bmatrix}, \quad (2.21)
\]

where \( V_I = \Sigma \otimes P_{n,N} \), \( V_{II} = \Sigma \otimes \left( -\frac{1}{N} J_{n,s(N-n)} \right) \) and \( V_{II} = \Sigma \otimes P_{n-N,N} \).

Further, \( P_{n,N} = I_n - \frac{1}{N} J_n = \left( I_n - \frac{J_n}{n} \right) + \left( \frac{J_n}{n} - \frac{J_n}{N} \right) = P_n + (1-f) \frac{J_n}{n} \), and

\( P_{n-N,N} = I_{N-n} - \frac{1}{N} J_{N-n} = P_{N-n} + f \frac{J_{N-n}}{N-n} \), where \( f = n/N \) is the sampling fraction.

Consequently,

\[
V = \begin{bmatrix} \Sigma \otimes \left( P_n + (1-f) n^{-1} J_n \right) & \Sigma \otimes \left( -N^{-1} J_{n,s(N-n)} \right) \\ \Sigma \otimes \left( -N^{-1} J_{n,s(N-n)} \right) & \Sigma \otimes \left( P_{n-N} + f \left( N-n \right)^{-1} J_{N-n} \right) \end{bmatrix}
\]

\[
= \begin{bmatrix} \Sigma \otimes P_n & 0 \\ 0 & \Sigma \otimes P_{n-N} \end{bmatrix} + \begin{bmatrix} \Sigma \otimes \left( (1-f) n^{-1} J_n \right) & \Sigma \otimes \left( -N^{-1} J_{n,s(N-n)} \right) \\ \Sigma \otimes \left( -N^{-1} J_{n,s(N-n)} \right) & \Sigma \otimes \left( f \left( N-n \right)^{-1} J_{N-n} \right) \end{bmatrix},
\]

and
\[ V_I = \Sigma \otimes \left( P_n + \left(1 - f\right) \frac{J_n}{n} \right), \]

\[ V_{I,II} = -\frac{1}{N} \Sigma \otimes J_{n \times (N-n)}, \]

\[ V_{II} = \Sigma \otimes \left( P_{N-n} + f \frac{J_{N-n}}{N-n} \right). \]

Furthermore, it is shown that,

\[ V_I^{-1} = \Sigma^{-1} \otimes P_{n,N}^{-1} = \Sigma^{-1} \otimes \left( I_n + \frac{1}{N-n} J_n \right) \]

and

\[ V_{II}^{-1} = \Sigma^{-1} \otimes P_{N-N,n,N}^{-1} = \Sigma^{-1} \otimes \left( I_{N-n} + \frac{1}{n} J_{N-n} \right). \]
CHAPTER 3

GENERAL ESTIMATION STRATEGY

This Chapter proposes a general method for estimating population parameters based on survey samples. This method connects permutation models, estimation functions and seemingly unrelated regression. In later Chapters, we will apply this method to develop various estimators for scenarios that are common in practical settings. The basic strategy for mathematical derivation is to:

1) Represent the randomness arising from sampling and relate samples to the population using a random permutation model;

2) Define target parameters as linear or nonlinear functions of population totals;

3) Represent joint permutations of response and auxiliary variables in a seemingly unrelated regression framework;

4) Represent estimation of population totals as predictors of the remainder based on linear functions of the sample;

5) Define unbiasedness constraints as parameter-wise or average unbiasedness;

6) Introduce auxiliary information as additional linear constraints to estimating equations using reparameterization or additional Lagrangian functions;

7) Optimize estimation by minimizing the generalized mean squared errors, or trace of variance-covariance matrix of the population totals subject to various unbiasedness or calibration constraints.

Since many population parameters can be defined as functions of population totals, estimating population totals is fundamental to the estimation of other parameters.
The following sections outline several strategies to obtain unbiased and minimum variance estimators for population totals. We will limit our discussion to the settings of simple random sampling without replacement.

Figure 3.1. General Estimation Strategy

[Diagram of General Estimation Strategy]

- Finite Population
- Parameters represented as functions of totals
- SRSWOR
- Random Permutation
- Sample
- Response variable (Measured w/o error)
- Auxiliary variables
- RPSUR Model
- Define Target Random Variables
- Sample + Remainder
- Criteria:
  1. GMSE, 2. Trace
  3. Penalty function
- Constraints:
  1. Parameter-wise
  2. Average unbiasedness
  3. Auxiliary information
- Methods:
  1. Reparameterization
  2. Lagrangian function
- Optimized estimators
3.1. Parameters of interest

In Chapter 2, we have shown that the joint distribution of \( Y^{(0)}, Y^{(1)}, \ldots, Y^{(p)} \) can be obtained through simultaneous permutation of the underlying variables \( (y^{(0)}, y^{(1)}, \ldots, y^{(p)}) \). A \((p + 1) \times 1\) vector \( T \) of population totals of interest can be represented as a set of \( p + 1 \) linear combinations of the random variables arising from the permutations,

\[
T = L'Z, \quad \text{where } L' \text{ is a } (p + 1) \times (p + 1)N \text{ matrix, } Z = \begin{pmatrix} Y^{(0)}r & Y^{(1)}r & \cdots & Y^{(p)}r \end{pmatrix}'.
\]

The elements of \( Z \) can be partitioned into two parts that represent the sample \( (Z_I) \) and the remaining \( (Z_{II}) \) by premultiplication by the permutation matrix \( K^* \) given by (2.9),

\[
K^*Z = \begin{pmatrix} Z_I' & Z_{II}' \end{pmatrix}'.
\]

Since \( K^*K^* = I \), \( T \) can be represented using following identities,

\[
T = L'Z = L'K''K^*Z = L'_IZ_I + L'_{II}Z_{II},
\]

where \( L_I \) and \( L_{II} \) are \((p + 1)n \times q\) and \((p + 1)(N - n) \times (p + 1)\) matrices respectively. Once \( Z_I \) is observed after sampling, estimating \( T \) is equivalent to predicting \( L'_{II}Z_{II} \).

3.2. Random permutation seemingly unrelated regression model

Based on the explicit presentation of simultaneous permutation of multiple variables from Section 2.4, a \((p + 1)\)-equation simple location seemingly unrelated regression model can be defined as
\[
\begin{pmatrix}
Y^{(0)} \\
Y^{(1)} \\
\vdots \\
Y^{(p)}
\end{pmatrix} =
\begin{pmatrix}
1_N & 0 & \cdots & 0 \\
0 & 1_N & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1_N
\end{pmatrix}
\begin{pmatrix}
\mu_0 \\
\mu_1 \\
\vdots \\
\mu_p
\end{pmatrix} + E,
\]

where \( E \) is a \((p + 1)N \times 1\) column vector of residuals. Since \( Y^{(0)}, Y^{(1)}, \ldots, Y^{(p)} \) arise from exactly the same permutation of \( y^{(0)}, y^{(1)}, \ldots, y^{(p)} \) as \( U \), these \( p + 1 \) equations are indeed related to each other and this relationship is captured through the identical permutation represented by \( \text{vec}(U) \). Following conventional notation, Model (3.2) can be expressed more compactly as,

\[
Z = X\mu + E
\]

where \( \mu_{(p+1)\times 1} = (\mu_0, \mu_1, \ldots, \mu_p)' \), \( X = I_{p+1} \otimes I_N \) is the design matrix for the model.

Hereafter, Model (3.3) is referred as Random Permutation Seemingly Unrelated Regression (RPSUR) model. Accordingly, the RPSUR model can be written in terms of sample and remaining parts,

\[
\begin{pmatrix}
Z_I \\
Z_{II}
\end{pmatrix} =
\begin{pmatrix}
X_I \\
X_{II}
\end{pmatrix}\mu + E.
\]

It is readily shown that \( E(Z_I) = X_J\mu \) and \( E(Z_{II}) = X_{II}\mu \).

With (3.4), linear estimators of \( T \) can be derived under design-based framework (Stanek, Singer and Lencina 2003) using a method similar to the prediction-based approach (Royall 1976; Bolfarine and Zacks 1992; Valliant, Dorfman and Royall 2000). We will illustrate several methods in the following sections.
3.3. Simultaneous estimation of $p+1$ population totals

The $p+1$ population totals of interest can be represented as a vector

$$ T = \begin{pmatrix} T_0 & T_1 & \cdots & T_p \end{pmatrix}' $$

By definition, $T = I_{p+1} \otimes 1_n' \mathbf{Z} = L'_j \mathbf{Z}_j + L'_{ij} \mathbf{Z}_{ij}$, where

$L'_j = I_{p+1} \otimes 1_n'$ and $L'_{ij} = I_{p+1} \otimes 1_{n-n}'$. We are interested in deriving an estimator for $T$ that is a linear function of $\mathbf{Z}_j$ and unbiased for $T$ while a chosen optimization criterion is satisfied. A linear estimator of $T$ can be defined as,

$$ \hat{T} = (L'_j + W') \mathbf{Z}_j, \quad (3.5) $$

where $W$ is a $n(p+1) \times (p+1)$ matrix of weights (or coefficients). There are many possible ways to define $W$. Usually, $W$ is chosen to satisfy certain optimization criteria, such as to minimize trace of the variance-covariance matrix, generalized mean squared error or M-estimation criteria. An intuitive but arbitrary way to define $W$ is to define $W = I_{p+1} \otimes w_*$, where $w_*$ is $n \times 1$ column vector, that is, a single weight system is applied to all variables of interest. Once $W$ is defined, the corresponding estimation error can be defined as

$$ \hat{T} - T = (L'_j + W') \mathbf{Z}_j - L' \mathbf{Z} = W' \mathbf{Z}_j - L'_{ij} \mathbf{Z}_{ij}. \quad (3.6) $$

The unbiasedness of $\hat{T}$ for $T$ requires that

$$ E(\hat{T} - T) = E(W' \mathbf{Z}_j - L'_{ij} \mathbf{Z}_{ij}) = 0. \quad (3.7) $$

Since $E(\mathbf{Z}_j) = X_j \mu$ and $E(\mathbf{Z}_{ij}) = X_{ij} \mu$, for (3.7) to hold for any $T$, it is required that

$$ W'X_j - L'_{ij}X_{ij} = 0, \quad (3.8) $$

where $X_j = I_{p+1} \otimes 1_n$ and $X_{ij} = I_{p+1} \otimes 1_{n-n}$.
The variance-covariance matrix of $\hat{T}$ under Model (3.4) is defined as

$$E(\hat{T}-T)'(T-T),$$

which can be represented in terms of the variance-covariance matrices of the sample $(V_i)$ and remaining parts $(V_{II})$, and the covariance matrix between the sample and remaining parts $(V_{I,II})$,

$$\text{cov}(\hat{T}-T) = (W' - L_{II}') \begin{pmatrix} V_I & V_{I,II} \\ V_{II,I} & V_{II} \end{pmatrix} \begin{pmatrix} W \\ -L_{II} \end{pmatrix}. \quad (3.9)$$

To find the best estimate for $T$ is to find an estimate for $W$ that satisfies (3.8) and results in minimizing the expected value of a chosen function of $\text{cov}(\hat{T}-T)$.

Commonly used optimization criteria include minimizing $\text{cov}(\hat{T}-T)$ in terms of its

1) trace, or

2) quadratic function $q' \text{cov}(\hat{T}-T)q$, or M-estimation; when $q = 1_{p+1}$, this criterion is referred to as the generalized mean squared error (GMSE).

For example, if minimizing the trace of $\text{cov}(\hat{T}-T)$ is the criterion, we can define the optimization function as (3.8),

$$\Phi(W) = \text{trace}(W'V_I W) - 2\text{trace}(W'V_{I,II}L_{II}) + 2 \sum_{i=1}^{p+1} u_i' (W'X_i - L_{II}'X_{II}) \lambda, \quad (3.10)$$

where $u_i$ is the $i$-th row of identity matrix $I_{p+1}$ and $\lambda$ is a $(p+1) \times 1$ vector of Lagrangian multipliers. Differentiating (3.10) with respect to elements of $W$ and $\lambda$ and setting the derivatives to zero will yield two sets of estimation equations. The estimation equations can then be used to identify $W$. We pursue this approach in Chapter 4.
3.4. Example: best linear unbiased estimator of a population total

A special application of the method proposed in Section 3.3 is to derive a linear unbiased minimum variance estimator (BLUE) of \( T^{(0)} \) when no auxiliary information is used. By definition, \( T^{(0)} = l'_I Z_I + l'_II Z_{II} \), where \( l'_I = \begin{pmatrix} 1_n' \\ 0 \end{pmatrix} \) and \( l'_II = \begin{pmatrix} 1'_{N-n} \\ 0 \end{pmatrix} \). Using (3.5), a linear estimator of \( T^{(0)} \) is defined as

\[
\hat{T}^{(0)} = (l'_I + w')Z_I,
\]

where \( w \) is a \( n(p+1) \times 1 \) column vector. Following (3.6), the estimation error is defined as \( \hat{T}^{(0)} - T^{(0)} = w'Z_I - l'II Z_{II} \). Applying (3.8), the unbiasedness of \( \hat{T}^{(0)} \) for \( T^{(0)} \) becomes

\[
w'X_I - l'II X_{II} = 0,
\]

where \( X_I = I_{p+1} \otimes 1_n \) and \( X_{II} = I_{p+1} \otimes 1_{N-n} \).

Under the unbiasedness constraint (3.12), the mean squared error of \( \hat{T}^{(0)} \) is the same as its variance. As an analogue to (3.10), minimizing \( \text{var}(\hat{T}^{(0)}) \) subject to (3.12) is equivalent to minimize

\[
\Phi(w) = w'V_I w - 2w'V_{I,II}l_{II} + 2(w'X_I - l'II X_{II})\lambda,
\]

where \( \lambda \) is a \( (p+1) \times 1 \) vector of Lagrangian multipliers. Differentiating (3.10) and setting the derivatives to zero results in a set of estimating equations,

\[
\begin{pmatrix} V_I & X_I \\ X'_I & 0 \end{pmatrix} \begin{pmatrix} \hat{w} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} V_{I,II}l_{II} \\ X'_{II}l_{II} \end{pmatrix},
\]

which has a unique solution,
\[ \hat{w} = V^{-1}_{l,l} \left( V_{l,l} - X_l \left( X_l'V^{-1}_{l,l} X_l \right)^{-1} X_l' \left( V^{-1}_{l,l} V_{l,l} - I \right) \right) i_{l,l}. \]  

(3.15)

By applying (3.15) to (3.11), we have the BLUE for \( T^{(0)} \),

\[ \hat{T}_0 = l'_l Z_l + l''_l \left[ X_{l,l} \hat{\beta} + V_{l,l}^{-1} \left( Z_l - X_l \hat{\beta} \right) \right], \]

(3.16)

where \( \hat{\beta} = (X_l'V^{-1}_{l,l} X_l)^{-1} X_l'V^{-1}_{l,l} Z_l \). The variance of \( \hat{T}^{(0)} \) can be evaluated as

\[ \text{var} \left( \hat{T}^{(0)} \right) = \text{var} \left( (l_l + \hat{w})' Z_l \right) = (l_l + w)' V_l (l_l + w). \]

(3.17)

After algebraic simplification,

\[ \hat{\beta} = \frac{1}{n} \left( I_n' Y^{(0)} \ 1_n' Y^{(1)} \ \cdots \ 1_n' Y^{(p)} \right)' \left( \bar{Y}^{(0)} \ \bar{Y}^{(1)} \ \cdots \ \bar{Y}^{(p)} \right)' , \]

and

\[ \hat{T}^{(0)} = n \bar{Y}^{(0)} + (N - n) \bar{Y}^{(0)} = N \bar{Y}^{(0)}, \]

where \( \bar{Y}^{(0)} = \frac{1}{n} 1_n' Y^{(0)} \) is the sample mean of permuted response variable. The variance of \( \hat{T}^{(0)} \) is

\[ \text{var} \left( \hat{T}_0 \right) = N^2 \frac{1 - f}{n} \sigma^2, \]

where \( f = n/N \) is the sampling fraction.

3.5. Estimation of a population total using auxiliary information

In this section, we propose a method of using auxiliary information to improve precision of linear unbiased estimators of population totals through a simple reparameterization. We illustrate this method using a simple case that has one response and one auxiliary variable and assuming the auxiliary total \( T^{(1)} \) is known.
Our target parameter is population total of the response variable,

\[ T^{(0)} = 1_N' Y^{(0)} = t'Z, \]  

(3.18)

where \( t = (1_N' \ 0_{1\times N})' \). We are interested in deriving an estimator for \( T^{(0)} \) that is a linear function of the sample data and unbiased for \( T^{(0)} \), and has minimum mean squared error.

We transform the vector of auxiliary variables \( Y^{(i)} \) by pre-multiplying

\[ I_N = \frac{1_N}{N} 1_N' P_N, \]

such that \( Y^{(i)} = \frac{1_N}{N} \left( I_N' Y^{(i)} \right) + P_N Y^{(i)} \). The first term in this expression is a constant vector \( (\mu^{(i)} I_N) \), and the second term is \( Y^{(i)*} = P_N Y^{(i)} = Y^{(i)} - \mu^{(i)} I_N \). It is clear that

\[
\begin{pmatrix}
Y^{(0)} \\
Y^{(i)}
\end{pmatrix} =
\begin{pmatrix}
I_N & 0 \\
0 & P_N
\end{pmatrix}
\begin{pmatrix}
0 & 0 \\
0 & \frac{1_N}{N} I_N'
\end{pmatrix}
\begin{pmatrix}
Y^{(0)} \\
Y^{(i)}
\end{pmatrix} =
\begin{pmatrix}
Y^{(0)} \\
Y^{(i)*} + \mu^{(i)} I_N
\end{pmatrix} \tag{3.19}
\]

Applying identity (3.19) to Model (3.2) (here \( p = 1 \)) leads to a reparameterized model,

\[
\begin{pmatrix}
Y^{(0)} \\
Y^{(i)*}
\end{pmatrix} = (I_2 \otimes I_N) \mu^* + E, \tag{3.20}
\]

where \( \mu^* = (\mu^{(0)} \ 0)' \). Since this transformation does not have impact on \( E \), its variance-covariance matrix remains the same as \( \text{cov} \left( Y^{(0)}, Y^{(i)*} \right) = \Sigma \otimes P_N \), where \( \Sigma \) is given in Section 2.2.
Using the same method as shown in previous section, \( \begin{pmatrix} Y^{(0)r} & Y^{(1)r} \end{pmatrix} \)' can be partitioned as 
\[ Z^* = K^* \begin{pmatrix} Y^{(0)r} & Y^{(1)r} \end{pmatrix}' = \begin{pmatrix} Z_I' & Z_{II}' \end{pmatrix}', \]
where \( K^* \) is a permutation matrix as given by (2.9), \( Z_I = \begin{pmatrix} Y_I^{(0)r} & Y_I^{(1)r} \end{pmatrix}' \) represents the sample and \( Z_{II} = \begin{pmatrix} Y_{II}^{(0)r} & Y_{II}^{(1)r} \end{pmatrix}' \) represents the remaining part. It immediately follows that 
\[
E(Z_I) = (I_z \otimes I_n)\mu^*,
\]
\[
E(Z_{II}) = (I_z \otimes I_{N-n})\mu^*, \quad V_I = \Sigma \otimes P_{n,N} \quad \text{and} \quad V_{II,I} = V_{II,J} = \Sigma \otimes \left(-\frac{1}{N} J_{n,(N-n)}\right).
\]

Similar to (3.5), \( T^{(0)} \) can be represented as a linear combination of the sampled and remaining parts,

\[
T^{(0)} = I_I'Z_I + I_{II}'Z_{II},
\]
where \( I_I = (1 \ 0)' \otimes I_n \) and \( I_{II} = (1 \ 0)' \otimes I_{N-n} \). Its linear estimator is defined as,

\[
\hat{T}^{(0)} = (I_I' + w')Z_I,
\]
where \( w = \begin{pmatrix} w_0' & w_1' \end{pmatrix}' \) is a \( 2n \times 1 \) vector. Using (3.21) and (3.22), the unbiasedness for \( \hat{T}^{(0)} \) implies

\[
E\left( w'Z_I - I_{II}'Z_{II} \right) = 0.
\]

Since \( E(Z_I) = (I_z \otimes I_n)\mu^*, \ E(Z_{II}) = (I_z \otimes I_{N-n})\mu^* \) and \( \mu^* = \begin{pmatrix} \mu^{(0)} & 0 \end{pmatrix}' \), (3.23) is equivalent to \( (I_n w_0 - (N-n))\mu^{(0)} = 0 \). The constraint is satisfied for any \( \mu^{(0)} \) simply by

\[
w_0'1_n - (N-n) = 0.
\]
The mean squared error of $\hat{T}^{(0)}$ is given as,

$$\text{var}\left(\hat{T}^{(0)}\right) = \begin{pmatrix} w' & -I_n' \end{pmatrix} \begin{pmatrix} V_I & V_{I,II} \\ V_{II,I} & V_{II} \end{pmatrix} \begin{pmatrix} w \\ -I_n \end{pmatrix}.$$ (3.25)

To minimize (3.25) subject to (3.24), we define the optimization function as

$$\Phi(w) = \begin{pmatrix} w_0' & w_1' \end{pmatrix} V_I \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} - 2 \begin{pmatrix} w_0' & w_1' \end{pmatrix} V_{I,II} I_n' + 2 \left( w_0' 1_n - (N - n) \right) \lambda,$$ (3.26)

where $\lambda$ is a Lagrangian multiplier. Differentiating (3.26) and setting the derivatives to zero leads to a set of estimating equations that can be used to identify $w$.

Unlike the usual calibration technique of incorporating auxiliary information, this approach leads a set of weights that do not depend on the sample, and yields a linear unbiased minimum variance estimator of $T_0$. We will explore this approach further in Chapter 5, and extend it to settings when multiple auxiliary totals are known.

### 3.6. Enumeration and simulation

Comparisons between candidate estimators are sometimes difficult through mathematical derivation. When mathematical comparisons are not feasible, we will use enumeration and simulation methods.

When the population size is small, it is possible to enumerate all possible samples and compute the estimators based on each sample. The exact distribution of the estimators can be obtained. Thus comparisons between various types of estimators are possible. We will use this approach to illustrate the performance of estimators for small samples from finite populations of small size.
When population size is large, the number of possible samples is so large that it may be impossible to obtain the exact moments of the estimators. The exact values of an expected value, the bias and the variance of estimator for population totals can not be computed through enumeration. In such cases, Monte Carlo simulations can be used to evaluate the properties of the estimates if theoretical derivation is not possible. The finite population and the sampling design can be held constant, a large number of samples can be repeatedly drawn from the finite population using the specified sampling design. For every realized sample, the estimators can be calculated. If the number of samples is large enough, the distributions of the calculated estimators will closely approximate the exact sampling distribution to a given level of precision.

3.7. Research objectives

In this dissertation, we will primarily consider estimation of population quantities based on samples drawn from a real-valued finite population of known size, using simple random sampling without replacement (SRSWOR). The sample size is assumed fixed. To simplify the problems without losing generality, we assume that response for all sampled subjects can be measured without error.

Discussions will be restricted to the scenarios involving univariate response variable that may be either continuous or binomial. In design-based approaches, estimation of population totals or means of continuous and binomial response variable are not notably different. Unless otherwise noted, we will consider binomial response variables, and refer to the mean estimators as population rates (following the terminology in Epidemiology). The problems involving multinomial or ordinal response
variables may be reduced to estimation using a series of binomial variables, which will not be addressed in detail.

We will develop methods of improving estimation of population quantities (totals, means or rates) by incorporating auxiliary information. The values of auxiliary variables are known either for all population subjects, or at least some of its population quantities (such as population totals) are known. We will consider several simple problems that correspond to various scenarios that may arise according to the nature of response and auxiliary variables, as well as number and information availability of auxiliary variables.

We will define population totals as linear function of random variables arising from permutations. To obtain estimators of target parameters for each scenario, we will follow the strategies outlined in Section 3.3 through 3.5. First, we will derive estimators of population totals and their variance-covariance matrix using a RPSUR model described in 3.2. We optimize the estimators by minimizing either Generalized Mean Squared Errors (GMSE) or the trace of the variance-covariance matrix of population totals.

The properties of the derived estimators will be evaluated through mathematical derivation whenever possible. We will try to represent the estimator in terms of a naïve estimator (without incorporating auxiliary information) plus or multiplied by an adjustment factor. When analytic derivations are not possible, we will evaluate their properties using enumeration and simulation methods.

Each Chapter will start with a simple case that involves single binomial response and single binomial auxiliary variable, therefore the parameters have
straightforward meaning. Results will be gradually extended to more complicated cases that involves continuous response and auxiliary variables or multiple auxiliaries.

Finally, the derived results will be applied and related to direct and indirect adjustment methods in Epidemiology and vital statistics.
CHAPTER 4

ESTIMATION OF POPULATION TOTALS

4.1. Motivations

Public health surveys often involve estimating one or more population totals, such as the total number of persons who contracted a certain disease or the total number of retirees having prescription drug coverage. There may also be interest in joint estimates of multiple population totals when categorical outcomes are involved. For example, the status of a chronic disease is classified as three categories according to severity, that is, None, Stages I and II. Health policy researchers may be interested in estimating the total numbers of persons who contracted the disease of Stages I or II, or of those who did not contract the disease. One may question whether identical estimation methods should be applied to all three categories, and whether the methods are optimal for all three outcomes, or if methods can be identified to satisfy varying precision requirements for different outcomes.

In this Chapter, we derive the best linear unbiased estimator for the vector $\mathbf{T}$ of population totals of multiple variables, and its variance-covariance matrix, $\text{cov}(\hat{\mathbf{T}})$ assuming no auxiliary totals are known. We derive two classes of linear unbiased estimators based on either the minimum trace of $\text{cov}(\hat{\mathbf{T}})$, or the minimum of a specified quadratic function $\mathbf{q}'\text{cov}(\hat{\mathbf{T}})\mathbf{q}$, where $\mathbf{q}$ is a $(p + 1) \times 1$ non-null column vector.
4.2. Parameters of interest

Many estimators can be expressed as a linear function of estimated population totals (Thompson 1999), which is equivalent to prediction of a linear combination of random variables. The estimators for target parameters and their variances can be readily computed based on the estimators and their variance-covariance matrix of relevant population totals. Therefore, the estimation of population totals is the cornerstone of estimating other target parameters.

Suppose that there are $p + 1$ population totals of interest, the vector of population totals can be represented as a set of $p + 1$ linear combinations, $T = L'Z$, where $L = I_{p+1} \otimes I_N$ and $Z$ is a vector that represent the $N(p + 1) \times 1$ random variables arising from joint permutation of the $p + 1$ variables. After sampling is completed, $Z$ can be partitioned into sample and remaining parts such that $Z = (Z'_1 \quad Z''_1)$. Thus $T$ can be written as a set of linear combinations of $Z'_1$ and $Z''_1$ that represent the random variables for the sample and remaining parts,

$$T = \left( L'_1 \mid L''_1 \right) \left( \frac{Z'_1}{Z''_1} \right), \quad (4.1)$$

where $L'_1 = I_{p+1} \otimes I_n'$ and $L''_1 = I_{p+1} \otimes I'_{N-n}$.

4.3. Formulation of a RPSUR model

We will derive estimators of population totals using a simple location seemingly unrelated regression model based on a random permutation framework (RPSUR). Similar to Model 3.4, we define a RPSUR model,
\[
\begin{pmatrix}
Z_I
\end{pmatrix} = \begin{pmatrix}
X_I
\end{pmatrix} \mu + E,
\]

where \( X_I = I_{p+1} \otimes I_n \) and \( X_H = I_{p+1} \otimes I_{N-n} \), \( E(Z_I) = \frac{1}{N} X_I T \) and \( E(Z_H) = \frac{1}{N} X_H T \).

4.4. Two classes of estimators for population total vector

We are interested in developing two classes of estimators of population totals of response variables that is,

1) A linear function of the sample, \( \hat{T} = A'Z_I \)

2) Unbiased for \( T \), i.e., \( E(\hat{T} - T) = 0 \).

4.4.1. Two classes of linear estimators of \( T \)

A linear estimator of \( T = L_I'Z_I + L_H'Z_H \) is defined as a linear function of the sample, e.g., \( \hat{T} = A'Z_I \) or equivalently as,

\[
\hat{T} = (L'_I + W')Z_I,
\]

where \( W' = A' - L'_I \) is a \((p+1) \times n(p+1)\) matrix. In general, all elements of \( W' \) may differ. We limit our development to two classes of estimators, 1) \( W = \bigoplus_{k=0}^{p} w_k \), and 2) \( W = \bigoplus_{k=0}^{p} w_k \), where \( w_k, k = 0, 1, \ldots, p \) are \( n \times 1 \) vectors.

4.4.2. Unbiasedness for \( T \)

The estimation error is defined as \( \hat{T} - T = WZ_I - L'_HZ_H \). The unbiasedness of \( \hat{T} \) implies that \( E(\hat{T} - T) = E[(L'_I + W')Z_I - T] = 0 \) holds for any \( T \). Since
\[ E \left[ (L'_i + W')Z_i - T \right] = (W'X_i - L'_{ii}X_{ii}) \left( N^{-1}T \right), \]

unbiasedness will hold for all \( T \) if and only if

\[
W'X_i - L'_{ii}X_{ii} = 0 .
\] (4.4)

Constraint (4.4) is a set of \( p + 1 \) constraints:

\[
w'_01_n - (N - n) = 0, \\
w'_11_n - (N - n) = 0, \\
\vdots \\
w'_{p}1_n - (N - n) = 0.
\]

For convenience, the set of constraints can be rewritten alternatively as

\[
u'_i (W'X_i - L'_{ii}X_{ii}) = 0 ,
\]

where \( u_i \) is the \( i \)-th row of identity matrix \( I_{p+1} \), \( i = 1, 2, \ldots, p + 1 \).

4.4.3. Optimization criteria

We find the “best” estimators of \( T \) by evaluating \( W \) such that it attains a minimum for a specified optimization criteria. Many optimization criteria are proposed in literature (Deville and Särndal 1992; Singh and Mohl 1996). In this Chapter, we will consider the following optimization criteria:

1) Trace of \( \text{cov} \left( \hat{T} \right) \),

2) M-estimation, \( q' \text{cov} \left( \hat{T} \right)q \), where \( q \) is an arbitrary column vector; when

\[
q = 1_{p+1} ,
\]

it is generalized mean squared error (GMSE), \( 1'_{p+1} \text{cov} \left( \hat{T} \right)1_{p+1} \).
4.5. Estimating multiple population totals

In this section we develop two classes of estimators using the two optimization criteria outlined above. Four cases of interest are presented in Table 1.

Table 4.1. Two classes of unbiased estimators of population totals under two optimization criteria

<table>
<thead>
<tr>
<th>Class</th>
<th>Definition of $W$</th>
<th>Optimization Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$W = \bigoplus_{k=0}^p w_k$</td>
<td>$\text{trace}\left(\text{cov}\left(\hat{T}\right)\right)$</td>
</tr>
<tr>
<td>2</td>
<td>$W = \bigoplus_{k=0}^p w_*$</td>
<td>$q'\text{cov}\left(\hat{T}\right)q$</td>
</tr>
<tr>
<td>3</td>
<td>$W = \bigoplus_{k=0}^p w_*$</td>
<td>$\text{trace}\left(\text{cov}\left(\hat{T}\right)\right)$</td>
</tr>
<tr>
<td>4</td>
<td>$W = \bigoplus_{k=0}^p w_*$</td>
<td>$q'\text{cov}\left(\hat{T}\right)q$</td>
</tr>
</tbody>
</table>

4.5.1. Estimators with minimum $\text{trace}\left(\text{cov}\left(\hat{T}\right)\right)$ when $W = \bigoplus_{k=0}^p w_k$

Minimizing trace of $\text{cov}\left(\hat{T} - T\right)$ subject to the unbiasedness constraints (4.4) is equivalent to minimizing the following Lagrangian function,

$$\Phi(W) = tr(W'V_i W - W'V_{i,ll} L_{ll} - L_{ll}' V_{i,l} W) + \sum_{i=0}^p u_i' \left( W'X_i - L_{ll}' X_{l} \right) \lambda_i, \quad (4.5)$$

where $u_i'$ is the $i$-th row of $I_{p+1}$, $\lambda = \left( \lambda_0 \ldots \lambda_k \ldots \lambda_p \right)'$, $k = 0,1,\ldots,p$, is a $(p+1) \times 1$ column vector of Lagrangian multipliers. Differentiating (4.5) with respect to elements of $W$ and $\lambda$, setting the derivatives to zeros, results in the following estimation equations,

$$\begin{pmatrix} D_i & X_i \\ X_i' & 0 \end{pmatrix} \begin{pmatrix} \hat{W} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} d_{i,ll} \\ X_{l}' L_{ll} 1_{p+1} \end{pmatrix}. \quad (4.6)$$
where \( D_i = \left( \bigoplus_{k=0}^{p} \sigma_k^2 \right) \otimes P_{n,N} \) and \( d_{i,II} = \frac{N-n}{N} \left( \left( \bigoplus_{k=0}^{p} \sigma_k^2 \right) I_p \right) \otimes 1_n \). Estimation equations (4.6) have the following solutions,

\[
\hat{w} = D_i^{-1} d_{i,II} - D_i^{-1} X_i' \left( X_i' D_i^{-1} X_i \right)^{-1} \left( X_i' D_i^{-1} d_{i,II} - X_{i,II}' L_{II} 1_{p+1} \right).
\]

After algebraic simplification, the above solution is equivalent to

\[
\hat{w} = \frac{N-n}{n} 1_n \quad \text{or} \quad \hat{W} = \frac{N-n}{n} \left( I_{p+1} \otimes 1_n \right),
\]

Consequently, the best linear unbiased estimator for \( T \) and its variance are

\[
\hat{T} = N \hat{\mu} \quad \text{and} \quad \text{var}(\hat{T}) = \frac{N(N-n)}{n} \Sigma,
\]

where \( \hat{\mu} = \left( \bar{Y}^{(0)} \quad \bar{Y}^{(1)} \quad \cdots \quad \bar{Y}^{(p)} \right)' \) is the vector of sample means, \( Y^{(k)} = \frac{1}{n} \sum_{s=1}^{n} Y^{(k)}_s \),

\( k = 0,1,\ldots,p \).

Derivation of the above results is presented in Section 4.7.

4.5.2. Estimators with minimum \( q' \text{cov} \left( \hat{T} \right) q \) when \( W = \bigoplus_{k=0}^{p} w_k \)

Another optimization criterion is to minimize an arbitrary quadratic function of \( \text{cov} \left( \hat{T} \right) \), i.e., \( q' \text{cov} \left( \hat{T} \right) q \), where \( q \) is \( (p+1) \times 1 \) column vector of rank 1. When \( q = 1_{p+1} \), \( q' \text{cov} \left( \hat{T} \right) q \) is the same as generalized mean squared error (GMSE). Thus, an optimization function \( \Phi(w) \) can be defined using \( q' \text{cov} \left( \hat{T} \right) q \) and include the unbiased constraint for \( \hat{T} \), such that

\[
\Phi(w) = q' \left( W' V_i W - W' V_{i,II} L_{II} - L_{II}' V_{i,II}' W \right) q + \sum_{i=1}^{n+1} u_i' \left( W' X_i - L_{II}' X_{II} \right) \lambda,
\]

(4.8)
where $\lambda_{(p+1)\times 1} = (\lambda_0, \lambda_1, \ldots, \lambda_p)'$ is a $(p+1) \times 1$ column vector of Lagrangian multipliers. Differentiating (4.8) and setting both sets of derivatives to zero yields the following estimation equations,

$$\begin{pmatrix} V_{qI} & X_I \\ X_I' & 0 \end{pmatrix} \begin{pmatrix} \hat{w} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} V_{qI,II} L_{II} q \\ X_{II}' L_{II} 1_{(p+1)} \end{pmatrix},$$

(4.9)

where $V_{qI}^{-1} = \left( \bigotimes q_k^{-1} \right) \Sigma^{-1} \left( \bigotimes q_k^{-1} \right)$ and $V_{qI,II} = \left( \bigotimes q_k \right) \Sigma \left( -\frac{1}{N} J_{n(N-n)} \right)$.

Estimation equations (4.9) have the following unique solutions,

$$\hat{w} = V_{qI}^{-1} V_{qI,II} L_{II} q - V_{qI}^{-1} X_I \left( X_I' V_{qI}^{-1} X_I \right)^{-1} \left( X_I' V_{qI}^{-1} V_{qI,II} L_{II} q - X_{II}' L_{II} 1_{(p+1)} \right),$$

(4.10)

After some algebraic arrangement, (4.10) can be simplified as,

$$\hat{w} = 1_{p+1} \otimes \left( \frac{N-n}{n} 1_n \right),$$

(4.11)

and $\hat{w}_k = \frac{N-n}{n} 1_n$, for $k = 0, 1, \ldots, p$, $\hat{W} = \frac{N-n}{n} \left( I_{p+1} \otimes 1_n \right)$. Thus, we have the following expressions,

$$\hat{T} = N\hat{u} \quad \text{and} \quad \text{var}(\hat{T}) = \frac{N(N-n)}{n} \Sigma.$$

Derivation of the above results is presented in Section 4.8.

4.5.3. Estimators with minimum trace($\text{cov}(\hat{T})$) when $W = I_{p+1} \otimes w_*$

Since $W = I_{p+1} \otimes w_*$, the unbiased constraints (4.4) reduce to,

$$X_I' w_* - (N-n) = 0,$$

(4.12)

where $X_I = 1_n$. The Lagrangian function (4.5) is simplified to,

$$\Phi(W) = \text{tr} \left( W' V_{lI} W - W' V_{lII} L_{II} - L_{II}' V_{lII} W \right) + \left( X_I' w_* - (N-n) \right) \lambda.$$

(4.13)
Differentiating (4.28) with respect to $w_*$ and $\lambda$ and setting the derivatives to equal zero yields the following estimating equations

\[
\begin{pmatrix}
V_{q,l} & X_l \\
X_l' & 0
\end{pmatrix}
\begin{pmatrix}
\hat{w}_* \\
\hat{\lambda}
\end{pmatrix} =
\begin{pmatrix}
V_{q,l,l} I_l \\
(N-n)
\end{pmatrix},
\]

(4.14)

where $V_{q,l} = tr(\Sigma) P_{q,n}, V_{q,l,l} = -\frac{N-n}{N} tr(\Sigma)$ and $I_l = I_n$. The solutions of (4.14) are

\[
\hat{w}_* = V_{q,l}^{-1} V_{q,l,l} I_l - V_{q,l}^{-1} X_l \left( X_l' V_{q,l}^{-1} X_l \right)^{-1} \left( X_l' V_{q,l}^{-1} V_{q,l,l} I_l - (N-n) \right).
\]

(4.15)

After algebraic simplification, (4.15) becomes

\[
\hat{w}_* = \frac{N-n}{n} I_n,
\]

or equivalently,

\[
\hat{W} = \frac{N-n}{n} (I_{p+1} \otimes I_n).
\]

Consequently, $\hat{\Sigma} = N\hat{\mu}$ and $\text{var}(\hat{\Sigma}) = \frac{N(N-n)}{n} \Sigma$.

Detailed derivations are given in Section 4.9.

4.5.4. Estimators with minimum $q' \text{cov}(\hat{\Sigma}) q$ when $W = I_{p+1} \otimes w_*$

Using constraint (4.12), the Lagrangian function of this case can be defined as

\[
\Phi(w) = q'W'V_l W q - 2q'W'V_{l,l} L_{\mu} q + (X_l' w_* - (N-n)) \lambda.
\]

(4.16)

Differentiating (4.16) and setting the derivatives to zero yields the following estimating equations,

\[
\begin{pmatrix}
V_{q,l} & X_l \\
X_l' & 0
\end{pmatrix}
\begin{pmatrix}
\hat{w}_* \\
\hat{\lambda}
\end{pmatrix} =
\begin{pmatrix}
V_{q,l,l} I_l \\
(N-n)
\end{pmatrix},
\]

(4.17)
where $V_{q,l} = (q\Sigma q)P_{q,l,n}$, $V_{q,l,l} = -\frac{N-n}{N}(q\Sigma q)$, $I_l = 1_n$ and $X_l = 1_n$. The solutions of (4.17) are

$$\hat{w}_* = V_{q,l}^{-1}V_{q,l,l}I_l - V_{q,l}^{-1}X_l\left(X_l'V_{q,l}^{-1}V_{q,l,l}I_l - (N-n)\right).$$  (4.18)

After simplification, (4.18) becomes,

$$\hat{w}_* = \frac{N-n}{n}1_n,$$

or equivalently,

$$\hat{W} = \frac{N-n}{n}(I_{p+1} \otimes 1_n).$$

Consequently,

$$\hat{T} = N\left(I_{p+1} \otimes 1_n\right)Z_l = N\hat{u}, \text{ and } \text{var}(\hat{T}) = \frac{N(N-n)}{n} \Sigma.$$  

Detailed derivations are given in Section 4.10.

4.6. Remarks

In this Chapter, we derived two classes of estimators that are defined by two different sets of coefficients, i.e., $W = \bigoplus_{k=0}^{p} w_k$ and $W = I_{p+1} \otimes w_*$. The estimators were optimized by minimizing either the trace of $\text{cov}(\hat{T})$ or $q'\text{cov}(\hat{T})q$. Under SRSWOR and unbiasedness constraints for every element of $\hat{T}$, the estimators for all four cases are identical and shown to be simple expansion estimators. It is apparent that neither the method of defining estimators nor the optimization criterion had impact on the estimation.
4.7. Derivations of results in Section 4.5.1.

This section provides proof of the results in Section 4.5.1. A class of estimators defined by \( W = \bigoplus_{k=0}^{p} w_k \), where \( w_k \) may or may not be equal to each other, is derived by minimizing \( \text{trace} \left( \text{cov} \left( \hat{T} \right) \right) \) subject to the unbiasedness constraints for every element of \( T \).

As shown in Section 4.4.2., the unbiasedness constraints are

\[
W'X_I - L''_I X_{II} = 0.
\]  
(4.19)

Therefore, minimizing \( \text{trace} \left( \text{cov} \left( \hat{T} \right) \right) \) subject to unbiased constraints (4.19) is equivalent to minimizing the following Lagrangian function,

\[
\Phi(W) = tr \left( W'V_I W - W'V_{I,II}L_{II} - L''_I V_{I,II}' W \right) + 2\sum_{i=1}^{p+1} u_i \left( W'X_I - L''_I X_{II} \right) \lambda,
\]  
(4.20)

where \( \lambda \) is a \((p+1) \times 1\) vector of Lagrangian multipliers.

Since \( V_I = \Sigma \otimes P_{n,N}, \ P_{n,N} = I_n - \frac{1}{N} J_n \), \( W = \bigoplus_{k=0}^{p} w_k \),

\[
W'V_I W = \left( \bigoplus_{k=0}^{p} w_k \right) \left( \Sigma \otimes P_{n,N} \right) \left( \bigoplus_{k=0}^{p} w_k \right)
\]

\[
= \begin{pmatrix}
\sigma_0^2 w_0' P_{n,N} w_0 & \sigma_0 w_0' P_{n,N} w_1 & \cdots & \sigma_0 w_0' P_{n,N} w_p \\
\sigma_0 w_0' P_{n,N} w_0 & \sigma_0^2 w_1' P_{n,N} w_1 & \cdots & \sigma_1 w_1' P_{n,N} w_p \\
\cdots & \cdots & \cdots & \cdots \\
\sigma_0 w_p' P_{n,N} w_0 & \sigma_p w_p' P_{n,N} w_1 & \cdots & \sigma_p^2 w_p' P_{n,N} w_p
\end{pmatrix}.
\]

It is obvious that \( \text{trace} \left( W'V_I W \right) = \sum_{k=1}^{p} \sigma_k^2 w_k' P_{n,N} w_k \), \( \frac{\partial}{\partial w_k} \text{trace} \left( W'V_I W \right) = 2\sigma_k^2 P_{n,N} w_k \)

and
\[
\frac{\partial}{\partial \mathbf{w}} \text{tr} \left( \mathbf{W}' \mathbf{V}_J \mathbf{W} \right) = \begin{bmatrix}
\frac{\partial}{\partial \mathbf{w}_0} \sum_{k=1}^{p} \sigma_k^2 \mathbf{w}'_k \mathbf{P}_{n,N} \mathbf{w}_k \\
\frac{\partial}{\partial \mathbf{w}_2} \sum_{k=1}^{p} \sigma_k^2 \mathbf{w}'_k \mathbf{P}_{n,N} \mathbf{w}_k \\
\vdots \\
\frac{\partial}{\partial \mathbf{w}_p} \sum_{k=1}^{p} \sigma_k^2 \mathbf{w}'_k \mathbf{P}_{n,N} \mathbf{w}_k
\end{bmatrix} = \begin{bmatrix}
2\sigma_0^2 \mathbf{P}_{n,N} \mathbf{w}_0 \\
2\sigma_1^2 \mathbf{P}_{n,N} \mathbf{w}_1 \\
\vdots \\
2\sigma_p^2 \mathbf{P}_{n,N} \mathbf{w}_p
\end{bmatrix} = 2 \left( \bigoplus_{k=0}^{p} \sigma_k^2 \right) \otimes \mathbf{P}_{n,N} \begin{bmatrix}
\mathbf{w}_0 \\
\mathbf{w}_1 \\
\vdots \\
\mathbf{w}_p
\end{bmatrix}
\]
\[
= 2 \left( \mathbf{D}_\Sigma \otimes \mathbf{P}_{n,N} \right) \mathbf{w}
\]
\[
= 2 \mathbf{D}_I \mathbf{w},
\]
where \( \mathbf{D}_\Sigma = \bigoplus_{k=0}^{p} \sigma_k^2 \) and \( \mathbf{D}_I = \mathbf{D}_\Sigma \otimes \mathbf{P}_{n,N} \).

Since \( \text{tr}(\mathbf{W}' \mathbf{V}_J \mathbf{L}_{II}) = -\frac{N-n}{N} \sum_{k=0}^{p} \sigma_k^2 \mathbf{w}'_k \mathbf{1}_n \),
\[
\frac{\partial}{\partial \mathbf{w}} \text{tr} \left( \mathbf{W}' \mathbf{V}_J \mathbf{L}_{II} \right) = -\frac{N-n}{N} \mathbf{D}_\Sigma \mathbf{1}_p \otimes \mathbf{1}_n
\]
we have \( \frac{\partial}{\partial \mathbf{w}} \text{tr} \left( \mathbf{W}' \mathbf{V}_J \mathbf{L}_{II} \right) = -\frac{N-n}{N} \left( \mathbf{D}_\Sigma \mathbf{1}_p \right) \otimes \mathbf{1}_n \). In addition,
\[
\frac{\partial}{\partial \mathbf{w}} \sum_{k=0}^{p} \text{tr} \left( \mathbf{u}'_{k+1} \mathbf{W}' \mathbf{X}_J \mathbf{X}_J \right) = \sum_{k=0}^{p} \frac{\partial}{\partial \mathbf{w}} \text{tr} \left( \mathbf{u}'_{k+1} \left( \sum_{k=0}^{p} \mathbf{w}' \left( \left( \mathbf{u}'_{k+1} \mathbf{u}'_{k+1} \right) \otimes \mathbf{1}_n \right) \mathbf{X}_J \right) \right)
\]
\[
= \sum_{k=0}^{p} \sum_{k=0}^{p} \frac{\partial}{\partial \mathbf{w}} \text{tr} \left( \mathbf{u}'_{k+1} \mathbf{u}'_{k+1} \left( \mathbf{w}' \left( \left( \mathbf{u}'_{k+1} \mathbf{u}'_{k+1} \right) \otimes \mathbf{1}_n \right) \mathbf{X}_J \right) \right)
\]
\[
= \sum_{k=0}^{p} \sum_{k=0}^{p} \left( \mathbf{u}'_{k+1} \mathbf{u}'_{k+1} \right) \frac{\partial}{\partial \mathbf{w}} \left( \mathbf{w}' \left( \left( \mathbf{u}'_{k+1} \mathbf{u}'_{k+1} \right) \otimes \mathbf{1}_n \right) \mathbf{X}_J \right)
\]
\[
= \left( \sum_{k=0}^{p} \sum_{k=0}^{p} \left( \mathbf{u}'_{k+1} \mathbf{u}'_{k+1} \right) \left( \left( \mathbf{u}'_{k+1} \mathbf{u}'_{k+1} \right) \otimes \mathbf{1}_n \right) \right) \mathbf{X}_J
\]
\[
= \mathbf{X}_J \mathbf{X}_J.
\]

Therefore, differentiating (4.20) with respect to elements of \( \mathbf{w} \) and \( \lambda \) gives
\[
\frac{\partial}{\partial \mathbf{w}} \Phi (\mathbf{w}) = 2 \mathbf{D}_I \mathbf{w} - 2 \mathbf{D}_{II} - 2 \mathbf{X}_J \lambda,
\]
(4.21)
\[
\frac{\partial}{\partial \lambda} \Phi (w) = (X'_i W - X'_i L_{ii}) \sum_{i=1}^{p+1} u_i = X'_i w - X'_i L_{ii} 1_{p+1},
\] (4.22)

where \( D_j = D_{\Sigma} \otimes P_{n, N} \) and \( d_{i, ii} = -\frac{N - n}{N} (D_{\Sigma} 1_p) \otimes 1_n \). Setting both (4.21) and (4.22) to zero results in the following estimating equations,

\[
\begin{pmatrix}
D_i & X_i \\
X'_i & 0
\end{pmatrix}
\begin{pmatrix}
\hat{w} \\
\hat{\lambda}
\end{pmatrix} =
\begin{pmatrix}
d_{i, ii} \\
X'_i L_{ii} 1_{p+1}
\end{pmatrix}.
\] (4.23)

The solutions to (4.23) are

\[
\hat{w} = D_i^{-1} d_{i, ii} - D_i^{-1} X_i \left( X'_i D_i^{-1} X_i \right)^{-1} \left( X'_i D_i^{-1} d_{i, ii} - X'_i L_{ii} 1_{p+1} \right).
\]

Since \( P_{n, N}^{-1} = \left( I_n + \frac{1}{N - n} J_n \right) \), \( P_{n, N}^{-1} 1_n = -\frac{N}{N - n} 1_n \), \( 1_n P_{n, N}^{-1} 1_n = \frac{NN}{N - n} \), \( D_i^{-1} d_{i, ii} = -1_{p+1} \otimes 1_n \), we have

\[
\begin{pmatrix}
D_i^{-1} X_i \\
X'_i
\end{pmatrix}
\begin{pmatrix}
\left( X'_i D_i^{-1} X_i \right)^{-1} \\
\left( X'_i D_i^{-1} d_{i, ii} - X'_i L_{ii} 1_{p+1} \right)
\end{pmatrix}
\]

\[
= \left( D_{\Sigma}^{-1} D_{\Sigma} \right) \otimes \left( P_{n, N}^{-1} 1_n \left( I_{n} P_{n, N}^{-1} 1_n \right)^{-1} \right)
\]

\[
= I_{p+1} \otimes \left( \frac{1}{n} 1_n \right).
\]

\[
X'_i D_i^{-1} d_{i, ii} - X'_i L_{ii} 1_{p+1} = -\frac{N - n}{N} \left( I_{p+1} \otimes 1_n \right) \left( D_{\Sigma}^{-1} \otimes P_{n, N}^{-1} \right) \left( D_{\Sigma} \otimes 1_n \right) \left( I_{p+1} \otimes 1_n \right)
\]

\[
- \left( I_{p+1} \otimes 1_{N-n} \right) \left( I_{p+1} \otimes 1_{N-n} \right) 1_{p+1}
\]

\[
= -n I_{p+1} - \left( N - n \right) 1_{p+1} = -n \left( I_{p+1} + \frac{N - n}{n} 1_{p+1} \right),
\]

\[
\hat{w} = D_i^{-1} d_{i, ii} - D_i^{-1} X_i \left( X'_i D_i^{-1} X_i \right)^{-1} \left( X'_i D_i^{-1} d_{i, ii} - X'_i L_{ii} 1_{p+1} \right)
\]

\[
= -I_{p+1} \otimes 1_n + \left( I_{p+1} \otimes 1_n \right) \left( I_{p+1} + \frac{N - n}{n} 1_{p+1} \right)
\]

\[
= \frac{N - n}{n} 1_{p+1} \otimes 1_n.
\]
Consequently, \( \hat{W} = \frac{N-n}{n} (I_{p+1} \otimes I_n) \), and
\[
\hat{T} = \left( I_{p+1} \otimes I_n' + \frac{N-n}{n} \left( I_{p+1} \otimes I_n' \right) \right) Z_f = N \left( I_{p+1} \otimes I_n' \right) Z_f = N \hat{\mu}.
\]
\[
\text{var} \left( \hat{T} \right) = \left( \frac{N-n}{n} \right) (I_p \otimes I_{n'})(\Sigma \otimes P_{n,n'}) \left( \frac{N-n}{n} \right) (I_p \otimes I_n) = \frac{N(N-n)}{n} \Sigma.
\]

4.8. Derivations of results in Section 4.5.2.

This section presents the derivations for a class of estimators defined by
\[ W = \bigoplus_{k=0}^{p} w_k \] and optimized by minimizing \( q' \text{cov} \left( \hat{T} \right) q \) subject to the unbiasedness constraints for every element of \( T \), where \( q \) is an \( n \times 1 \) vector.

As shown in Section 4.4.2., the unbiasedness constraints are
\[ W'X_I - L'_H X_{II} = 0. \quad (4.24) \]
Minimizing \( q' \text{cov} \left( \hat{T} \right) q \) subject to unbiased constraints (4.24) is equivalent to minimizing the following Lagrangian function,
\[
\Phi(W) = q' \left[ W'V_I W - W' \left( V_{I,II} L_{II} - (V_{I,II} L_{II})' \right) W \right] q + 2 \sum_{i=1}^{p+1} u_i' (W'X_I - L'_H X_{II}) \lambda_i,
\]
where \( \lambda_{(p+1) \times 1} = (\lambda_0, \lambda_1, \ldots, \lambda_p)' \) is a \( (p+1) \times 1 \) vector of Lagrangian multipliers.

For convenience, we denote \( w = (w_0', w_1', \ldots, w_p')' \). It is clear that
\[ w = \left( \bigoplus_{k=0}^{p} w_k \right) I_{p+1} \] and \[ W = \sum_{i=1}^{p+1} \left( (u_i u_i') \otimes I_n \right) w_i', \] where \( u_i \) and \( u_i' \) are the \( i \)-th column or row of identity matrix \( I_{p+1} \). Therefore, \( \Phi(w) = \Phi(W) \).
Since $Wq = \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n$ and $q'W' = w' \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n$,

$$\frac{\partial}{\partial w} q'W'V_i w q = \frac{\partial}{\partial w} w' \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n w' = 2 \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n w,$$

$$\frac{\partial}{\partial w} q'W'(V_{l,ll} L_{ll}) q = \frac{\partial}{\partial w} w' \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n V_{l,ll} L_{ll} q = \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n V_{l,ll} L_{ll} q.$$

As shown in Section 4.9, $\sum_{i=1}^{p+1} u_i' (W'X_l - L_{ll}'X_{ll}) \lambda = w'X_l \lambda - L_{ll}'X_{ll} \lambda$,

$$\frac{\partial}{\partial w} \sum_{i=1}^{p+1} u_i' (W'X_l - L_{ll}'X_{ll}) \lambda = X_l \lambda, \quad \frac{\partial}{\partial \lambda} \sum_{i=1}^{p+1} u_i' (W'X_l - L_{ll}'X_{ll}) \lambda = X_l'w - X_{ll}'L_{ll} 1_{p+1}.$$

Consequently, the two sets of derivatives are,

$$\frac{\partial \Phi}{\partial w} = 2 \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n w \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n V_{l,ll} L_{ll} q + 2X_l \lambda,$$

$$\frac{\partial \Phi}{\partial \lambda} = X_l'w - X_{ll}'L_{ll} 1_{p+1}.$$

Setting both sets of derivatives to zeros, we have,

$$\begin{bmatrix} \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n V_l \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n X_l \hat{w} \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X_l' \lambda \\ 0 \end{bmatrix} = \begin{bmatrix} X_l' \lambda \end{bmatrix}. \quad (4.25)$$

Denote $V_{ql} = \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n V_l \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n$ and $V_{ql,ll} = \left( \bigoplus_{k=0}^{p} q_k \right) \otimes I_n V_{l,ll}$, the solutions to (4.25) can be written as

$$\begin{bmatrix} \hat{\lambda} \end{bmatrix} = \left( X_l' V_{ql}^{-1} X_l \right)^{-1} \left[ X_l' V_{ql}^{-1} V_{ql,ll} X_{ll} - X_l' \right] L_{ll} 1_{p+1},$$

$$\begin{bmatrix} \hat{w} \end{bmatrix} = V_{ql}^{-1} V_{ql,ll} L_{ll} 1_{p+1} - V_{ql}^{-1} X_l \left( X_l' V_{ql}^{-1} X_l \right)^{-1} \left[ X_l' V_{ql}^{-1} V_{ql,ll} X_{ll} - X_l' \right] L_{ll} 1_{p+1}. \quad (4.26)$$
The explicit expression for $\hat{w}$ can be derived as follows,

$$\hat{w} = V^{-1}_{ql}V_{ql,II}L^{-1}_{II}q - V^{-1}_{ql}X_q \left(X'_qV^{-1}_{ql}X_q\right)^{-1} \left[X'_qV^{-1}_{ql}V_{ql,II}L^{-1}_{II}q - X'_{II}L^{-1}_{II}I_{p+1}\right],$$

where

$$V^{-1}_{ql} = \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes I_n \left[\Sigma^{-1} \otimes \mathbf{P}^{-1}_{n,N}\right] \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes I_n = \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \Sigma^{-1} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes \mathbf{P}^{-1}_{n,N},$$

$$V_{ql,II} = \left(\bigotimes_{k=0}^{p} q_k\right) \Sigma \otimes \left(-\frac{1}{N} \mathbf{J}_{n\times(N-n)}\right), \quad V^{-1}_{ql}V_{ql,II} = \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes \left(-\frac{1}{N-n} \mathbf{J}_{n\times(N-n)}\right),$$

$$V^{-1}_{ql}X_q = \frac{N}{N-n} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \Sigma^{-1} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes I_n,$$

$$\left(X'_qV^{-1}_{ql}X_q\right)^{-1} = \left(I_{p+1} \otimes I'_n\right) \left(\left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \Sigma^{-1} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes \mathbf{P}^{-1}_{n,N}\right) \left(I_{p+1} \otimes I'_n\right)^{-1}$$

$$= \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \Sigma^{-1} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \left(I'_n\mathbf{P}^{-1}_{n,N}I_n\right)^{-1} = \frac{N-n}{N} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \Sigma^{-1} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right)^{-1}$$

$$X'_{II}V^{-1}_{ql}V_{ql,II} = \left(I_{p+1} \otimes I'_n\right) \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes \left(-\frac{1}{N-n} \mathbf{J}_{n\times(N-n)}\right) = -\frac{n}{N-n} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes I'_n.$$

Therefore,

$$\hat{w} = \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes \left(-\frac{1}{N-n} \mathbf{J}_{n\times(N-n)}\right) \left(I_{p+1} \otimes I'_{N-n}\right)q$$

$$-\frac{N}{N-n} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \Sigma^{-1} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes I_n \frac{N-n}{N} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \Sigma^{-1} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right)^{-1}$$

$$\times \left[-\frac{n}{N-n} \left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes I'_{N-n} \left(I_{p+1} \otimes I'_{N-n}\right) - \left(I_{p+1} \otimes I'_{N-n}\right) \left(I_{p+1} \otimes I'_{N-n}\right) - \left(I_{p+1} \otimes I'_n\right) \right]$$

$$= -\left(\bigotimes_{k=0}^{p} q_k^{-1}\right) \otimes I_n \left[-\frac{1}{n} \left(I_{p+1} \otimes I_n\right) \right] - n \left(\bigotimes_{k=0}^{p} q_k^{-1}\right)q - (N-n)I_{p+1}$$

$$= -\left(I_{p+1} \otimes I_n\right) + \frac{N}{n} \left(I_{p+1} \otimes I_n\right),$$

or simply,
\[
\hat{w} = \frac{N-n}{n}(I_{p+1} \otimes I_n) \quad \text{and} \quad \hat{W} = I_{p+1} \otimes \left( \frac{N-n}{n}I_n \right).
\]

### 4.9. Derivations of results in Section 4.5.3.

This section provides proof of the results in Section 4.5.3. A class of estimators defined by \( W = I_{p+1} \otimes w_* \) is derived by minimizing trace of \( \text{cov}(\hat{T}) \) subject to the unbiasedness constraints for every element of \( T \).

Since \( W = I_{p+1} \otimes w_* \), the unbiasedness constraints are reduced to

\[
w_*'I_n - (N-n) = 0. \tag{4.27}
\]

Minimizing the trade of \( \text{cov}(\hat{T}) \) subject to (4.27) is equivalent to minimizing the following Lagrangian function,

\[
\Phi(W) = tr\left(W'V_i W - W'V_{i,ll} L_{ll} - L_{ll}' V_i' W\right) + 2\left(w_*'I_n - (N-n)\right)\lambda, \tag{4.28}
\]

where \( \lambda \) is a Lagrangian multiplier.

Since \( W'V_i W = (I_{p+1} \otimes w_*')V_i (I_{p+1} \otimes w_*) \), \( V_i = \Sigma \otimes P_{n,N} \),

\[
tr\left(W'V_i W\right) = tr\left(\left(I_{p+1} \otimes w_*'\right)\left(\Sigma \otimes P_{n,N}\right)\left(I_{p+1} \otimes w_*\right)\right) = tr(\Sigma)\left(w_*'P_{n,N}w_*\right),
\]

\[
\frac{\partial}{\partial w_*} tr\left(W'V_i W\right) = tr(\Sigma)\frac{\partial}{\partial w_*}\left(w_*'P_{n,N}w_*\right) = 2tr(\Sigma)P_{n,N}w_*,
\]

\[
\text{trace}\left(W'V_{i,ll} L_{ll}\right) = \text{trace}\left(\left(I_{p+1} \otimes w_*'\right)\left(\Sigma \otimes \left(-\frac{1}{N}J_{n\times(N-n)}\right)\right)\left(I_{p+1} \otimes I_{N-n}\right)\right),
\]

\[
= w_*'\left(-tr(\Sigma)\frac{N-n}{N}I_n\right)
\]

\[
\frac{\partial}{\partial w_*} tr\left(W'V_{i,ll} L_{ll}\right) = -\frac{N-n}{N}tr(\Sigma)I_n,
\]

64
\[
\frac{\partial}{\partial w_*} \left( w_*'I_n - (N - n) \right) \lambda = I_n' \lambda ,
\]

\[
\frac{\partial}{\partial \lambda} \Phi(w^*) = I_n'w_* - (N - n) .
\]

Therefore, differentiating (4.28) with respect to \( w_* \) and \( \lambda \) and setting the derivatives to equal zero gives the following estimating equations

\[
\left( \begin{array}{c}
V_{w,l} X_l \\
X_l'
\end{array} \right) \begin{pmatrix}
\dot{w}_*
\
\dot{\lambda}
\end{pmatrix} = \begin{pmatrix}
V_{w,l,l} I_l \\
N - n
\end{pmatrix},
\]

(4.29)

where \( V_{w,l} = tr(\Sigma)P_{n,N} , X_l = I_n \), \( V_{w,l,l} = -\frac{N - n}{N}tr(\Sigma) \) and \( I_l = I_n \). The solutions of (4.29) are

\[
\hat{w}_* = V_{w,l}^{-1} V_{w,l,l} I_l - V_{w,l}^{-1} X_l \left( X_l' V_{w,l}^{-1} X_l \right)^{-1} \left( X_l' V_{w,l}^{-1} V_{w,l,l} I_l \right) - (N - n) .
\]

(4.30)

After algebraic simplification, (4.30) is equivalent to \( \hat{w}_* = (N - n) \frac{I_n}{n} \) or

\[
\hat{W} = \frac{N - n}{n} \left( I_{p+1} \otimes I_n \right) .
\]

Consequently,

\[
\hat{T} = N \left( I_{p+1} \otimes \frac{I_n}{n} \right) Z_l = N\hat{\mu} , \quad \text{and} \quad \text{var}(\hat{T}) = \frac{N(N - n)}{n} \Sigma .
\]

4.10. Derivations of results in Section 4.5.4.

This section provides derivations for the class of estimators defined by

\[
W = I_{p+1} \otimes w_* \quad \text{and optimized by minimizing} \quad q' \text{cov}(T)q \quad \text{subject to the unbiasedness constraints for every element of} \quad T , \quad \text{where} \quad q \quad \text{is an} \quad n \times 1 \quad \text{vector.}
\]

Since \( W = I_{p+1} \otimes w_* \), the unbiasedness constraints reduce to

\[
I_n'w_* - (N - n) = 0 .
\]

(4.31)
Minimizing $q'\text{cov}(\tilde{T})q$ subject to (4.24) is equivalent to minimize the following Lagrangian function,

$$
\Phi(W) = q'W'V_1W - 2q'W'(V_{1,H}L_H)q + 2(1_n'w_* - (N-n))\lambda,
$$

where $\lambda$ is a $(p+1)\times 1$ column vector of Lagrangian multipliers. The Lagrangian function can be written alternatively as,

$$
\Phi(w_*) = q'W'V_1Wq - 2q'W'(V_{1,H}L_H)q + 2(1_n'w_* - (N-n))\lambda.
$$

Let us evaluate them by part.

$$
\frac{\partial}{\partial w_*} q'W'V_1Wq = \frac{\partial}{\partial w_*} (q' \otimes w_*) \left( \Sigma \otimes P_{n,N} \right) (q \otimes w_*) = \frac{\partial}{\partial w_*} (q' \Sigma q) w'_* P_{n,N} w_*
$$

$$
= 2(q' \Sigma q) P_{n,N} w_*
$$

$$
\frac{\partial}{\partial w_*} q'W'(V_{1,H}L_H)q = \frac{\partial}{\partial w_*} q' \left( I_{p+1} \otimes w_* \right) \left( V_{1,H}L_H \right) q
$$

$$
= \frac{\partial}{\partial w_*} \left( q' \otimes w_* \right) \left( -\frac{1}{N} \Sigma \otimes J_{n\times(N-n)} \right) \left( I_{p+1} \otimes 1_{N-n} \right) q
$$

$$
= -\frac{N-n}{N} (q' \Sigma q) \frac{\partial}{\partial w_*} (1_n'w_*)
$$

$$
= -\frac{N-n}{N} (q' \Sigma q) 1_n
$$

$$
\frac{\partial}{\partial w_*} (1_n'w_* - (N-n))\lambda = 1_n \lambda,
$$

$$
\frac{\partial}{\partial \lambda} (1_n'w_* - (N-n))\lambda = 1_n'w_* - (N-n).
$$

Therefore,

$$
\frac{\partial}{\partial w_*} \Phi(w_*) = 2(q' \Sigma q) P_{n,N} w_* - 2\frac{N-n}{N} (q' \Sigma q) 1_n + 21_n \lambda \quad (4.32)
$$

$$
\frac{\partial}{\partial \lambda} \Phi(w_*) = 1_n'w_* - (N-n) \quad (4.33)
$$
Set both (4.32) and (4.33) to zero, we have,

\[
\begin{pmatrix}
V_{q,I} & X_I \\
X_I' & 0
\end{pmatrix}
\begin{pmatrix}
\hat{w}_* \\
\hat{\lambda}
\end{pmatrix}
= \begin{pmatrix}
V_{q,I,H} I_I \\
(N - n)
\end{pmatrix},
\tag{4.34}
\]

where \( V_{q,I} = (q'\Sigma q) P_{n,N} \), \( V_{q,I,H} = -\frac{N - n}{N}(q'\Sigma q) \), \( I_I = 1_n \) and \( X_I = 1_n \). The solutions of (4.34) are

\[
\hat{w}_* = V_{q,I}^{-1} V_{q,I,H} I_I - V_{q,I}^{-1} X_I \left( X_I' V_{q,I}^{-1} X_I \right)^{-1} \left( X_I' V_{q,I}^{-1} V_{q,I,H} I_I - (N - n) \right).
\tag{4.35}
\]

After simplification, (4.35) becomes,

\[
\hat{w}_* = \frac{N - n}{n} 1_n,
\]

or equivalently,

\[
\hat{W} = \frac{N - n}{n} (I_{p + I} \otimes 1_n).
\]

Consequently,

\[
\hat{T} = N \left( I_{p + I} \otimes \frac{1_n}{n} \right) Z_I = N \hat{\mu}, \quad \text{and} \quad \text{var}(\hat{T}) = \frac{N(N - n)}{n} \Sigma.
\]
CHAPTER 5

ESTIMATING POPULATION TOTALS BY REPARAMETERIZATION

5.1. Motivation

In many practical situations, auxiliary information such as gender, age, income and chronic disease-bearing history are completely or partially known either from a sampling frame or publicly available census data. Although such information is commonly used in survey design, its use in estimation after sampling is equally important. In this Chapter, we demonstrate several methods of incorporating auxiliary information in estimation through reparameterization. We will consider two scenarios, i.e.,

1) Values of auxiliary variable are known for all units in the population,
2) Values of auxiliary variable are known only for all units in the sample while its population total (or mean) is known for the auxiliary variable.

5.2. Formulation of the model through reparameterization

In Chapter 4, we defined a random permutation seemingly unrelated regression model,

\[
\begin{pmatrix}
Y^{(0)} \\
Y^{(1)}
\end{pmatrix} = \left( I_2 \otimes 1_N \right) \mu + \varepsilon. \tag{5.1}
\]

In both scenarios considered here, the auxiliary total \( T^{(1)} \) (or the mean \( \mu^{(1)} \)) is known. As a result, the \( N \) random variables representing the permutation of auxiliary variable \( Y^{(1)} \) is constrained by \( 1_N' Y^{(1)} = T^{(1)} \). We transform the vector of auxiliary variables
by pre-multiplying $I_N = \frac{1}{N} \mathbf{1}_N \mathbf{1}_N' + P_N$, such that $Y^{(l)} = \frac{1}{N} \left( I_N Y^{(l)} \right) + P_N Y^{(l)}$. The first term in this expression is a vector of constant $\left( \mu^{(l)} \mathbf{1}_N \right)$. We define the second term as

$$Y^{(l)*} = P_N Y^{(l)} = Y^{(l)} - \mu^{(l)} \mathbf{1}_N.$$  

This transformation can be represented jointly in terms of both $Y^{(0)}$ and $Y^{(l)}$,

$$\Pi \begin{pmatrix} Y^{(0)} \\ Y^{(l)} \end{pmatrix} = \begin{pmatrix} Y^{(0)} \\ Y^{(l)*} \end{pmatrix} + \begin{pmatrix} 0 \\ \mu^{(l)} \mathbf{1}_N \end{pmatrix},$$

where $\Pi = \begin{pmatrix} I_N & 0 \\ 0 & P_N \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \frac{1}{N} \mathbf{1}_N \mathbf{1}_N' \end{pmatrix}$. Therefore, Model (5.1) becomes,

$$\begin{pmatrix} Y^{(0)} \\ Y^{(l)*} \end{pmatrix} + \begin{pmatrix} 0 \\ \mu^{(l)} \mathbf{1}_N \end{pmatrix} = \begin{pmatrix} \mu^{(0)} \mathbf{1}_N & 0 \\ 0 & \mu^{(l)} \mathbf{1}_N \end{pmatrix} + E,$$

or simply

$$\begin{pmatrix} Y^{(0)} \\ Y^{(l)*} \end{pmatrix} = (I_2 \otimes \mathbf{1}_N) \mu^* + E,$$

where $\mu^* = \begin{pmatrix} \mu^{(0)} \\ 0 \end{pmatrix}$. As shown in A2.1.15, $\text{cov} \begin{pmatrix} Y^{(0)} \\ Y^{(l)*} \end{pmatrix} = \Sigma \otimes P_N$, where

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{pmatrix}.$$

The next step is to partition $\begin{pmatrix} Y^{(0)*} \\ Y^{(l)*} \end{pmatrix}'$ into two parts that represent sample and the remaining parts,
\[ Z' = K^* \begin{pmatrix} Y^{(0)}_t \\ Y^{(1)*}_t \end{pmatrix} = \begin{pmatrix} Z_I \\ Z_{II} \end{pmatrix}, \]

where \( K^* \) is a permutation matrix as given by (2.9) in Chapter 2, \( Z_I = \begin{pmatrix} Y^{(0)r}_t \\ Y^{(1)r}_t \end{pmatrix} \) is the sample and \( Z_{II} = \begin{pmatrix} Y^{(0)r}_{II} \\ Y^{(1)r}_{II} \end{pmatrix} \) is the remaining part. It immediately follows that

\[
E(Z_I) = (I_2 \otimes 1_n) \begin{pmatrix} \mu^{(0)}_t \\ 0 \end{pmatrix}, \quad E(Z_{II}) = (I_2 \otimes 1_{N-n}) \begin{pmatrix} \mu^{(0)}_{II} \\ 0 \end{pmatrix}, \quad V_I = \Sigma \otimes P_{n,N} \text{ and } V_{II} = V_{II,I} = \Sigma \otimes \left( -\frac{1}{N} J_{n \times (N-n)} \right). \]

5.3. Estimating \( T^{(0)} \) and \( \mu^{(0)} \) with a known auxiliary population total

In this Section, we will derive an estimator of response population total (and mean) using the method outlined in Section 3.5. of Chapter 3. First, we briefly reiterate the research question and then derive the estimator.

5.3.1. Parameters of interest and their linear estimators

The parameter of interest is the population total of the response variable,

\[ T^{(0)} = l'_N Y^{(0)} = l' Z', \quad (5.3) \]

where \( l = (1'_N \ 0_{1 \times N})' \). We are interested in deriving an estimator for \( T^{(0)} \) that is a linear function of the sample data and unbiased for \( T^{(0)} \), and has minimum mean squared error. After partitioning, \( T^{(0)} \) can be represented as a linear combination of the sampled and remaining parts,

\[ T^{(0)} = l'_I Z_I + l'_II Z_{II}, \quad (5.4) \]
where \( I_I = (1 \ 0) \otimes 1_n \) and \( I_{II} = (1 \ 0) \otimes 1_{N-n} \).

The linear estimator of \( T^{(0)} \) is defined as linear function of \( Z_I \),

\[
\hat{T}^{(0)} = (I_I' + w')Z_I, \tag{5.5}
\]

where \( w = (w_0' \ w_i')' \) is a \( 2n \times 1 \) vector.

5.3.2. Unbiasedness constraint and optimization criterion

As shown in Section 3.5, the unbiasedness for \( \hat{T}^{(0)} \) implies constraint (3.24), or namely,

\[
w_0'1_n - (N-n) = 0. \tag{5.6}
\]

Under this constraint, the mean square error of \( \hat{T}^{(0)} \) is the same as it variance,

\[
\text{var} (\hat{T}^{(0)}) = (w' - I_{II'}) \begin{pmatrix} V_I & V_{II} \\ V_{II,I} & V_{II} \end{pmatrix} \begin{pmatrix} w \\ -I_{II} \end{pmatrix}. \tag{5.7}
\]

To minimize (3.25) subject to (3.24), we define the optimization function as

\[
\Phi(w) = (w_0' \ w_i')V_I \begin{pmatrix} w_0 \\ w_i \end{pmatrix} - 2(w_0' \ w_i')V_{II,I}I_{II} + 2(w_0'1_n - (N-n))\hat{\lambda}, \tag{5.8}
\]

where \( \hat{\lambda} \) is a Lagrangian multiplier, and \( V_I \) is the expected variance-covariance matrix of the sample. Differentiating (3.26) and setting the derivatives to zero leads to the following estimating equations,

\[
\begin{pmatrix} \Sigma \otimes P_{n,N} \\ (1 \ 0) \otimes 1_n' \end{pmatrix} \begin{pmatrix} \hat{w}_0 \\ \hat{w}_i \end{pmatrix} = \begin{pmatrix} \Sigma \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ (1 - f) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes 1_n \end{pmatrix},
\]

where \( \Sigma \) is given in Section 2.2. This leads to the unique solution,
\[
\begin{bmatrix}
\hat{w}_0 \\
\hat{w}_1
\end{bmatrix} = -\begin{bmatrix} 1_n \\
0
\end{bmatrix} + \frac{N}{n} \begin{bmatrix} 1_n \\
-\beta_{01} 1_n
\end{bmatrix},
\]

where $\beta_{01} = \sigma_{01}/\sigma_1^2$ is the linear regression coefficient of $y^{(0)}$ on $y^{(1)}$ in the finite population (Cochran 1977). Consequently, the estimator for $T^{(0)}$ is,

\[
\hat{T}^{(0)} = \left( I_n' + \left( \hat{w}_0 \hat{w}_1 \right) \right) \begin{bmatrix} Y_i^{(0)} \\
Y_i^{(1)}
\end{bmatrix}
\]

\[
= 1_n' Y_i^{(0)} + (N - n) \left[ \frac{1_n'}{n} Y_i^{(0)} - \frac{N}{N - n} \beta_{01} \left( \frac{1_n'}{n} Y_i^{(1)} \right) \right]
\]

\[
= n \bar{Y}^{(0)} + (N - n) \left( \bar{Y}^{(0)} - \beta_{01} \left( \bar{Y}^{(1)} - \mu^{(1)} \right) \right) - n \beta_{01} \left( \bar{Y}^{(1)} - \mu^{(1)} \right),
\]

where $\bar{Y}^{(0)} = \frac{1}{n} 1_n' Y_i^{(0)}$ and $\bar{Y}^{(1)} = \frac{1}{n} 1_n' Y_i^{(1)}$ are the sample means of the response and auxiliary variables, respectively. Let us write (5.9) alternatively as,

\[
\hat{T}^{(0)} = N \beta_{01} \mu^{(1)} + N \left( \bar{Y}^{(0)} - \beta_{01} \bar{Y}^{(1)} \right).
\]

Identity (5.10) implies that, in estimating $T^{(0)}$, we would be estimating $T^{(0)}$ as a constant $\left( N \beta_{01} \mu^{(1)} \right)$ plus the finite population estimating function

\[
N \left( \mu^{(0)} - \beta_{01} \mu^{(1)} \right),
\]

for which the best sample-based estimating function is

\[
N \left( \bar{Y}^{(0)} - \beta_{01} \bar{Y}^{(1)} \right).
\]

The variance of $\hat{T}^{(0)}$ is

\[
\text{var} \left( \hat{T}^{(0)} \right) = N^2 \frac{1 - f}{n} \sigma_0^2 \left( 1 - \rho_{01}^2 \right),
\]
where \( \rho_{01} = \sqrt{\sigma_{01}^2 / \sigma_0^2 \sigma_1^2} \) is the population correlation coefficient between the response and auxiliary variable.

Similarly, the estimator of the mean for the response variable and its variance are

\[
\hat{\mu}^{(0)} = \bar{y}^{(0)} - \beta_{01} \left( \bar{y}^{(1)} - \hat{\mu}^{(1)} \right)
\]

and

\[
\text{var}(\hat{\mu}) = \frac{1 - f}{n} \sigma_0^2 \left(1 - \rho_{01}^2\right),
\]

respectively.

Derivations of the above results are given in Section 5.7.

5.3.3. Special case: \( \mathbf{w} = (\mathbf{w}_0' \ 0)' \)

A special case is when \( \mathbf{w}_1 \) is forced to be a null vector, such that \( \mathbf{w} = (\mathbf{w}_0' \ 0)' \).

Using a similar derivation as above, it is shown that, \( \hat{\mathbf{w}}_0 = \frac{N - n}{n} \mathbf{1}_n \), which is the weight of the simple expansion estimator. Therefore, the estimator for \( T^{(0)} \) when \( \mathbf{w} = (\mathbf{w}_0' \ 0)' \) is \( \hat{T}^{(0)} = N \bar{y}^{(0)} \), and its variance is \( \text{var}(\hat{T}^{(0)}) = N^2 \left(\frac{1 - f}{n}\right) \sigma_0^2 \).

5.3.4. Remarks

Unlike the usual calibration technique of incorporating auxiliary information, this approach leads to a set of weights that do not depend on sample, and yields a linear unbiased minimum variance estimator of \( T^{(0)} \). Estimator (5.10) is seen to be a generalized regression estimator (GREG); this has been shown to be both design- and model-unbiased (Särndal, Swensson and Wretman 1989; Särndal, Swensson and
That is, we have derived under a design-based framework an estimator that is exactly the same as a GREG estimator. This estimator depends on neither a superpopulation model nor an assumption about a regression model about the response and auxiliary variables.

5.4. **Estimating $T^{(0)}$ and $\mu^{(0)}$ with several known auxiliary totals**

In this section, we extend the results in Section 5.3 to scenarios with multiple auxiliary variables. Similar to Model (5.1), we define a random permutation seemingly unrelated regression model that includes $p$ auxiliary variables,

$$
\begin{pmatrix}
Y^{(0)} \\
Y^{(*)}
\end{pmatrix} = (I_p \otimes I_N) \begin{pmatrix}
\mu^{(0)} \\
\mu^{(*)}
\end{pmatrix} + E,
$$

(5.14)

where $Y^{(*)} = \left( Y^{(1)^T} \quad Y^{(2)^T} \quad \cdots \quad Y^{(p)^T} \right)^T$ is the vector of all $p$ auxiliary variables and $\mu^{(*)} = \left( \mu^{(1)} \quad \mu^{(2)} \quad \cdots \quad \mu^{(p)} \right)^T$ is the vector of auxiliary means.

Since the vector of auxiliary totals $T^{(*)} = \left( T^{(1)} \quad T^{(2)} \quad \cdots \quad T^{(p)} \right)^T$ (and thus $\mu^{(*)}$) is known, the $N$ random variables arising from permuting the $k$-th auxiliary variable have a constraint $I_N' Y^{(k)} = T^{(k)}$, where $k = 1, 2, \ldots, p$. We transform $Y^{(*)}$ by pre-multiplying $I_p \otimes \left( \frac{1_N}{N} I_N' + P_N \right)$, such that,

$$
Y^{(*)} = I_p \otimes \left( \frac{1_N}{N} I_N' \right) Y^{(*)} + \left( I_p \otimes P_N \right) Y^{(*)},
$$

(5.15)
The first component of (5.15) is a vector of constants equal to \( \left( I_p \otimes \frac{1_N}{N} \right) T^{(*)} \). We denote the second component of (5.15) as

\[
Y^{(*)^*} = \left( I_p \otimes P_N \right) Y^{(*)} = Y^{(*)} - \left( I_p \otimes I_N \right) \mu^{(*)}.
\]

This transformation can be represented jointly in terms of both \( Y^{(0)} \) and \( Y^{(*)} \),

\[
\begin{bmatrix}
I_N & 0 \\
0 & I_p \otimes P_N
\end{bmatrix}
\begin{bmatrix}
0 & 0 \\
0 & I_p \otimes \left( \frac{1_N}{N} I_N \right)
\end{bmatrix}
\begin{bmatrix}
Y^{(0)} \\
Y^{(*)^*}
\end{bmatrix}
= \begin{bmatrix}
0 \\
0
\end{bmatrix}
+ \begin{bmatrix}
\left( I_p \otimes \frac{1_N}{N} \right) T^{(*)}
\end{bmatrix}
\]

After this transformation, Model (5.14) becomes,

\[
\begin{bmatrix}
Y^{(0)} \\
Y^{(*)^*}
\end{bmatrix}
= \begin{bmatrix}
I_{p+1} \otimes I_N
\end{bmatrix}
\begin{bmatrix}
\mu^{(0)} \\
0
\end{bmatrix}
+ E.
\]

(5.16)

As shown in Chapter 2, Equation A2.1.15, \( \text{cov} \begin{bmatrix}
Y^{(0)} \\
Y^{(*)^*}
\end{bmatrix}
= \Sigma \otimes P_N \), where

\[
\Sigma = \begin{bmatrix}
\sigma_0^2 & \sigma_{01} & \cdots & \sigma_{0p} \\
\sigma_{10} & \sigma_1^2 & \cdots & \sigma_{1p} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{p0} & \sigma_{p1} & \cdots & \sigma_p^2
\end{bmatrix}
= \begin{bmatrix}
\sigma_0^2 & \sigma_{0*} \\
\sigma_{0*} & \Sigma^{(*)}
\end{bmatrix}.
\]

After partitioning into sample and remaining parts, \( Z = \left( Z'_{\mathcal{I}} \quad Z'_{\mathcal{II}} \right)' \), where

\[
Z_{\mathcal{I}} = \begin{bmatrix}
Y_{\mathcal{I}}^{(0)*} & Y_{\mathcal{I}}^{(*)^*}
\end{bmatrix}'
\]
is the sample and \( Z_{\mathcal{II}} = \begin{bmatrix}
Y_{\mathcal{II}}^{(0)*} & Y_{\mathcal{II}}^{(*)^*}
\end{bmatrix}' \) is the remaining part. It immediately follows that

\[
E \left( Z_{\mathcal{I}} \right) = \left( I_{p+1} \otimes I_n \right) \begin{bmatrix}
\mu^{(0)} \\
0
\end{bmatrix}
\]
and

\[
E \left( Z_{\mathcal{II}} \right) = \left( I_{p+1} \otimes I_{N-n} \right) \begin{bmatrix}
\mu^{(0)} \\
0
\end{bmatrix}.
\]

\[
V_{\mathcal{I}} = \begin{bmatrix}
\sigma_0^2 & \sigma_{0*} \\
\sigma_{0*} & \Sigma^{(*)}
\end{bmatrix} \otimes P_{n,N} \quad \text{and} \quad V_{\mathcal{II},\mathcal{I}} = V_{\mathcal{II},\mathcal{I}}' = \begin{bmatrix}
\sigma_0^2 & \sigma_{0*} \\
\sigma_{0*} & \Sigma^{(*)}
\end{bmatrix} \otimes \begin{bmatrix}
-1 \\
\frac{1}{N} J_{n \times (N-n)}
\end{bmatrix}.
\]

75
The population total of the response variable, the population parameter of our primary interest, can be defined as,

\[ T^{(0)} = \mathbf{1}'_N \mathbf{Y}^{(0)} = \mathbf{l}' \begin{pmatrix} \mathbf{Y}^{(0)} \\ \mathbf{y}^{(*)} \end{pmatrix} = \mathbf{l}' \begin{pmatrix} \mathbf{Y}^{(0)} \\ \mathbf{y}^{(*)} \end{pmatrix}, \tag{5.17} \]

where \( \mathbf{l} = \left( \mathbf{1}'_N \left( \mathbf{1}'_p \otimes \mathbf{0}_{1 \times N} \right) \right)' \). After partitioning, \( T^{(0)} \) can be represented as a linear combination of the sampled and remaining parts,

\[ T^{(0)} = \mathbf{l}'_{I} \mathbf{Z}_{I} + \mathbf{l}'_{II} \mathbf{Z}_{II}, \]

where \( \mathbf{l}_{I} = \left( \mathbf{1} - \mathbf{0}_{1 \times p} \right)' \otimes \mathbf{1}_n \) and \( \mathbf{l}_{II} = \left( \mathbf{1} - \mathbf{0}_{1 \times p} \right)' \otimes \mathbf{1}_{N-n} \).

We derive an estimator for \( T^{(0)} \) that is a linear function of \( \mathbf{Z}_I \), unbiased for \( T^{(0)} \) and has minimum variance. The linear estimator of \( T^{(0)} \) is defined as linear function of the sample,

\[ \hat{T}^{(0)} = \left( \mathbf{l}'_I + \mathbf{w}' \right) \mathbf{Z}_I, \tag{5.18} \]

where \( \mathbf{w} = \left( \mathbf{w}'_0 \mathbf{w}'_1 \ldots \mathbf{w}'_p \right)' \) is a \( pn \times 1 \) vector of weights, and \( \mathbf{w}_* = \left( \mathbf{w}'_1 \mathbf{w}'_2 \ldots \mathbf{w}'_p \right)' \).

The unbiasedness for \( T^{(0)} \) requires that \( E(\mathbf{w}' \mathbf{Z}_I - \mathbf{l}'_I \mathbf{Z}_{II}) = 0 \). Since

\[ E(\mathbf{Z}_I) = \left( \mathbf{1}_{p+1} \otimes \mathbf{1}_n \right) \left( \mu^{(0)} \mathbf{0} \right)' \], \( E(\mathbf{Z}_{II}) = \left( \mathbf{1}_{p+1} \otimes \mathbf{1}_{N-n} \right) \left( \mu^{(0)} \mathbf{0} \right)' \),

this constraint is equivalent to

\[ \left( \mathbf{1}'_n \mathbf{w}_0 - (N-n) \right) \mu^{(0)} + \sum_{k=1}^p \mathbf{1}'_n \mathbf{w}_{(k)} \mu^{(k)*} = 0. \tag{5.19} \]

Since \( \mu^{(k)*} = 0 \) for any \( k = 1,2,\ldots, p \), (5.19) is satisfied for any \( \mu^{(0)} \) simply by

\[ \mathbf{1}'_n \mathbf{w}_0 - (N-n) = 0. \tag{5.20} \]
Similar to (3.25), the variance of \( T^{(0)} \) can be represented in terms of partitioned matrices that include multiple auxiliary totals,

\[
\operatorname{var}\left(\hat{T}^{(0)}\right) = \mathbf{w}' \mathbf{V}_I \mathbf{w} - 2 \mathbf{w}' \mathbf{V}_{I,II} \mathbf{l}_II + \mathbf{L}'_{II} \mathbf{V}_{II} \mathbf{l}_{II}.
\]

where \( \mathbf{w} = \left( \mathbf{w}_0' \quad \mathbf{w}_1' \quad \cdots \quad \mathbf{w}_p' \right)' \) is an \( n(p+1) \times 1 \) vector, \( \mathbf{l}_{II} = \left( \mathbf{1} \quad \mathbf{0}_{1 \times p} \right)' \otimes \mathbf{1}_{N-n} \). Under unbiasedness constraint (5.20), the Lagrangian function is defined as

\[
\Phi (\mathbf{w}) = \mathbf{w}' \mathbf{V}_I \mathbf{w} - 2 \mathbf{w}' \mathbf{V}_{I,II} \mathbf{l}_{II} + 2 \left( \mathbf{w}_0' \mathbf{1}_n - (N-n) \right) \lambda,
\]

where \( \lambda \) is a Lagrangian multiplier. Differentiating (5.21) with respect to \( \mathbf{w} \) and \( \lambda \) results in the following estimating equations,

\[
\begin{pmatrix}
\mathbf{V}_I \\
\mathbf{u}_i' \otimes \mathbf{1}_n \\
0
\end{pmatrix}
\begin{pmatrix}
\mathbf{w} \\
\lambda
\end{pmatrix}
= \begin{pmatrix}
\mathbf{V}_{I,II} \left( \mathbf{u}_i \otimes \mathbf{1}_{N-n} \right) \\
N-n
\end{pmatrix},
\]

(5.22)

where \( \mathbf{u}_i = \left( \mathbf{1} \quad \mathbf{0}_{1 \times p} \right)' \) is the first column of identity matrix \( \mathbf{I}_{p+1} \). The unique solution to (5.22) is

\[
\hat{\mathbf{w}} = -\mathbf{u}_i \otimes \mathbf{1}_n + \frac{N}{n} \left( \mathbf{\Sigma}^{-1} \mathbf{u}_i \left( \mathbf{u}_i' \mathbf{\Sigma}^{-1} \mathbf{u}_i \right)^{-1} \right) \otimes \mathbf{1}_n
\]

\[
= -\mathbf{u}_i \otimes \mathbf{1}_n + \frac{N}{n} \left( \frac{1}{-\mathbf{\beta}_{0*}} \right) \otimes \mathbf{1}_n,
\]

(5.23)

where \( \mathbf{\beta}_{0*} = \mathbf{\Sigma}^{-1} \mathbf{\sigma}_{0*} \). Consequently,

\[
\hat{T}^{(0)} = N \bar{\mathbf{Y}}^{(0)} - N \mathbf{\beta}_{0*}' \left( \mathbf{\bar{Y}}^{(*)} - \bar{\mathbf{\mu}}^{(*)} \right),
\]

\[
\operatorname{var}\left(\hat{T}^{(0)}\right) = \left(1-\rho_{0*}^2\right) \left(N^2 \left(\frac{1-f}{n}\right) \sigma_{0*}^2\right).
\]

where \( \mathbf{\bar{Y}}^{(*)} \) is a \( p \times 1 \) column vector of sample means of the \( p \) auxiliary variables, \( \bar{\mathbf{\mu}}^{(*)} \) is the column vector of known auxiliary means, and \( \rho_{0*} = \sqrt{\sigma_{0*} \mathbf{\Sigma}_{0*}^{-1} \mathbf{\sigma}_{0*} / \sigma_{0*}^2} \) is the
multiple correlation coefficient of $Y^{(0)}$ on $Y^{(*)}$. This expression is similar to those expressions commonly seen in multiple linear regression models (Graybill 1976), except its finite population correction factor. It is convenient to show

$$\hat{\mu}_0 = \bar{Y}^{(0)} - \beta_0^* \left( \bar{Y}^{(*)} - \mu^{(*)} \right), \quad (5.24)$$

$$\text{var} \left( \hat{\mu}_0 \right) = \left( 1 - \rho_{0(0)}^2 \right) \left( \frac{1 - f}{n} \sigma_0^2 \right). \quad (5.25)$$

Detailed derivations of the above results are summarized in Section 5.8.

5.5. Estimation of variance and confidence intervals

The estimators derived in this Chapter depend on the knowledge of the covariance vector $\sigma_{00}$ between the response variable and the auxiliary variables, and the variance-covariance matrix $\Sigma_0$ of auxiliary variables. They are often unknown to surveyors in practice.

1) When auxiliary values for all units in the population are known, $\Sigma_0$ is known, we only need to approximate $\sigma_{00}$.

2) When individual auxiliary values are only known for the sampled units, we may need to approximate both $\Sigma_0$ and $\sigma_{00}$.

The problem of associating standard errors and confidence intervals with the estimators can be assessed in several ways. One commonly used approach is to replace $\sigma_{00}$ and $\Sigma_0$ with their sample estimates, which is referred as methods of moments in literature.
5.5.1. Normal approximation

To compute the variance of the estimators (5.10) and (5.24), one needs to know \( \sigma^2_0 \), \( \sigma_{01} \) (or \( \sigma_{0*} \)), and \( \sigma^2_1 \) (or \( \Sigma_\cdot \)). Using the method of moments, one may simply use their sample estimates, i.e.,

\[
\hat{\sigma}^2_0 = \frac{1}{n-1} Y_i^{(0)'} \left( I_n - \frac{1}{n} J_n \right) Y_i^{(0)},
\]

\[
\hat{\sigma}_{01} = \frac{1}{n-1} Y_i^{(0)'} \left( I_n - \frac{1}{n} J_n \right) Y_i^{(1)},
\]

\[
\hat{\sigma}^2_1 = \frac{1}{n-1} Y_i^{(1)'} \left( I_n - \frac{1}{n} J_n \right) Y_i^{(1)}.
\]

However, if \( \sigma^2_1 \) (or \( \Sigma_\cdot \)) is provided or auxiliary values are known for all subjects, one only need to use plug-in estimates for \( \sigma^2_0 \) and \( \sigma_{01} \) (or \( \sigma_{0*} \)). For example, using methods of moment estimates for \( \hat{\sigma}^2_0 \) and \( \hat{\sigma}_{01} \), (5.10) leads to

\[
\hat{T}^{(0)} = N \left( \bar{Y}^{(0)} + \hat{\beta}_{01} \left( \mu^{(1)} - \bar{Y}^{(1)} \right) \right),
\]

where \( \hat{\beta}_{01} = \frac{\sigma^2_{01}}{\sigma^2_1} \) if \( \sigma^2_1 \) is unknown or \( \hat{\beta}_{01} = \hat{\sigma}_{01} / \sigma^2_1 \) if \( \sigma^2_1 \) is known. The estimated variance of \( \hat{T}^{(0)} \) is,

\[
\text{var} \left( \hat{T}^{(0)} \right) = N^2 \frac{1 - f}{n} \hat{\sigma}^2_0 \left( 1 - \hat{\rho}_{01}^2 \right),
\]

where \( \hat{\rho}_{01}^2 = \frac{\sigma^2_{01}}{\sigma^2_0 \sigma^2_1} \) if \( \sigma^2_1 \) is unknown or \( \hat{\rho}_{01}^2 = \frac{\sigma^2_{01}}{\sigma^2_0 \sigma^2_1} \) if \( \sigma^2_1 \) is known. An *ad hoc* confidence interval based on appropriate normal-based prediction interval for \( \hat{T}^{(0)} \) would be

\[
N \bar{Y}^{(0)} \pm z_{1-\alpha} \sqrt{\text{var} \left( \hat{T}_0 \right)}.
\]
5.5.2. Jackknife variance

Another widely recommended variance estimation procedure is Jackknife variance estimation (Royall and Cumberland 1978; Wolter 1985; Särndal, Swensson and Wretman 1992; Stukel, Hidiroglou and Särndal 1996; Duchesne 2000; Valliant 2002). Suppose we are interested in estimating the variance of \( \hat{\mu}_0 \) using (5.12); the process of Jackknife variance estimation is as follows:

1) Partition the sample into \( B \) random groups of equal size \( m = n/B \), assuming each of the \( m \) groups is a SRSWOR subsample from the sample;

2) For each group, compute a pseudo estimator \( \left( \hat{\mu}^{(0)}_{(b)}, b = 1, 2, \ldots, m \right) \) of \( \mu^{(0)} \) with the same function form as (5.12) based on the remaining units of the sample after omitting the units in the group, \( \hat{\mu}^{(0)}_{(b)} = \bar{Y}^{(0)}_{(b)} - \beta_{01} \left( \bar{Y}^{(1)}_{(b)} - \mu^{(1)} \right) \).

We then define \( \hat{\mu}^{(b)} = m\hat{\mu}^{(0)} - (m-1)\hat{\mu}^{(0)}_{(b)} \), where \( \hat{\mu}^{(0)} \) is computed with (5.12) based on full sample.

Then the Jackknife estimator of \( \mu^{(0)} \) is

\[
\hat{\mu}^{(jk)} = \frac{1}{m} \sum_{b=1}^{m} \hat{\mu}^{(b)}
\]

and the Jackknife variance estimator is defined as

\[
\hat{V}_{jk} = \frac{1}{m(m-1)} \sum_{b=1}^{m} \left( \hat{\mu}^{(b)} - \hat{\mu}^{(jk)} \right)^2.
\]

An appropriate normal-based \((1-\alpha) \times 100\% \) confidence interval for \( \hat{T}^{(0)} \) would be

\[
N \hat{\mu}_0 \pm z_{1-\alpha} \sqrt{\hat{V}_{jk} \left( T^{(0)} \right)}.
\]
5.5.3. Bootstrap variance estimator and confidence intervals

Standard error and confidence intervals can be also computed using Bootstrap method (Efron 1979; Efron and Tibshirani 1993). To apply bootstrap method in finite sampling, modifications must be made (McCarthy and Snowden 1985). Suppose that a sample \( Z_I = \left( Y_I^{(0)r} \ Y_I^{(1)r} \right)' \) has been drawn and observed. A bootstrap sample

\[
Z^b_I = \left( Y_I^{(0)b} \ Y_I^{(1)b} \right)' \]

is obtained by randomly sampling with replacement \( n^* \) times from \( Z_I \), where \( n^* = (n - 1)/(1 - f) \), \( f = n/N \) (McCarthy and Snowden 1985). In matrix format, each bootstrap sampling process may be represented as

\[
Z^b_I = (I_2 \otimes Q^b) Z_I,
\]

where \( Q^b = \begin{pmatrix} u_{m_1} & u_{m_2} & \cdots & u_{m_i} \end{pmatrix} \), \( u_{m_i} \) is any \( m_i \)-th column of identity matrix of \( I_n \), \( i = 1, 2, \ldots, n^* \) and \( 1 \leq m_i \leq n^* \). The bootstrap process begins by generating a large number \( (B) \) of independent bootstrap samples \( Z^1_I, Z^2_I, \ldots, Z^b_I, \ldots, Z^B_I \). For each bootstrap replication \( Z^b_I, b = 1, 2, \ldots, B \), the estimator \( T^{(0)b} \) is calculated using (5.10) while replacing \( \beta_{01}, \bar{Y}^{(0)} \) and \( \bar{Y}^{(1)} \) with the estimates from \( Z^b_I \). The bootstrap estimate of variance of \( \hat{T}^{(0)} \) is the variance of the bootstrap replications,

\[
\hat{V}^b_\text{es}(T^{(0)}) = \frac{1}{B-1} \sum_{b=1}^{B} \left( T^{(0)b} - \bar{T}^{(0)b} \right)^2,
\]

where \( T^{(0)b} = B^{-1} \sum_{b=1}^{B} T^{(0)b} \). To construct a \( 100(1 - \alpha)% \) confidence intervals, one may use \( \hat{V}^b_\text{es}(T^{(0)}) \) and normal approximation, such that

\[
\hat{T}^{(0)} \pm z^{(1-\alpha)} \sqrt{\hat{V}^b_\text{es}(T^{(0)})}.
\]
The method by McCarthy and Snowden (1985) has a weakness that
\[ n^* = \frac{n - 1}{1 - f} \]
needs to be approximated if it is not an integer, the impacts of such
approximation on estimation have not investigated. Rao and Wu (1988) developed a
method to the bootstrap estimation of the standard error of a population mean, which
overcomes the non-integer \( n^* \) issue of McCarthy-Snowden method.

5.6. Remarks

The estimators presented in Section 5.3 and 5.4 are identical to the estimator by
(Cochran 1977) (page 191-192), and GREG estimators derived from model-assisted
approach (Särndal, Swensson and Wretman 1992)(page 273). In addition, the estimators
and their variances derived in Section 5.3 and 5.4 are the same as the asymptotically
design-unbiased estimator and its asymptotic design-variance derived from a model-
calibration approach based on a superpopulation framework (Wu and Sitter 2001).

They have a functional form similar to those of calibration estimators (Deville
and Särndal 1992) and estimators derived from model-based prediction approach
(Valliant, Dorfman and Royall 2000)(page 32); (Bolfarine and Zacks 1992) (page37).
When the variance-covariance of the auxiliary variables \( \Sigma_r \) and the covariance
between response and auxiliary variables \( \sigma_{ri} \) are unknown, we apply these results by
substituting their sample estimates (method of moments). The resulting estimators will
be the same as usual calibration estimators (Deville and Särndal 1992).

However, our results do not rely on model assumptions that are necessary for
superpopulation models and model-assisted approaches. In addition, these estimators
and their variance are exact but not approximated.
These estimators require the knowledge of both $\sigma_{01}$ and $\sigma_1^2$, and their variances are always smaller than variance of simple expansion estimator. The gain in the precision of prediction is proportional to the squared (multiple) correlation coefficient of the response variable and auxiliary variable(s).

5.7. Derivations of results in Section 5.3.

As shown in Section 5.2, the reparameterized model is

$$\begin{pmatrix} Y^{(0)} \\ Y^{(1)*} \end{pmatrix} = \left( I_2 \otimes I_n \right) \begin{pmatrix} \mu^{(0)} \\ 0 \end{pmatrix} + E.$$  

The parameter of interest is population total of the response variable that is defined as,

$$T^{(0)} = l'Z = \begin{pmatrix} 1_N' & 0 \\ Y^{(0)} & Y^{(1)*} \end{pmatrix},$$

where $l = (1_N' \ 0_{1 \times N})'$. After partitioning, we denote the partitioned vector as $Z$, and the sample and remaining parts as $Z_I$ and $Z_{II}$ respectively, such that $Z = (Z_I' \ Z_{II}')'$. Thus, $T^{(0)}$ can be expressed as $T^{(0)} = l_I'Z_I + l_{II}'Z_{II}$, where $l_I' = (1 \ 0) \otimes 1_n'$ and $l_{II}' = (1 \ 0) \otimes 1_{N-n}'$.

We define a linear estimator for $T^{(0)}$ as a linear function of $Z_I$, such that

$$\hat{T}^{(0)} = (l_I' + w')Z_I,$$

where $w$ can be defined in two possible ways, i.e., $w = (w_0' \ w_1')'$ and $w = (w_0' \ 0)'$, $w_0$ and $w_1$ are vectors of weights corresponding to the response and auxiliary variable.
Scenario 1: \( \mathbf{w} = (\mathbf{w}_0' \quad \mathbf{w}_1')' \)

The unbiasedness for \( T^{(0)} \) requires \( E(\mathbf{w}' Z_I - I_n' \mathbf{Z}_II) = 0 \), equivalently,
\[
(1_n' \mathbf{w}_0 - (N-n))\mu^{(0)} + 1_n' \mathbf{w}_i \mu^{(i)*} = 0.
\]
Since \( \mu^{(i)*} = 0 \), this constraint is satisfied for any \( \mu^{(0)} \) simply by \( \mathbf{w}_0' 1_n = N-n \). Therefore, the optimization function is
\[
\Phi = (\mathbf{w}_0' \quad \mathbf{w}_1') \begin{pmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \end{pmatrix} - 2(\mathbf{w}_0' \quad \mathbf{w}_1') \mathbf{V}_{I,II} 1_n + 2(\mathbf{w}_0' 1_n - (N-n)) \lambda,
\]
which can be simplified to
\[
\Phi = \sigma_0^2 \mathbf{w}_0' \mathbf{P}_{n,N} \mathbf{w}_0 + 2\sigma_{01} \mathbf{w}_1' \mathbf{P}_{n,N} \mathbf{w}_0 + \sigma_1^2 \mathbf{w}_1' \mathbf{P}_{n,N} \mathbf{w}_1
\]
\[
+ 2(1-f)(\sigma_0^2 \mathbf{w}_0' 1_n + \sigma_{01} \mathbf{w}_1' 1_n) + 2(\mathbf{w}_0' 1_n - N(1-f)) \lambda
\]
Differentiating (5.26) with respect to \( \mathbf{w}_0 \), \( \mathbf{w}_1 \) and \( \lambda \) yields following estimating equations,
\[
\begin{pmatrix}
\sigma_0^2 \mathbf{P}_{n,N} & \sigma_{01} \mathbf{P}_{n,N} & 1_n \\
\sigma_{01} \mathbf{P}_{n,N} & \sigma_1^2 \mathbf{P}_{n,N} & 0 \\
1_n' & 0 & 0
\end{pmatrix}
\begin{pmatrix}
\mathbf{w}_0 \\
\mathbf{w}_1 \\
\lambda
\end{pmatrix} = \begin{pmatrix}
-(1-f)\sigma_0^2 1_n \\
-(1-f)\sigma_{01} 1_n \\
N(1-f)
\end{pmatrix},
\]
or
\[
\begin{pmatrix}
\Sigma \otimes \mathbf{P}_{n,N} & 1 \otimes 1_n \\
1 \otimes 1_n' & 0
\end{pmatrix}
\begin{pmatrix}
\mathbf{w}_0 \\
\mathbf{w}_1 \\
\lambda
\end{pmatrix} = \begin{pmatrix}
-(1-f)\Sigma 1_n \\
N(1-f)
\end{pmatrix}.
\]
Solving these estimating equations leads to
\[
\hat{\lambda} = -(1 - f) \left\{ (1 - 0) \Sigma^{-1} \left( \begin{array}{c} 1 \\ 0 \end{array} \right) \otimes I_n, P_{n,n}^{-1} 1_n \right\} - N \left\{ (1 - 0) \Sigma^{-1} \left( \begin{array}{c} 1 \\ 0 \end{array} \right) \otimes \left( I_n' P_{n,n}^{-1} 1_n \right) + N \right\} \\
= -n \frac{N - n}{n} \left( \begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array} \right) \Sigma^{-1} \left( \begin{array}{c} 1 \\ 0 \end{array} \right) \\
= -n \frac{N - n}{n} \left( \sigma_0^2 - \frac{\sigma_{01}^2}{\sigma_1^2} \right)
\]

\[
\left( \hat{w}_0, \hat{w}_1 \right) = -\left( \begin{array}{c} 1 \\ 0 \end{array} \right) \otimes I_n + \frac{N}{n} \left( 1 - \frac{\sigma_{01}^2}{\sigma_1^2} \right) \otimes I_n = \frac{N}{n} \left( 1 - \frac{\sigma_{01}^2}{\sigma_1^2} \right) \otimes I_n
\]

\[
\hat{t}^{(0)} = \left( I_n' + \left( w_0' \quad w_1' \right) \right) Z_l = \left\{ \left( I_n' \quad 0 \right) - \left( I_n' \quad 0 \right) + \frac{N}{n} \left( I_n' \quad - \frac{\sigma_{01}^2}{\sigma_1^2} I_n' \right) \right\} \left( \begin{array}{c} Y_l^{(0)} \\ Y_l^{(1)*} \end{array} \right)
\]

\[
= \frac{N}{n} \left( 1' Y_l^{(0)} - \frac{\sigma_{01}^2}{\sigma_1^2} 1' Y_l^{(1)*} \right)
\]

\[
= N \left( \bar{y}_0 - \frac{\sigma_{01}^2}{\sigma_1^2} (\bar{y}_1 - \mu_1) \right)
\]

\[
\text{var} \left( \hat{t}^{(0)} \right) = \left( I_n' + \left( w_0' \quad w_1' \right) \right) \text{cov} (Z_l) \left( I_n' + \left( w_0' \quad w_1' \right) \right)
\]

\[
= \left( \frac{N}{n} \right)^2 \left\{ (1 - \beta_{01}) \otimes I_n' \right\} \left( \Sigma \otimes P_{n,n} \right) \left\{ \frac{1}{n} \right\} \otimes 1_n \right\} = \frac{N(N - n)}{n} \left\{ \sigma_0^2 - \beta_{01}^2 \right\}
\]

\[
= N^2 \left( \frac{1 - f}{n} \right) (1 - \rho_{01}^2) \sigma_0^2,
\]

where $\beta_{01} = \sigma_{01}/\sigma_1^2$ and $\rho_{01} = \sqrt{\frac{\sigma_{01}^2}{\sigma_0^2 \sigma_1^2}}$ is the correlation coefficient between the response and auxiliary variable.
Scenario 2: \( w = \left( w'_0 \quad 0 \right)' \)

The unbiasedness for \( T_0 \) requires \( E \left( w'Z_f - t'_0 Z_{0f} \right) = 0 \), equivalently,

\[
\left( 1'_n w'_0 - (N - n) \right) \mu^{(0)} + 1'_n 0 \mu^{(1)} = 0.
\]

This constraint is satisfied for any \( \mu_0 \) simply by \( w'_0 1_n = N - n \). Therefore, the optimization function is

\[
\Phi = w'_0 \left( 1_n \quad 0 \right) V_f \left( 1_n \quad 0 \right) w'_0 - 2w'_0 \left( 1_n \quad 0 \right) V_{f, 0} I_n + 2 \left( w'_0 1_n - (N - n) \right) \lambda,
\]

which can be simplified to

\[
\Phi = \sigma^2_0 w'_0 P_{n, N} w_0 - 2(1 - f) \sigma^2_0 w'_0 1_n + 2 \left( w'_0 1_n - (N - n) \right) \lambda \quad (5.27)
\]

Differentiating (5.26) with respect to \( w_0 \) and \( \lambda \) and setting the derivatives to zero yields following estimating equations,

\[
\left( \frac{\sigma^2_0 P_{n, N}}{1'_n} \right) \left( \begin{array}{c} 1_n \\ 0 \end{array} \right) \lambda = \left( \begin{array}{c} -(1 - f) \sigma^2_0 1_n \\ N(1 - f) \end{array} \right),
\]

Solving these estimating equations leads to

\[
\hat{w}_0 = \frac{N - n}{n} 1_n,
\]

which is the weight of simple expansion estimator. Therefore, the estimator for \( T^{(0)} \) is

\[
\hat{T}^{(0)} = \frac{N}{n} 1'_n Y_t^{(0)} = \frac{N}{n} \overline{y}_0,
\]

and its variance is

\[
\text{var} \left( \hat{T}^{(0)} \right) = N^2 \left( \frac{1 - f}{n} \right) \sigma^2_0.
\]
5.8. Derivations of results in Section 5.4.

This section provides proof of the results in Section 5.4. The derivations are analogues to those presented in Section 5.7. The population total of the response variable that is defined as,

\[ T^{(0)} = l'Z = \begin{pmatrix} u_i' \otimes I_{n}^\prime \\ Y^{(0)} \\ Y^{(0)*} \end{pmatrix}, \]

where \( l = u_i' \otimes I_{n}^\prime \) and \( u_i = \begin{pmatrix} 1 & 0_{1 \times p} \end{pmatrix}' \). After partitioning, we denote the sample and remaining parts as \( Z_I \) and \( Z_{II} \), and \( T^{(0)} \) can be expressed as \( T^{(0)} = l'_IZ_I + l'_IZ_{II} \), where \( l'_I = u_i' \otimes I_{n}^\prime \) and \( l'_II = u_i' \otimes I_{n-n}^\prime \).

The estimator for \( T^{(0)} \) as \( \hat{T}^{(0)} = (l'_I + \mathbf{w}')Z_I \), where \( \mathbf{w} = \begin{pmatrix} w_0' & w_1' & \cdots & w_p' \end{pmatrix}' \), \( w_k \), \( k = 0,1,2,\ldots, p \) are vectors of weights corresponding to the response and the \( p \) auxiliary variables.

The unbiasedness for \( T^{(0)} \) requires \( E(\mathbf{w}'Z_I - l'_IZ_{II}) = 0 \), equivalently,

\( (1'_Iw_0 - (N-n))\mu^{(0)} + \sum_{k=1}^{p} I_n'w_k\mu^{(k)*} = 0 \). Since \( \mu^{(k)*} = 0 \) for any \( k = 1,2,\ldots, p \), this constraint is satisfied for any \( \mu^{(0)} \) simply by \( w_0'1_n = N-n \).

Moreover, the variance of \( \hat{T}^{(0)} \) can be represented as,

\[ \text{var}(\hat{T}^{(0)}) = \mathbf{w}'V_I\mathbf{w} - 2\mathbf{w}'V_{I,II}l_{II} + l_{II}'V_{II}l_{II}. \]

The Lagrangian function is defined as

\[ \Phi(\mathbf{w}) = \mathbf{w}'V_I\mathbf{w} - 2\mathbf{w}'V_{I,II}(u_i \otimes 1_{N-n}) + 2\left\{w_0'1_n - (N-n)\right\}\lambda, \quad (5.28) \]
where $\lambda$ is a Lagrangian multiplier. Differentiating (5.28) with respect to $w$ and $\lambda$ gives,

$$\begin{align*}
\frac{\partial}{\partial w} \Phi(w) &= 2V_{I,I}w - 2V_{I,II}(u_i \otimes 1_{N-n}) + 2(u_i \otimes 1_n)\lambda \\
\frac{\partial}{\partial \lambda} \Phi(w) &= w'1_n - (N-n)
\end{align*}$$

(5.29)

Setting both equations of (5.29) to zeros results in the following estimating equations

$$\begin{pmatrix}
V_{I,II} \\
(u_i' \otimes 1'_{n})
\end{pmatrix}
\begin{pmatrix}
(u_i \otimes 1_n) \\
0
\end{pmatrix}
\lambda
= \begin{pmatrix}
V_{I,II}(u_i \otimes 1_{N-n}) \\
N-n
\end{pmatrix}$$

(5.30)

The solutions of (5.30) are

$$\begin{align*}
\hat{w} &= V_{I,II}^{-1}(u_i \otimes 1_{N-n}) - V_{I,II}^{-1}(u_i \otimes 1_n)\hat{\lambda} \\
\hat{\lambda} &= \left((u_i' \otimes 1'_{n})V_{I,II}^{-1}(u_i \otimes 1_n)\right)^{-1}\left((u_i' \otimes 1'_{n})V_{I,II}^{-1}(u_i \otimes 1_{N-n}) - (N-n)\right)
\end{align*}$$

(5.31)

Since $V_{I}^{-1} = \Sigma^{-1} \otimes P_{n,N}^{-1}$, $V_{I,II} = \Sigma \otimes \left(-\frac{1}{N}J_{n(N-n)}\right)$, $V_{I,II}^{-1}V_{I,II} = -\frac{1}{N-n}I_{p+1} \otimes J_{n(N-n)}$, $P_{n,N}^{-1}1_n = \frac{N}{N-n}1_n$ and $1'_n P_{n,N}^{-1}1_n = \frac{Nn}{N-n}$, (5.31) can be simplified as

$$\hat{w} = -u_i \otimes 1_n - \frac{N}{N-n}\left((\Sigma^{-1}u_i) \otimes 1_n\right)\hat{\lambda} \quad \text{and} \quad \hat{\lambda} = -\frac{N-n}{n}(u_i'\Sigma^{-1}u_i)^{-1}.$$

Therefore,

$$\hat{w} = -u_i \otimes 1_n + \frac{N}{n}(\Sigma^{-1}u_i(u_i'\Sigma^{-1}u_i)^{-1}) \otimes 1_n,$$

(5.32)

Since $\Sigma^{-1} = \begin{pmatrix}
\sigma_0^2 & \sigma_{\sigma_0}' \\
\sigma_{\sigma_0} & \Sigma_*
\end{pmatrix}^{-1} = \begin{pmatrix}
(\sigma_0^2 - \sigma_{\sigma_0}'\Sigma_*^{-1}\sigma_{\sigma_0})^{-1} & -\sigma_0^2\sigma_{\sigma_0}'(\Sigma_* - \sigma_{\sigma_0}\sigma_{\sigma_0}'\sigma_0^{-2})^{-1} \\
-(\Sigma_* - \sigma_{\sigma_0}\sigma_{\sigma_0}'\sigma_0^{-2})^{-1}\sigma_{\sigma_0}\sigma_0^{-2} & (\Sigma_* - \sigma_{\sigma_0}\sigma_{\sigma_0}'\sigma_0^{-2})^{-1}
\end{pmatrix}$,

$$(u_i'\Sigma^{-1}u_i)^{-1} = \sigma_0^2 - \sigma_{\sigma_0}'\Sigma_*^{-1}\sigma_{\sigma_0} \quad \text{and} \quad \Sigma^{-1}u_i = \begin{pmatrix}
(\sigma_0^2 - \sigma_{\sigma_0}'\Sigma_*^{-1}\sigma_{\sigma_0})^{-1} \\
-(\Sigma_* - \sigma_{\sigma_0}\sigma_{\sigma_0}'\sigma_0^{-2})^{-1}\sigma_{\sigma_0}\sigma_0^{-2}
\end{pmatrix},$$

we have
\[
\hat{w} = -u_i \otimes 1_n + \frac{N}{n} \left( \frac{1}{1 - \sigma_{e0}' \Sigma_0^{-1} \sigma_{e0} / \sigma_0^2} \left( \Sigma_0 - \sigma_{e0}' \sigma_{e0} / \sigma_0^2 \right)^{-1} \sigma_{e0} \right) \otimes 1_n .
\] (5.33)

Denote \( \beta_{0*} = \left( 1 - \sigma_{e0}' \Sigma_0^{-1} \sigma_{e0} / \sigma_0^2 \right) \left( \Sigma_0 - \sigma_{e0}' \sigma_{e0} / \sigma_0^2 \right)^{-1} \sigma_{e0} \), it can be shown that

\[
\beta_{0*} = \left( 1 - \sigma_{e0}' \Sigma_0^{-1} \sigma_{e0} / \sigma_0^2 \right) \left( \Sigma_0 - \sigma_{e0}' \sigma_{e0} / \sigma_0^2 \right)^{-1} \sigma_{e0} \\
= (I_p - \Sigma_0^{-1} \sigma_{e0} / \sigma_0^2) \left( \Sigma_0 (I_p - \Sigma_0^{-1} \sigma_{e0} / \sigma_0^2) \right)^{-1} \sigma_{e0} \\
= (I_p - \Sigma_0^{-1} \sigma_{e0} / \sigma_0^2) (I_p - \Sigma_0^{-1} \sigma_{e0} / \sigma_0^2)^{-1} \Sigma_0^{-1} \sigma_{e0} \\
= \Sigma_0^{-1} \sigma_{e0} .
\]

Therefore, (5.33) is simply

\[
\hat{w} = -u_i \otimes 1_n + \frac{N}{n} \left( \frac{1}{-\beta_{0*}} \right) \otimes 1_n .
\] (5.34)

Subsequently, the estimator for \( T^{(0)} \) is

\[
\hat{T}^{(0)} = N\bar{Y}^{(0)} - N\beta_{0*}' \left( \bar{Y}^{(*)} - \mu^{(*)} \right) ,
\] (5.35)

which has variance of

\[
\text{var} \left( \hat{T}^{(0)} \right) = \left( \frac{N}{n} \right)^2 \left\{ \left( 1 - \beta_{0*} \right) \otimes 1_n \right\} \text{cov} \left( Z_i \right) \left\{ \left( \frac{1}{-\beta_{0*}} \right) \otimes 1_n \right\} \\
= \left( 1 - \rho_{0(*)}^2 \right) \left( N^2 \left( \frac{1 - f}{n} \right) \sigma_0^2 \right).
\]

where \( \rho_{0(*)} = \sqrt{\sigma_{e0}' \Sigma_0^{-1} \sigma_{e0} / \sigma_0^2} \) is the multiple correlation coefficient of \( Y^{(0)} \) on \( Y^{(*)} \).
CHAPTER 6

ESTIMATORS BASED ON RANDOM PERMUTATION MODELS
UNDER STRATIFIED SIMPLE RANDOM SAMPLING

6.1. Motivation

In this Chapter, we extend the results based on RPSUR for SRSWOR to settings of stratified simple random sampling without replacement (STSRS). We consider two scenarios where 1) no auxiliary information is available, and 2) one auxiliary variable is known to some extent. For each scenario, we first derive general results when sampling fractions of strata may or may not be equal, and then consider a special case when sampling fractions are proportional to stratum sizes.

6.2. Definition and notations

Strata are $H$ nonoverlapping subpopulations within the population of size $N$, each subpopulation, referred to as a stratum, has a known size of $N_1, N_2, \ldots, N_H$, respectively, and $\sum_{h=1}^{H} N_h = N$.

The stratified simple random sampling without replacement (hereafter, STSRS) refers to the procedure that a sample of fixed sizes is drawn from each stratum independently using SRSWOR scheme, and the $H$ samples across strata constitute a STSRS sample. The sample sizes within strata are denoted by $n_1, n_2, \ldots, n_H$, respectively.
Following conventional notation for stratified sampling (Cochran 1977; Bolfarine and Zacks 1992; Särndal, Swensson and Wretman 1992; Thompson 1997), we adopt the following notations in this Chapter.

Suffix \( h \), \( h = 1,2,\ldots,H \), denotes the stratum. The following notations all refer to stratum \( h \).

\[
\begin{align*}
N_h & \quad \text{total number of subjects} \\
n_h & \quad \text{number of subjects in sample} \\
y_{hs}^{(k)} & \quad \text{value of variable } k \text{ of subject } s \\
a_h & = N_h / N \quad \text{stratum weight} \\
f_h & = n_h / N_h \quad \text{sampling fraction in the stratum} \\
\mu_h^{(k)} & = N_h^{-1} \sum_{s=1}^{N_h} y_{hs}^{(k)} \quad \text{stratum mean of the } k \text{-th variable} \\
\overline{Y}_h^{(k)} & = n_h^{-1} \sum_{i=1}^{n_h} Y_{hi}^{(k)} \quad \text{stratum sample mean of the } k \text{-th variable} \\
\mathbf{\mu}_h & = \left( \mu_h^{(0)}, \mu_h^{(1)}, \ldots, \mu_h^{(p)} \right)' \quad \text{vector of stratum means for all } p + 1 \text{ variables} \\
\Sigma_h & \quad \text{stratum variance-covariance matrix}
\end{align*}
\]

In addition, the vector of stratum means of the \( k \)-th variable is denoted as

\[
\mathbf{\mu}^{(k)} = \left( \mu_1^{(k)}, \mu_2^{(k)}, \ldots, \mu_H^{(k)} \right)' .
\]

For matrix operations, we represent a block diagonal matrix as a \textit{direct sum},

\[
\bigoplus_{h=1}^{H} \mathbf{B}_h , \text{ of matrices } \mathbf{B}_h , \ h = 1,2,\ldots,H . \quad \text{A matrix consists of a column or row of matrices}
\]

91
or vectors is represented by \( \{ c B_h \}_{h=1}^H = (B_1' \ B_2' \ \cdots \ B_H') \) and
\[ \{ c B_h \}_{h=1}^H = (B_1 \ B_2 \ \cdots \ B_H), \]
respectively (Searle, Cassella and McCulloch 1992).

6.3. Independent permutation of one response variable across \( H \) strata

Applying results of Chapter 2, the random variables arising from permuting the response variable \( y \) in the \( h \)-th stratum can be represented as
\[ Y_h = U_h y_h, \]
where \( h = 1, 2, \ldots, H \), \( U_h \) is of dimension \( N_h \times N_h \). The independent permutations of \( H \) strata can be represented as
\[
\begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_H
\end{bmatrix}
= 
\begin{bmatrix}
U_1 & 0 & \cdots & 0 \\
0 & U_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & U_H
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_H
\end{bmatrix}. \tag{6.1}
\]
Denote \( Y = \{ c Y_h \}_{h=1}^H \) and \( y = \{ c y_h \}_{h=1}^H \), (6.1) can be simply represented as
\[ Y = \bigoplus_{h=1}^H U_h y. \tag{6.2} \]
It immediately follows that,
\[ E(Y) = E\left( \bigoplus_{h=1}^H U_h y \right) = \bigoplus_{h=1}^H \mu_h', \]
where \( \mu = (\mu_1 \ \mu_2 \ \cdots \ \mu_H)' \).

Since \( Y \) representing the joint permutations of the \( H \) strata can be rewritten as
\[ Y = \left( \bigoplus_{h=1}^H (y_h' \otimes I_{N_h}) \right) \{ c \ vec(U_h) \}_{h=1}^H, \tag{6.3} \]
the variance-covariance matrix of $Y$ is thus,

$$\text{cov}(Y) = \left( \bigoplus_{h=1}^{H} \left( y_h' \otimes I_{N_h} \right) \right) \text{cov} \left( \{ c \text{ vec}(U_h) \}_{h=1}^{H} \right) \left( \bigoplus_{h=1}^{H} \left( y_h' \otimes I_{N_h} \right) \right).$$

Since the permutations of the $H$ strata are independent, $\text{cov} \left( \text{vec}(U_h), \text{vec}(U_{h'}) \right) = 0$ when $h \neq h'$. Consequently, we have

$$\text{cov} \left( \{ c \text{ vec}(U_h) \}_{h=1}^{H} \right) = \bigoplus_{h=1}^{H} \{ \text{cov}(\text{vec}(U_h)) \}$$

and thus

$$\text{cov}(Y) = \bigoplus_{h=1}^{H} \left( \sigma_h^2 P_{N_h} \right),$$

where $\sigma_h^2 = \frac{1}{N_h - 1} y' P_{N_h} y$ is the population variance of $y$ in $h$-th stratum.

### 6.4. Simultaneous permutations of one response variable and $p$ auxiliary variables across $H$ strata

The $(p+1) \times N_h$ random variables arising from permuting variables $y^{(0)}, y^{(1)}, \ldots, y^{(p)}$ in the $h$-th stratum can be represented as

$$\{ c Y_h^{(k)} \}_{k=0}^{p} = \left( I_{p+1} \otimes U_h \right) \{ c y_h^{(k)} \}_{k=0}^{p}, \quad (6.4)$$

where $h = 1, 2, \ldots, H$, $U_h$ is of dimension $N_h \times N_h$. We denote $Y_h = \{ c Y_h^{(k)} \}_{k=0}^{p}$ and

$$y_h = \{ c y_h^{(k)} \}_{k=0}^{p}$$

and rewrite identity (6.4) as

$$Y_h = \left( I_{p+1} \otimes U_h \right) y_h. \quad (6.5)$$

By applying (6.5) to (6.2), the joint permutations of $p+1$ variables in all $H$ strata can be represented as
\[
\begin{align*}
&\begin{pmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_H
\end{pmatrix} = \left( \bigoplus_{h=1}^{H} \left( I_{p+1} \otimes U_h \right) \right) \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_H
\end{pmatrix}.
\end{align*}
\] (6.6)

Denote \( Y = \{ c^T y_h \}_{h=1}^H \) and \( y = \{ c^T y_h \}_{h=1}^H \), (6.6) can be simplified as

\[
Y = \left( \bigoplus_{h=1}^{H} \left( I_{p+1} \otimes U_h \right) \right) y.
\] (6.7)

It immediately follows that,

\[
E(Y) = E\left( \left( \bigoplus_{h=1}^{H} \left( I_{p+1} \otimes U_h \right) \right) y \right) = \left( \bigoplus_{h=1}^{H} \left( I_{p+1} \otimes 1_{N_h} \right) \right) \mu,
\]

where \( \mu = \{ \mu_h \}_{h=1}^H \), \( \mu_h = \left( \mu_{h}^{(0)} \quad \mu_{h}^{(1)} \quad \cdots \quad \mu_{h}^{(p)} \right)' \), \( h = 1, 2, \ldots, H \).

Similar to (6.3), re-express the permutations of the response and auxiliary variable in the \( h \)-th stratum as

\[
Y_h = \left\{ c^T y_h^{(k)} \right\}_{k=0}^p = \left\{ c^T y_h^{(k)} \right\}_{k=0}^p \otimes I_{(p+1)N_h} vec \left( I_{p+1} \otimes U_h \right).
\]

In Chapter 2, we have shown that

\[
\text{cov}(Y_h) = \Sigma_h \otimes P_{N_h},
\]

where \( \Sigma_h \) is the population variance-covariance matrix of the \( h \)-th stratum. Therefore, the joint variance-covariance matrix for all \( H \) strata is

\[
\text{cov}(Y) = \bigoplus_{h=1}^{H} \text{cov}(Y_h) = \bigoplus_{j=1}^{H} \left( \Sigma_h \otimes P_{N_h} \right).
\]
6.5. Stratified random sampling and partition of random matrices

A sample of size \( n_h, n_h \leq N_h \), is drawn by SRSWOR from the \( h \)-th stratum, where \( h = 1, 2, \ldots, H \), can be viewed as the first \( n_h \) subjects in a random permutation within the \( h \)-th stratum. The elements of random vector \( Y_h \) arising from the permutation can be partitioned into a sample (indexed as \( I \)) and the remaining (indexed as \( II \)) parts by pre-multiplication by a permutation matrix \( K_h \), i.e.,

\[
Z_h = K_h Y_h = \begin{pmatrix} Z_{hl} \\ Z_{hII} \end{pmatrix},
\]

where \( K_h \) is defined similarly as (2.9) of Section 2.5, \( Z_{hl} = \left( Y_{h(0)^p}^{(0)} \ Y_{h(1)^p}^{(1)} \cdots Y_{h(p)^p}^{(p)} \right)' \) and \( Z_{hII} = \left( Y_{hII(0)^p}^{(0)} \ Y_{hII(1)^p}^{(1)} \cdots Y_{hII(p)^p}^{(p)} \right)' \). The expected values of \( Z_h \) are,

\[
E(Z_h) = E \begin{pmatrix} Z_{hl} \\ Z_{hII} \end{pmatrix} = \begin{pmatrix} (p+1) \otimes 1_{n_h} \\ (p+1) \otimes 1_{N_h-n_h} \end{pmatrix} \mu_h.
\]

Further, the partitioned variance-covariance matrix of the \( h \)-th stratum is

\[
V_h = \text{cov} \begin{pmatrix} Z_{hl} \\ Z_{hII} \end{pmatrix} = \begin{pmatrix} V_{hl} & V_{h(I,II)} \\ V_{h(II,I)} & V_{hII} \end{pmatrix},
\]

where \( V_{hl} = \Sigma_h \otimes P_{n_h, N_h} \), \( V_{h(I,II)} = \Sigma_h' \otimes \left( -\frac{1}{N_h} J_{n_h \times (N_h - n_h)} \right) \), \( V_{hII} = \Sigma_h \otimes P_{(N_h - n_h), N_h} \).

The partitioning of all \( H \) strata can be represented jointly as

\[
Z^* = \begin{pmatrix} Z_I \\ Z_{II} \end{pmatrix} = \left( \bigoplus_{h=1}^{H} K_h \right) Y,
\]

(6.8)
where \( Z_i = \left\{ \left\{ c \begin{pmatrix} Y_{hl}^H \\ Y_{hl}^H \\ Y_{hl}^H \end{pmatrix} \right\}_{h=1}^H \right\}, \ Z_{II} = \left\{ \left\{ c \begin{pmatrix} Y_{hll}^H \\ Y_{hll}^H \end{pmatrix} \right\}_{h=1}^H \right\}, \ R_f = \bigoplus_{h=1}^H \left( I_{p+1} \otimes I_{n_h} \ | \ I_{p+1} \otimes 0_{n_h \times (N_h - n_h)} \right) \)

and \( R_{II} = \bigoplus_{h=1}^H \left( I_{p+1} \otimes 0_{(N_h - n_h) \times n_h} \ | \ I_{p+1} \otimes I_{n_h - n_h} \right) \). After algebraic simplification, we have

\[
E(Z') = \left( \begin{array}{c} \bigoplus \mathbf{K}_h \\ \bigoplus \mathbf{K}_h \end{array} \right) \left( \begin{array}{c} \mathbf{R}_I \\ \mathbf{R}_{II} \end{array} \right) \left( \begin{array}{c} \mathbf{V}_I \\ \mathbf{V}_{II} \end{array} \right) \mathbf{Y},
\]

\[
\text{cov}(Z') = \mathbf{R} \left( \bigoplus \mathbf{K}_h \right) \text{cov}(Y) \left( \bigoplus \mathbf{K}_h \right) \mathbf{R}',
\]

where \( V_I = \bigoplus_{h=1}^H V_{hl}, V_{II} = V_{II,l} = \bigoplus_{h=1}^H V_{h(l,II)} \) and \( V_{II} = \bigoplus_{h=1}^H V_{hII} \).

6.6. Parameter of interest

The population mean of the response variable can be defined as a weighted average of stratum means, for example,

\[
\mu^{(0)} = \left(\begin{array}{ccc}
(a_1 & 0_{1 \times p}) & (a_2 & 0_{1 \times p}) & \cdots & (a_H & 0_{1 \times p})
\end{array}\right) \mu
\]

\[
= \left(a' \otimes \begin{pmatrix} 1 & 0_{1 \times p} \end{pmatrix} \right) \mu,
\]

where \( a = (a_1, a_2, \ldots, a_H)' \), \( a_h = N_h/N \) is stratum weight of the \( h \)-th stratum, \( h = 1, 2, \ldots, H \). The vector of population means of the \( p + 1 \) variables can be written as

\[
\mu^{(*)} = \left\{ c \left( a' \otimes u_{k+1}' \right) \right\}_{k=0}^p \mu, \text{ where } u_{k+1}' \text{ is the } (k+1)-\text{th row of identity matrix } I_{p+1}.
\]

6.7. Formulation of the model when auxiliary variables are present

The general approach to analyze stratified sampling data with RPSUR models is to treat the sampling process in each stratum as an independent permutation process.
When one response variable \( y^{(0)} \) and \( p \) auxiliary variables \( y^{(1)}, y^{(2)}, \ldots, y^{(p)} \) are present, a RPSUR model similar to that for the simple random sampling can be defined as,

\[
\{ c \ Y_h \}^H_{h=1} = \left( \bigoplus_{h=1}^H \left( I_{p+1} \otimes 1_{N_h} \right) \right) \{ c \ \mu_h \}^H_{h=1} + E, \tag{6.9}
\]

where \( Y_h = \{ c \ Y^{(k)}_h \}_{k=1}^p \) is defined in (6.4). Model (6.9) can be written alternatively as

\[
Y = X \mu + E, \tag{6.10}
\]

where \( Y = \{ c \ Y_h \}^H_{h=1}, X = \bigoplus_{h=1}^H \left( I_{p+1} \otimes 1_{N_h} \right) \) and \( \mu = \{ c \ \mu_h \}^H_{h=1} \).

Suppose that a set of \( H \) samples of fixed size \( n_h, \ h = 1, 2, \ldots, H \), are drawn with SRSWOR scheme from the \( H \) strata. Using a similar derivation as in Section 4.3, we can partition model (6.10) into sample and remaining proportions, and write the partitioned model as

\[
\begin{pmatrix}
Z_I \\
Z_{II}
\end{pmatrix} = \begin{pmatrix} X_I \\ X_{II} \end{pmatrix} \mu + E, \tag{6.11}
\]

where \( Z_I = \{ c \ Z_{ih} \}^H_{h=1} \) and \( Z_{II} = \{ c \ Z_{ih} \}^H_{h=1} \) as defined in (6.8), \( X_I = \bigoplus_{h=1}^H \left( I_{p+1} \otimes 1_{N_h} \right) \) and \( X_{II} = \bigoplus_{h=1}^H \left( I_{p+1} \otimes 1_{N_h} \right) \), and \( E \) is the vector of residuals. It is evident that

\[
E(Z_I) = X_I \mu \text{ and } E(Z_{II}) = X_{II} \mu.
\]

6.8. **Estimating \( T^{(0)} \) when one auxiliary variable is present**

Without losing generality, we first consider a simple case when there are only three strata \( (H = 3) \) and one auxiliary variable \( (p = 1) \) is known. We will consider two possible scenarios of interest, i.e., 1) the stratum means of the auxiliary variable for all
$H$ strata are known, 2) some but not all of the auxiliary variable stratum means are known.

Based on Model (6.10), the population total $T^{(0)}$ of the response variable can be represented using the following identity,

$$T^{(0)} = I'Y = \left( I_1' \mid I_2' \mid I_3' \right)Y,$$

where $I = \left( I_1' \mid I_2' \mid I_3' \right)'$, $I_h = (1 \ 0)' \otimes 1_{N_h}$, $h = 1,2,3$. After sampling and partitioning $Z$ into $Z_I$ and $Z_{II}$, $T^{(0)}$ can be written as

$$T^{(0)} = I'_I Z_I + I'_{II} Z_{II},$$

where $I_I = \left( I_{1I}' \mid I_{2I}' \mid I_{3I}' \right)'$, $I_h = \left( (1 \ 0) \otimes 1_{n_h} \right)'$, and $I_{II} = \left( I_{1II}' \mid I_{2II}' \mid I_{3II}' \right)'$, $I_{hII} = \left( (1 \ 0) \otimes 1_{(N_h - n_h)} \right)'$.

The primary interest is to derive an estimator for $T^{(0)}$ that is a linear function of the sample, unbiased for $T^{(0)}$ and has minimum mean squared error.

6.8.1. Linear estimator for $T^{(0)}$

We define the linear estimator as,

$$\hat{T}^{(0)} = (I_I' + w')Z_I,$$

where $w = \left\{ \left( w_h^{(0)} \right)_{h=1}^{II} \right\}$, $w_h^{(0)}$ and $w_h^{(i)}$ have dimension of $n_h \times 1$.

6.8.2. Unbiasedness constraints

The unbiasedness for $T^{(0)}$ requires that $E \left( w'Z_I - I_I'Z_{II} \right) = 0$ for any possible $T^{(0)}$. This constraint is equivalent to,
\[
\sum_{h=1}^{3} \left( \left( w^{(0)}_h \mathbf{1}_{n_h} - (N_h - n_h) \right) \mu^{(0)}_h \right) + \sum_{h=1}^{3} w^{(1)}_h \mathbf{1}_{n_h} \mu^{(1)}_h = 0 .
\]

To satisfy this condition for all \( \mu^{(0)}_h \) and \( \mu^{(1)}_h \), it is required that both terms of the left-hand side equal 0, i.e.,

\[
q^{(0)}: \quad \left( w' \left( \oplus_{h=1}^{3} (1 \ 0) \otimes \mathbf{1}_{n_h} \right) - \left\{ c \left( N_h - n_h \right) \right\}_{h=1}^{3} \right) \mu^{(0)} = 0 ,
\]

\[
q^{(1)}: \quad w' \left( \oplus_{h=1}^{3} (0 \ 1) \otimes \mathbf{1}_{n_h} \right) \mu^{(1)} = 0 ,
\]

where \( \mu^{(k)} = \left( \mu^{(k)}_1 \mu^{(k)}_2 \mu^{(k)}_3 \right)' , k = 0,1 \).

To satisfy \( q^{(0)} \) for any \( \mu^{(0)} \), it is required that,

\[
\left( \oplus_{h=1}^{3} (1 \ 0) \otimes \mathbf{1}_{n_h} \right) w - \left\{ c \left( N_h - n_h \right) \right\}_{h=1}^{3} = 0 .
\]

Constraint \( q^{(1)} \) can be satisfied by either forcing \( w^{(1)}_h \) to be null vectors or through proper transformation such that \( \mu^{(1)}_h = 0 \), where \( h = 1,2,3 \). Transformation of random variables have been discussed in Chapter 5 in great detail, we apply those results here.

**Case 1:** If all stratum means of auxiliary variable are known, \( q^{(1)} \) will be satisfied by applying following transformation,

\[
\left\{ \left( \mathbf{I}_{N_h} \right) \mathbf{P}_{N_h} + \frac{1_{N_h}}{N_h} \mathbf{1}_{N_h}' \right\} \times \left\{ \left( Y_{h}^{(0)} \right) \right\}_{h=1}^{3} = \left\{ \left( Y_{h}^{(0)} \right) \right\}_{h=1}^{3} - \left\{ \left( \mathbf{0} \right) \right\}_{h=1}^{3} .
\]

99
Case 2: If some but not all stratum-specific auxiliary means are known while the population auxiliary mean is unknown, such incomplete auxiliary information can also be used to improve estimation. To illustrate this we consider a case where two stratum auxiliary means $\mu_1^{(i)}$ and $\mu_2^{(i)}$ are known. We apply the following transformation to incorporating the two known stratum auxiliary means,

$$
\begin{pmatrix}
\mathbf{Y}_1^{*(i)} \\
\mathbf{Y}_2^{*(i)} \\
\mathbf{Y}_3^{*(i)}
\end{pmatrix} =
\begin{pmatrix}
\mathbf{Y}_1^{(i)} - \mu_1^{(i)} \mathbf{1}_{N_1} \\
\mathbf{Y}_2^{(i)} - \mu_2^{(i)} \mathbf{1}_{N_2} \\
\mathbf{Y}_3^{(i)}
\end{pmatrix}.
$$

(6.17)

Therefore, constraint $\mathbf{q}^{(i)}$ is reduced to

$$
\left( \mathbf{1}_{n_3}^\prime \mathbf{w}_3^{(i)} \right) \mu_3^{(i)} = 0,
$$

to hold this condition for any $\mu_3^{(i)}$, it is required that $\mathbf{1}_{n_3}^\prime \mathbf{w}_3^{(i)} = \mathbf{0}$.

6.8.3. Estimators for $\hat{T}^{(0)}$ after reparameterization

Following the strategy outlined in Chapter 5, we transform the auxiliary variable $\mathbf{Y}^{(i)}$ to $\mathbf{Y}^{*(i)}$ using either (6.16) or (6.17) whenever appropriate. As an analogue to (5.2) of Chapter 5, it can be shown that $\text{cov}\left(\mathbf{Y}^{*(i)}\right) = \text{cov}\left(\mathbf{Y}^{(i)}\right)$.

Based on Model (6.11), the partitioned model can be written as

$$
\begin{pmatrix}
\mathbf{Z}_I^* \\
\mathbf{Z}_{II}^*
\end{pmatrix} = \begin{pmatrix}
\mathbf{X}_I \\
\mathbf{X}_{II}
\end{pmatrix} \mu^* + \mathbf{E},
$$

(6.18)

where $\mathbf{Z}_I = \left\{ \begin{pmatrix}
\mathbf{Y}^{(0)}_{hl} \\
\mathbf{Y}^{*(i)}_{hl}
\end{pmatrix}_{h=1}^c \right\}^3$, $\mathbf{Z}_{II}^* = \left\{ \begin{pmatrix}
\mathbf{Y}_{hl}^{(0)} \\
\mathbf{Y}_{hl}^{*(i)}
\end{pmatrix}_{h=1}^c \right\}^3$, $\mathbf{X}_I = \text{3} \left( \mathbf{I}_{p+1} \otimes \mathbf{1}_{n_k} \right)$, $\mathbf{X}_{II} = \text{3} \left( \mathbf{1}_{p+1} \otimes \mathbf{1}_{n_k} \right)$, and $\mu^* = \begin{pmatrix}
(\mu_1^{(0)})^t \\
(\mu_2^{(0)})^t 
\end{pmatrix}$.

The variance of $\hat{T}^{(0)}$ is thus
\[
\text{var} \left( \hat{T}_0 \right) = \left( w' - r'' \right) \begin{pmatrix}
V_f & V_{fII} \\
V'_{fII} & V_{II}
\end{pmatrix}^{-1} \begin{pmatrix} w \\
-1
\end{pmatrix}.
\]

6.8.4. Case 1: all stratum means of the auxiliary variable are known

When all stratum means of the auxiliary variable are known, we apply transformation (6.16), and define the optimization function as

\[
\Phi(w) = w'V_f w - 2w'V_{fII}I_{III} + 2\lambda' \left\{ X_f'w - \left\{ c \left( N_h - n_h \right) \right\}_{h=1}^3 \right\},
\]

where \( X_f' = \left( \bigoplus_{h=1}^3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) X_f' = \left( \bigoplus_{h=1}^3 \left( \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes 1_{n_h} \right) \right) \). Differentiating \( \Phi(w) \) with respect to \( w \) and \( \lambda \), and setting the derivatives to zero leads to the following solution,

\[
\hat{w} = -I_f + \left\{ \begin{pmatrix} 1 \\
-\beta_h/n_h\end{pmatrix}^{\otimes} \frac{N_h}{n_h} \right\}_{h=1}^3,
\]

(6.19)

where \( \beta_h = \frac{\sigma_{01,h}}{\sigma_{1,h}^2} \) is the linear regression coefficient of \( y^{(0)} \) on \( y^{(1)} = y^{(1)} - \mu^{(1)} 1_N \) in the \( h \)-th stratum of the finite population (Cochran 1977).

The estimator for \( T^{(0)} \) can be written as

\[
\hat{T}^{(0)} = \left\{ \begin{pmatrix} 1 \\
-\beta_h/n_h\end{pmatrix}^{\otimes} \frac{N_h}{n_h} \right\}_{h=1}^3 Z_f = \sum_{h=1}^3 N_h \bar{y}_h^{(0)} - \sum_{h=1}^3 N_h \beta_h \left( \bar{y}_h^{(1)} - \mu^{(1)} \right),
\]

where \( \bar{y}_h^{(0)} \) and \( \bar{y}_h^{(1)} \) are sample means of the response and auxiliary of the \( h \)-th stratum.

The variances of \( \hat{T}^{(0)} \) is

\[
\text{var} \left( \hat{T}^{(0)} \right) = \sum_{h=1}^3 N_h^2 \left( 1 - f_h \right) \left( 1 - \rho_h^2 \right) \sigma_{0,h}^2,
\]
where $\rho_h = \frac{\sigma_{01,h}}{\sigma_{0,h} \sigma_{1,h}}$ is the correlation coefficient between the response variable and auxiliary variable in the $h$-th stratum.

6.8.5. Case 2: some but not all stratum means of the auxiliary variable are known

When some but not all stratum means of the auxiliary variable are known, we apply transformation (6.17) and define the optimization function as

$$\Phi(w) = w'V_I w - 2w'V_{I,II} I_{II} + 2\lambda' \left( X''_I - \frac{\{ c (N_h - n_h) \}_{h=1}^3}{0} \right),$$

where $X''_I = \left( \begin{array}{c} \bigoplus_{h=1}^3 (1 \ 0) \\ (0 \ 1) \end{array} \right) X'_I = \left( \begin{array}{c} \bigoplus_{h=1}^3 ((1 \ 0) \otimes 1'_{n_h}) \\ 0 \ 1_{2(n_i + n_j) + n_i} \ 1'_{n_j} \end{array} \right) \left( \begin{array}{c} \bigoplus_{h=1}^3 (1 \ 0) \\ (0 \ 1) \end{array} \right).$

Differentiating $\Phi(w)$ with respect to $w$ and $\lambda$ and setting the derivatives to zero results in the following estimating equations,

$$\begin{pmatrix} V_I \\ X''_I \ 0 \end{pmatrix} \begin{pmatrix} \hat{w} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} V_{I,II} \\ \{ c (N_h - n_h) \}_{h=1}^3 \end{pmatrix}. \quad (6.20)$$

Solving (6.20) leads to

$$\begin{align*}
\hat{w} &= V_I^{-1} V_{I,II} I_{II} - V_I^{-1} X''_I \hat{\lambda} \\
\hat{\lambda} &= \left( X''_I V_I^{-1} X'_I \right)^{-1} \left( X''_I V_I^{-1} V_{I,II} I_{II} - \left\{ c (N_h - n_h) \right\}_{h=1}^3 \right)
\end{align*}$$

After algebraic simplification, we have
\[
\hat{w} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_{n_h} \right\}_{h=1}^3 + \frac{\left\{ \begin{pmatrix} \frac{N_h}{n_h} \left( 1 - \beta_h \right) \\ -\beta_h \end{pmatrix} \otimes I_{n_h} \right\}_{h=1}^2 \right\}^2
\]
\[
\frac{N_3}{n_3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_{n_3}
\]
(6.21)

where \( \beta_h = \sigma_{01,h} / \sigma^2_{1,h} \), \( h = 1,2 \). Consequently, the estimator for \( T_0 \) and its variance are

\[
\hat{T}_0 = \left( l'_i + \hat{w}'' \right) Z_j = \sum_{h=1}^2 \frac{N_h}{n_h} \left( l'_n Y_{i}^{(0)} - \beta_h Y_{i}^{(1)} \right) + \frac{N_3}{n_3} l'_n Y_{3i}^{(0)}
\]

\[
= \sum_{h=1}^2 \frac{N_h}{n_h} \left( \bar{y}_h^{(0)} - \beta_h \left( \bar{y}_h^{(1)} - \mu_h^{(1)} \right) \right) + \frac{N_3}{n_3} \bar{y}_3^{(0)}
\]

\[
\text{var} \left( \hat{T}_0 \right) = \left( l'_i + \hat{w}'' \right)^2 \text{cov} (Z_j) \left( l_i + w^* \right) = \sum_{h=1}^2 \left( \frac{N_h}{n_h} \left( 1 - \beta_h^2 \right) \right) \sigma^2_{0,h} + \frac{N_3}{n_3} \sigma^2_{0,3}.
\]

### 6.9. Derivations of results in Section 6.8.4.

**Case I:** All stratum means of the auxiliary variable are known

As shown in Section 6.7, under the unbiasedness constraint of \( \hat{T}_0 \) for \( T_0 \) and after transformation (6.15), the optimization function is defined as

\[
\Phi (w) = w'V_i w - 2w'V_{i,II}I_{II} + 2\lambda' \left\{ X_i''w - \left\{ c (N_h - n_h) \right\}_{h=1}^3 \right\} = 0,
\]

where \( w = \begin{pmatrix} w_1^{(0)} & w_1^{(1)} & w_2^{(0)} & w_2^{(1)} & w_3^{(0)} & w_3^{(1)} \end{pmatrix}' \), \( w_h^{(1)} \) has dimension of \( n_h \times 1 \);

\[
X_i'' = \begin{pmatrix} \bigotimes_{h=1}^3 (1 \ 0) \end{pmatrix} X_i' = \begin{pmatrix} \bigotimes_{h=1}^3 (1 \ 0) \otimes I_{n_h} \end{pmatrix} \). Differentiating \( \Phi (w) \) with respect to \( w \) and \( \lambda \) leads to the following estimation equations

\[
\begin{pmatrix} V_i & X_i'' \end{pmatrix} \begin{pmatrix} w \\ \lambda \end{pmatrix} = \begin{pmatrix} V_{i,II}I_{II} \\ \left\{ c (N_h - n_h) \right\}_{h=1}^3 \end{pmatrix}
\]
which has a unique solution,

\[ \hat{\lambda} = \left( X_{h}^{*} V_{I}^{-1} X_{I}^{*} \right)^{-1} \left( X_{h}^{*} V_{I}^{-1} V_{I,II} J_{I} - \left\{ c \left( N_{h} - n_{h} \right) \right\}_{h=1}^{3} \right) \]

\[ \hat{w} = V_{I}^{-1} V_{I,II} J_{II} - V_{I}^{-1} X_{I}^{*} \hat{\lambda}. \]

Since \( V_{I}^{-1} = \bigoplus_{h=1}^{3} V_{h}^{-1} \), \( V_{I,II} = V_{II,II} = \bigoplus_{h=1}^{H} V_{h(I,II)} \) and \( V_{I,II}^{-1} V_{I,II} = \bigoplus_{h=1}^{3} V_{h(I,II)} V_{h(I,II)} \), we have

\[ \hat{\lambda} = \left( \bigoplus_{h=1}^{3} \left( (1, 0) \otimes 1_{n_{h}} \right) \right) \left( \bigoplus_{h=1}^{3} V_{h}^{-1} \right) \left( \bigoplus_{h=1}^{3} \left( (1, 0)' \otimes 1_{n_{h}} \right) \right)^{-1} \]

\[ \times \left( \bigoplus_{h=1}^{3} \left( (1, 0) \otimes 1_{n_{h}} \right) \right) \left( \bigoplus_{h=1}^{3} V_{h}^{-1} V_{h(I,II)} \right) \left( (1, 0)' \otimes 1_{n_{h}} \right) \left\{ c \left( N_{h} - n_{h} \right) \right\}_{h=1}^{3} \]

\[ = \left( \bigoplus_{h=1}^{3} \left( (1, 0) \otimes 1_{n_{h}} \right) \right) V_{h}^{-1} \left( (1, 0)' \otimes 1_{n_{h}} \right)^{-1} \]

\[ \times \left( \bigoplus_{h=1}^{3} \left( (1, 0) \otimes 1_{n_{h}} \right) \right) V_{h}^{-1} \left( (1, 0)' \otimes 1_{n_{h}} \right) \left\{ c \left( N_{h} - n_{h} \right) \right\}_{h=1}^{3} \]

\[ \hat{w} = \left( \bigoplus_{h=1}^{3} V_{h}^{-1} \bigoplus_{h=1}^{3} V_{h(I,II)} \right) J_{II} - \left( \bigoplus_{h=1}^{3} V_{h}^{-1} \right) \left( (1, 0)' \otimes 1_{n_{h}} \right) \hat{\lambda}. \]

Since \( V_{h} = \bigoplus_{h=1}^{3} P_{n_{h}, N_{h}} \), \( V_{h}^{-1} = \bigoplus_{h=1}^{3} P_{n_{h}, N_{h}}^{-1} \), \( V_{h(I,II)} = V_{h(I,II)} = \bigoplus_{h=1}^{3} \left( -\frac{1}{N_{h}} J_{n_{h} \times \left( N_{h} - n_{h} \right)} \right) \),

\[ V_{h}^{-1} V_{h(I,II)} = I_{2} \otimes \left( -\frac{1}{N_{h} - n_{h}} J_{n_{h} \times \left( N_{h} - n_{h} \right)} \right) \), and \( J_{II} = \left\{ \frac{1}{c} \otimes 1_{N_{h} - n_{h}} \right\}_{h=1}^{3} \)

\[ \hat{\lambda} = \left( \bigoplus_{h=1}^{3} \left( (1, 0) \otimes 1_{n_{h}} \right) \left( \bigoplus_{h=1}^{3} P_{n_{h}, N_{h}}^{-1} \right) \left( (1, 0)' \otimes 1_{n_{h}} \right) \right)^{-1} \]

\[ \times \left( \bigoplus_{h=1}^{3} \left( (1, 0) \otimes \left( -\frac{n_{h}}{N_{h} - n_{h}} 1_{N_{h} - n_{h}} \right) \right) \right) J_{II}^{-1} \left\{ c \left( N_{h} - n_{h} \right) \right\}_{h=1}^{3} \]
\[
\hat{\lambda} = \left( \bigoplus_{h=1}^{3} \left( \left( \begin{array}{cc}
1 & 0 \\
0 & 0
\end{array} \right) \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \right) \otimes \left( \left( \begin{array}{c}
\mathbf{P}_{n_{h} \times N_{h}}^{-1} \mathbf{1}_{n_{h}} \end{array} \right) \right) \right)^{-1} \\
\times \left( \bigoplus_{h=1}^{3} \left( \left( \begin{array}{cc}
1 & 0 \\
0 & 0
\end{array} \right) \otimes \left( \frac{n_{h}}{N_{h} - n_{h}} \mathbf{1}_{N_{h} - n_{h}} \right) \right) \right) \left( \left( \begin{array}{c}
1 \\
0
\end{array} \right) \otimes \mathbf{1}_{N_{h} - n_{h}} \right) \right)_{h=1}^{3} - \left\{ c \left( N_{h} - n_{h} \right) \right\}_{h=1}^{3}
\]

\[
= \left( \bigoplus_{h=1}^{3} \left( \frac{N_{h} - n_{h}}{n_{h}} \right) \left( \left( \begin{array}{cc}
1 & 0 \\
0 & 0
\end{array} \right) \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \right) \right)^{-3} \left\{ c \left( -n_{h} \right) \right\}_{h=1}^{3} - \left\{ c \left( N_{h} - n_{h} \right) \right\}_{h=1}^{3}
\]

\[
= \left( \bigoplus_{c=1}^{3} \left( \frac{N_{h} - n_{h}}{n_{h}} \right)^{3} \right) \left( \left( \begin{array}{cc}
1 & 0 \\
0 & 0
\end{array} \right) \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \right)_{h=1}^{3}
\]

\[
\hat{w} = \left( \bigoplus_{h=1}^{3} \left( \mathbf{1}_{2} \otimes \left( \frac{-1}{N_{h} - n_{h}} \mathbf{J}_{n_{h} \times (N_{h} - n_{h})} \right) \right) \right) \left( \left( \begin{array}{c}
\mathbf{1}_{N_{h} - n_{h}} \end{array} \right) \right) - \left( \bigoplus_{h=1}^{3} \left( \frac{N_{h}}{N_{h} - n_{h}} \right) \left( \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \right) \otimes \mathbf{1}_{n_{h}} \right) \hat{\lambda}
\]

\[
= \left\{ \left( \begin{array}{c}
1 \\
0
\end{array} \right) \otimes \left( \frac{-1}{N_{h} - n_{h}} \mathbf{J}_{n_{h} \times (N_{h} - n_{h})} \right) \right\}_{h=1}^{3}
\]

\[
- \left( \bigoplus_{h=1}^{3} \left( \frac{N_{h}}{N_{h} - n_{h}} \right) \left( \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \otimes \mathbf{1}_{n_{h}} \right) \right) \left( \left( \begin{array}{c}
1 \\
0
\end{array} \right) \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \right) \right)_{h=1}^{3}
\]

\[
= \left\{ \left( \begin{array}{c}
1 \\
0
\end{array} \right) \otimes \left( -\mathbf{1}_{n_{h}} \right) \right\}_{h=1}^{3} - \left\{ \left( \frac{N_{h}}{n_{h}} \right) \left( \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \otimes \mathbf{1}_{n_{h}} \right) \right( \left( \begin{array}{c}
1 \\
0
\end{array} \right) \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \right) \right\}_{h=1}^{3}
\]

\[
= -l_{r} + \left\{ \left( \frac{N_{h}}{n_{h}} \right) \left( \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \otimes \mathbf{1}_{n_{h}} \right) \right( \left( \begin{array}{c}
1 \\
0
\end{array} \right) \Sigma_{h}^{-1} \left( \begin{array}{c}
1 \\
0
\end{array} \right) \right) \right\}_{h=1}^{3}
\]

105
\[\hat{w} = -l_r + \left\{ \left( -\left( \frac{N_h}{n_h} \right) \frac{1}{\Sigma_{h}^{-1}} \right) \otimes \mathbf{1}_{n_h} \right\} \left( 1 \right) \frac{1}{\Sigma_{h}^{-1}} \right\} \left( 1 \right) ^{-1} \right\} \right\}_{h=1}^3 \]

\[= -l_r + \left\{ \left( \frac{1}{
_{n_h}} \right) \right\}_{h=1}^3 \]

\[= -l_r + \left\{ \left( \frac{1}{\beta_h} \right) \right\}_{h=1}^3 , \]

where \( \beta_h = \frac{\sigma_{01,h}}{\sigma_{1,h}^2} \).

Therefore, the estimator for \( T^{(0)} \) can be written as

\[\hat{T}^{(0)} = \left\{ \left( 1 \mid -\beta_h \right) \otimes \left( \frac{N_h}{n_h} \mathbf{1}_{n_h'} \right) \right\}_{h=1}^3 \mathbf{Z}_t \]

\[= \sum_{h=1}^3 \frac{N_h}{n_h} \mathbf{Y}^{(0)}_{hl} - \sum_{h=1}^3 \frac{\beta_h}{n_h} \mathbf{Y}^{(1)}_{hl} \]

\[= \sum_{h=1}^3 \frac{N_h}{n_h} \bar{T}^{(0)}_h - \sum_{h=1}^3 \frac{\beta_h}{n_h} \left( \bar{T}^{(1)}_h - \mu \right) . \]

The variance of \( \hat{T}^{(0)} \) is

\[\text{var} \left( \hat{T}^{(0)} \right) = \left\{ \left( 1 \mid -\beta_h \right) \otimes \left( \frac{N_h}{n_h} \mathbf{1}_{n_h'} \right) \right\}_{h=1}^3 \text{cov} \left( \mathbf{Z}_t \right) \left\{ \left( 1 \mid -\beta_h \right) ^\prime \otimes \left( \frac{N_h}{n_h} \mathbf{1}_{n_h} \right) \right\}_{h=1}^3 \]

\[= \left\{ \left( 1 \mid -\beta_h \right) \otimes \left( \frac{N_h}{n_h} \mathbf{1}_{n_h'} \right) \right\}_{h=1}^3 \left( \bigoplus_{h=1}^3 \mathbf{V}_h \right) \left\{ \left( 1 \mid -\beta_h \right) ^\prime \otimes \left( \frac{N_h}{n_h} \mathbf{1}_{n_h} \right) \right\}_{h=1}^3 \]

\[= \sum_{h=1}^3 \left( \left( 1 \mid -\beta_h \right) \otimes \left( \frac{N_h}{n_h} \mathbf{1}_{n_h} \right) \right) \left( \frac{1}{\Sigma_{h} \otimes \mathbf{P}_{n_h,n_h}} \right) \left( -\beta_h \right) \otimes \left( \frac{N_h}{n_h} \mathbf{1}_{n_h} \right) \right\} \]
\[
\text{var}(\hat{T}^{(0)}) = \sum_{h=1}^{H} \left\{ \frac{N_h (N_h - n_h)}{n_h} \left( (1 | -\beta_h) \Sigma_h \left( \frac{1}{-\beta_h} \right) \right) \right\}
\]

\[
= \sum_{h=1}^{H} \left\{ \frac{N_h (N_h - n_h)}{n_h} \left( (1 | -\beta_h) \Sigma_h \left( \frac{1}{-\beta_h} \right) \right) \right\}
\]

\[
= \sum_{h=1}^{H} \left\{ \frac{N_h (N_h - n_h)}{n_h} \left( \sigma^2_{0,h} - 2\beta_h \sigma_{01,h} + \beta_h^2 \sigma^2_{1,h} \right) \right\}
\]

\[
= \sum_{h=1}^{H} \left\{ \frac{N_h (N_h - n_h)}{n_h} \left( \sigma^2_{0,h} \right) \right\}
\]

\[
= \sum_{h=1}^{H} \left\{ N_h^2 (1 - f_h) \left( 1 - \rho_h^2 \right) \sigma^2_{0,h} \right\},
\]

where \( \rho_h = \frac{\sigma_{01,h}}{\sigma_{0,h} \sigma_{1,h}} \) is the correlation coefficient between the response variable and auxiliary variable in the \( h \)-th stratum.

**6.10. Derivations of results in Section 6.8.5.**

When some but not all stratum means for the auxiliary variable are known, after application of transformation (6.16) to \( Y_h^{(1)} \) and partitioning into sample and remaining parts \( Z_I \) and \( Z_{II} \). The optimization function is defined as

\[
\Phi(\mathbf{w}) = \mathbf{w}' \mathbf{V}_I \mathbf{w} - 2\mathbf{w}' \mathbf{V}_{I,II} \mathbf{l}_{II} + 2\lambda' \left\{ \mathbf{X}_{c}' \mathbf{w} - \left\{ \left( N_h - n_h \right) \right\}_{h=1}^{H} \right\},
\]

where \( \mathbf{X}_{c}' = \left( \bigoplus_{h=1}^{H} \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \right) \mathbf{X}' = \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \mathbf{X}_c \). Differentiating \( \Phi(\mathbf{w}) \) with respect to \( \mathbf{w} \) and \( \lambda \) and setting the derivatives to zero results in the following estimating equations,
\[
\begin{align*}
\begin{pmatrix} V_I & X'_I \end{pmatrix} \begin{pmatrix} \dot{X}_I \\ \dot{\lambda} \end{pmatrix} &= \begin{pmatrix} V'_{I,II}I_{II} \\ \left\{ \frac{n_h}{c} - n_h \right\}_{h=1}^3 \end{pmatrix} \\
&= \frac{2}{h=1} \left( \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_{n_h} \right) \begin{pmatrix} 0 \\ 2(n_h + n_c) \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ I_2 \otimes I_{n_c} \end{pmatrix}
\end{align*}
\] (6.22)

After simplification, \( X'_I = \frac{2}{h=1} \left( \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_{n_h} \right) \begin{pmatrix} 0 \\ 2(n_h + n_c) \end{pmatrix} \) ...

Solving (6.22) leads to ...

\[
\begin{align*}
\dot{\hat{\lambda}} &= \left( X''_I V'^{-1} I_{II} \right) \begin{pmatrix} X''_I V'^{-1} I_{II} - \left\{ \frac{n_h}{c} - n_h \right\}_{h=1}^3 \end{pmatrix} \\
\hat{\lambda} &= \left( X''_I V'^{-1} I_{II} \right)^{-1} \begin{pmatrix} X''_I V'^{-1} I_{II} - \left\{ \frac{n_h}{c} - n_h \right\}_{h=1}^3 \end{pmatrix} \\
\end{align*}
\]

Since \( V_I = \bigoplus_{h=1}^H V_{hl} \), \( V'^{-1} = \Sigma^{-1} \otimes P^{-1} \), \( V_{II} = V'_{II} = \bigoplus_{h=1}^H V_{h(II)} \), \( V_{hl} = \bigoplus_{h=1}^H V_{h(II)} \), we have ...

\[
\begin{align*}
V_{h(II)} &= V'_{h(II)} = \Sigma_h \otimes \left( -\frac{1}{N_h} J_{n_h \times (N_h - n_h)} \right) \\
I_{II} &= \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_{n_h} \right\}_{h=1}^3 \\
\end{align*}
\]

and consequently ...

\[
\begin{align*}
X''_I V'^{-1} I_{II} &= \bigoplus_{h=1}^2 \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_{n_h} \right\} \begin{pmatrix} 0 \\ I_2 \otimes I_{n_h} \end{pmatrix} = \begin{pmatrix} \left\{ \frac{n_h}{c} \right\}_{h=1}^2 \\ 0 \end{pmatrix} \\
X''_I V'^{-1} I_{II} &= \begin{pmatrix} \left\{ \frac{n_h}{c} \right\}_{h=1}^3 \\ 0 \end{pmatrix} = \begin{pmatrix} \left\{ \frac{n_h}{c} \right\}_{h=1}^3 \\ 0 \end{pmatrix} \\
\end{align*}
\]

108
\[
\left( X_i^t V_i^{-1} X_i^* \right)^{-1} = \begin{pmatrix}
\frac{2}{\sigma_{0,h}} \left( \frac{N_h - n_h}{N_h n_h} \sigma_{0,h}^2 \left( 1 - \rho_{01,h}^2 \right) \right) & 0 \\
0 & \frac{N_3 - n_3 \Sigma_3}{n_3}
\end{pmatrix},
\]

\[
\hat{\lambda} = -\frac{\frac{N - n_h}{n_h} \sigma_{0,h}^2 \left( 1 - \rho_{01,h}^2 \right)}{\frac{N_3 - n_3}{n_3} \sigma_{0,3}^2 \left( \frac{1}{\sigma_{0,3}} \right)}
\]

\[
\hat{w} = \left\{ -\left( \begin{array}{c}
1 \\
0
\end{array} \right) \otimes 1_{n_h} \right\}^{3} - \left( \bigotimes_{h=1}^{3} \left( \Sigma_h^{-1} \otimes \mathbf{P}_{n_h, N_h}^{-1} \right) \right) \left\{ \bigotimes_{h=1}^{2} \left( \begin{array}{c}
0 \\
1_{n_h}
\end{array} \right) \otimes \mathbf{0}_{2(n_h+1)\times 2} \right\}
\]

\[
\times \left\{ -\frac{N - n_h}{n_h} \sigma_{0,h}^2 \left( 1 - \rho_{01,h}^2 \right) \right\}_{h=1}^{2}
\]

\[
\times \left\{ -\frac{N_3 - n_3}{n_3} \sigma_{0,3}^2 \left( \frac{1}{\sigma_{0,3}} \right) \right\}
\]

\[
= \left\{ -\left( \begin{array}{c}
1 \\
0
\end{array} \right) \otimes 1_{n_h} \right\}^{3} + \left\{ \left( \bigotimes_{h=1}^{2} \left( \Sigma_h^{-1} \otimes \mathbf{P}_{n_h, N_h}^{-1} \mathbf{1}_{n_h} \right) \right) \left( \frac{N - n_h}{n_h} \sigma_{0,h}^2 \left( 1 - \rho_{01,h}^2 \right) \right) \right\}_{h=1}^{2}
\]

\[
\times \left\{ -\frac{N_3 - n_3}{n_3} \left( \Sigma_3^{-1} \otimes \left( \mathbf{P}_{n_3, N_3}^{-1} \mathbf{1}_{n_3} \right) \right) \left( \frac{\sigma_{0,3}^2}{\sigma_{0,3}} \right) \right\}
\]

This can be further simplified as
\[
\hat{w} = \left\{ \left( \begin{array}{c}
\beta \\
0
\end{array} \right) \otimes 1_{n_h} \right\}_{h=1}^3 + \frac{\left\{ \left( \frac{N}{n_h} \left( 1 - \beta_{01,h} \right) \otimes 1_{n_h} \right)^2 \right\}_{h=1}}{N \left( \frac{1}{0} \otimes 1_{n_3} \right)}
\]

where \( \beta_{01,h} = \sigma_{01,h} / \sigma_{1,h}^2 \), \( h = 1, 2 \).
CHAPTER 7

APPLICATIONS TO RATE ESTIMATION

7.1. Motivations

This Chapter discusses applications of the methods developed in Chapters 4, 5 and 6 to rate estimation that is common in epidemiologic surveys. We will consider scenarios involving up to two auxiliary variables, either categorical or continuous.

7.2. Cases with single categorical auxiliary variable

We first consider the scenarios with single categorical auxiliary variable that has only two categories (binomial), then extend the results to the case with more than two categories.

7.2.1. Binomial auxiliary variable

Suppose that a sample survey is conducted to estimate the disease prevalence of a population. A SRSWOR sample of size \(n\) is drawn from the population, and the auxiliary variable is subjects’ gender. In practice, the following scenarios may occur,

1) Gender is known for all subjects, and

2) Only total numbers of males \((N_m)\) and females \((N_f)\) are known.

From Section 5.3 of Chapter 5, we know the estimator for population disease prevalence can be written as

\[
\hat{r}_0 = \bar{r}_0 + \beta_{01}(r_1 - \bar{r}_1),
\]  

(7.1)
where $\bar{r}_0$ is disease rate of the sample, $\bar{r}_1$ is the proportion of males in the sample, $r_1$ is the proportion of males in the population, and $\beta_{01} = \sigma_{01}/\sigma_1^2$ is the population regression coefficient of disease status on gender.

When either the gender of all subjects or the total numbers of males (females) of the population is known, $\sigma_1^2$ is known, that is,

$$\sigma_1^2 = \frac{N_mN_f}{N(N-1)}.$$  

For this scenario, knowledge of the total number of males or females is adequate.

By methods of moments, we replace other variance components, $\sigma_0^2$ and $\sigma_{01}$, with their sample estimates,

$$\hat{\sigma}_0^2 = \frac{n_0(n-n_0)}{n(n-1)} \quad \text{and} \quad \hat{\sigma}_{01} = \frac{n_mn_f}{n(n-1)} \left( \frac{n_{0m} - n_{0f}}{n_m/n_f} \right),$$

where $n_m$ and $n_f$ are numbers of males and females in the sample, $n_0 = n_{0m} + n_{0f}$, $n_{0m}$ and $n_{0f}$ are the numbers of male and female disease bearers in the sample, respectively. Consequently, using $\sigma_1^2$,

$$\hat{\beta}_{01} = \frac{N(N-1)}{n(n-1)} \left( \frac{n_mn_f}{N_mN_f} \right) \left( \frac{n_{0m} - n_{0f}}{n_m/n_f} \right).$$

The estimator for disease prevalence is simplified to

$$\hat{r}_0 = \frac{n_0 + N - 1}{n(n-1)} \left( \frac{n_mn_f}{n^2} \right) \left( \frac{n_f}{N_f} - \frac{n_m}{N_m} \right) \left( \frac{n_{0m} - n_{0f}}{n_m/n_f} \right).$$  \hspace{1cm} (7.2)

Its variance is
\[ V(\hat{p}_0) = \frac{1-f}{n} \left( \frac{n_0(n-n_f)}{n(n-1)} \right) \left( 1 - \hat{p}_0^2 \right) \]
\[ = \frac{1-f}{n} \left( \frac{n_0(n-n_f)}{n(n-1)} \right) + \frac{1-f}{n^2} \left( \frac{N(N-1)}{N_mN_f} \right) \left( \frac{\hat{p}_0^2}{n_m} \right)^2, \quad (7.3) \]

where \( \hat{p}_0^2 = \left( \frac{n_0(n-n_f)}{n(n-1)} \right) \left( \frac{n_m n_f}{N_m N_f} \right) \left( \frac{\hat{p}_0^2}{n_m} \right)^2. \)

If we use \( \hat{\sigma}_0^2 = n_0(n-n_f)/n(n-1), \) \( \hat{\sigma}_i^2 = n_m n_f / n(n-1) \) and

\[ \hat{\sigma}_{01}^2 = \frac{n_m n_f}{n(n-1)} \left( \frac{n_{0m} - n_{0f}}{n_m} \right), \text{ then using } \hat{\sigma}_{01}^2, \hat{\beta}_{01} = \frac{n_{0m}}{n_m} - \frac{n_{0f}}{n_f}, \text{ and } \]
\[ \hat{p}_0 = \frac{N_m}{N} \left( \frac{n_{0m}}{n_m} \right) + \frac{N_f}{N} \left( \frac{n_{0f}}{n_f} \right). \quad (7.4) \]

This is a “post-stratified” estimator of prevalence rate, which has been shown using calibration (Deville and Särndal 1992) and model-assisted approaches (Särndal, Swensson and Wretman 1992). Its variance is

\[ V_p(\hat{p}_0) = \frac{1-f}{n(n-1)} \left( n_0 - \left( \frac{n_{0m}^2}{n_m} + \frac{n_{0f}^2}{n_f} \right) \right). \quad (7.5) \]

Derivations of (7.2), (7.3), (7.4) and (7.5) are given in Section 7.5.

7.2.2. Relative performance of various estimation methods

We conducted a small Monte Carlo simulation to evaluate the performance of estimators (7.2) and (7.5). We created a series of hypothetical populations of sizes 50, 100, 200, 400, 800 and 1,600; each consists of 50% men and 50% women. In all populations, the prevalence of smokers is 60% for men and 20% for women. We use
the known gender proportions to improve our estimation based on estimators shown in the previous section.

The simulation study was carried out as follows. From each population of size $N = 25, 50, 100, 200, 400, 800$, a series of 5,000 SRSWOR samples of size $n$, where $n = 25, 50, 100, 200, 400, 800$ and $n < N$, were drawn. For each sample, we estimated the proportion of smokers, and its variance using the following methods:

1) **SEP**: Simple Expansion Predictor without auxiliary information.

2) **DPS**: Design-based Predictor using Sample estimate of population variance of gender ($\hat{\sigma}_1^2$) and estimated covariance ($\hat{\sigma}_{01}$), i.e., (7.4) and (7.5).

3) **DPP**: Design-based Predictor using known Population variance of gender ($\sigma_1^2$), i.e., (7.2) and (7.3), respectively.

4) **POP**: Design-based predictor using known Population variance and covariance, i.e., predictor (5.12).

5) **JKF**: The variance of design-based predictor (7.4) was estimated with modified delete-one Jackknife method (Jones 1974) using functional form (7.4).

The 95% confidence intervals (CIs) were computed using the variance estimators based on asymptotic normal approximation. We then compared the overall performance of these estimation methods in terms of mean squared error (MSE), coverage rates and average length of the confidence intervals. Simulation results are summarized in the following Tables 7.1, 7.2 and 7.3, and presented graphically in Figures 7.1, 7.2 and 7.3.

Based on these simulation data, we made the following observations.
1) As expected, these simulations confirm that all these estimators are asymptotically unbiased, or at least the biases are negligible; especially when sample size is not small.

2) We observed significant reduction in MSE of the estimators relative to simple expansion estimators for small and large sample sizes (Table 7.1, Figure 7.2), and thus narrower confidence intervals (Table 7.2).

3) From Figures 7.2 and 7.3, we concluded that DPP method is superior to DPS method in terms of MSE reduction, especially when sample size is relatively small (n<200).

4) The coverage rate of the confidence intervals constructed using the variance estimators of the SRS, DPS, DPP, POP and JKF methods are all very close to their nominal level (95%). When sample sizes are small (n ≤ 50), DPP and DPS confidence intervals tend to have highest coverage rates and the narrowest width. When sample sizes are moderate or large, there are little differences in coverage rates between the methods. However, the CIs based on Jackknife variance estimator tend to be much wider than POP, DPP and DPS methods.

5) From Table 7.3, the CIs based on Jackknife variance and DPS method appeared to have higher rate of right-miss for smaller sample sizes (n<=100).
Table 7.1. Mean squared error (x 10,000) of four estimators by population size and sample size

<table>
<thead>
<tr>
<th>Population size</th>
<th>SEP</th>
<th>DPS</th>
<th>DPP</th>
<th>POP</th>
<th>SEP</th>
<th>DPS</th>
<th>DPP</th>
<th>POP</th>
<th>SEP</th>
<th>DPS</th>
<th>DPP</th>
<th>POP</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>48.3</td>
<td>42.6</td>
<td>39.7</td>
<td>39.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>71.6</td>
<td>64.3</td>
<td>60.6</td>
<td>60.6</td>
<td>24.4</td>
<td>21.2</td>
<td>20.5</td>
<td>20.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>85.2</td>
<td>75.2</td>
<td>70.8</td>
<td>70.8</td>
<td>35.5</td>
<td>30.4</td>
<td>29.5</td>
<td>29.5</td>
<td>12.0</td>
<td>10.1</td>
<td>9.9</td>
<td>9.9</td>
</tr>
<tr>
<td>400</td>
<td>89.1</td>
<td>76.8</td>
<td>73.0</td>
<td>73.0</td>
<td>40.2</td>
<td>34.9</td>
<td>34.0</td>
<td>34.0</td>
<td>17.8</td>
<td>15.1</td>
<td>15.0</td>
<td>15.0</td>
</tr>
<tr>
<td>800</td>
<td>96.3</td>
<td>85.0</td>
<td>81.0</td>
<td>81.0</td>
<td>44.9</td>
<td>38.7</td>
<td>37.7</td>
<td>37.7</td>
<td>21.5</td>
<td>18.0</td>
<td>17.8</td>
<td>17.8</td>
</tr>
<tr>
<td>1600</td>
<td>95.6</td>
<td>84.1</td>
<td>80.8</td>
<td>80.8</td>
<td>47.1</td>
<td>40.8</td>
<td>39.9</td>
<td>39.9</td>
<td>21.8</td>
<td>18.7</td>
<td>18.5</td>
<td>18.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Population size</th>
<th>SEP</th>
<th>DPS</th>
<th>DPP</th>
<th>POP</th>
<th>SEP</th>
<th>DPS</th>
<th>DPP</th>
<th>POP</th>
<th>SEP</th>
<th>DPS</th>
<th>DPP</th>
<th>POP</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>5.9</td>
<td>4.9</td>
<td>4.9</td>
<td>4.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>9.2</td>
<td>7.6</td>
<td>7.6</td>
<td>7.6</td>
<td>3.0</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>10.8</td>
<td>9.1</td>
<td>9.0</td>
<td>9.0</td>
<td>4.6</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
<td>1.5</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
</tr>
</tbody>
</table>

SEP = Simple expansion predictor based SRSWOR, DPS = Estimates based on sample variance and covariance, DPP = Estimates based on sample covariances and population variance of the auxiliary variable, POP = Estimates based on population variance and covariance.
Table 7.2. Coverage rates and width of 95% confidence intervals

<table>
<thead>
<tr>
<th>Population size</th>
<th>Sample size and Estimation method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEP</td>
</tr>
<tr>
<td>50</td>
<td>96.0</td>
</tr>
<tr>
<td>100</td>
<td>93.4</td>
</tr>
<tr>
<td>200</td>
<td>91.0</td>
</tr>
<tr>
<td>400</td>
<td>94.7</td>
</tr>
<tr>
<td>800</td>
<td>93.6</td>
</tr>
<tr>
<td>1600</td>
<td>93.6</td>
</tr>
<tr>
<td>400</td>
<td>95.0</td>
</tr>
<tr>
<td>800</td>
<td>94.2</td>
</tr>
<tr>
<td>1600</td>
<td>94.5</td>
</tr>
</tbody>
</table>

SEP = Simple expansion predictor, DPS = DP based on sample variance and covariances, DPP = DP using sample covariances and population variance of the auxiliary variable, POP = Design-based predictor (DP) based on population variance and covariances, JKF = CI based on Jackknife variance estimator.
Table 7.3. Percentage of left- and right-misses of the 95% confidence intervals, by population size, sample size and estimation methods

<table>
<thead>
<tr>
<th>Population size</th>
<th>SEPL</th>
<th>DPSL</th>
<th>DPP</th>
<th>POPL</th>
<th>JKF</th>
<th>SEPR</th>
<th>DPSR</th>
<th>DPPR</th>
<th>POPR</th>
<th>JKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>2.0</td>
<td>2.3</td>
<td>2.4</td>
<td>2.4</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>3.5</td>
<td>2.5</td>
<td>2.5</td>
<td>3.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.7</td>
<td>2.2</td>
<td>3.3</td>
<td>3.3</td>
<td>2.7</td>
<td>3.2</td>
<td>2.4</td>
<td>2.7</td>
<td>2.7</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>4.9</td>
<td>4.4</td>
<td>2.7</td>
<td>2.7</td>
<td>3.9</td>
<td>3.5</td>
<td>3.5</td>
<td>2.6</td>
<td>2.6</td>
<td>2.9</td>
</tr>
<tr>
<td>200</td>
<td>2.6</td>
<td>2.6</td>
<td>3.3</td>
<td>3.3</td>
<td>2.7</td>
<td>1.3</td>
<td>2.2</td>
<td>2.4</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>6.5</td>
<td>4.8</td>
<td>2.7</td>
<td>2.7</td>
<td>3.6</td>
<td>3.3</td>
<td>3.4</td>
<td>2.5</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>400</td>
<td>2.7</td>
<td>2.4</td>
<td>3.0</td>
<td>3.0</td>
<td>2.5</td>
<td>2.0</td>
<td>2.5</td>
<td>2.7</td>
<td>2.7</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>4.1</td>
<td>2.2</td>
<td>2.2</td>
<td>3.5</td>
<td>3.9</td>
<td>3.0</td>
<td>2.3</td>
<td>2.3</td>
<td>2.8</td>
</tr>
<tr>
<td>800</td>
<td>3.3</td>
<td>2.6</td>
<td>3.7</td>
<td>3.7</td>
<td>2.7</td>
<td>2.8</td>
<td>2.6</td>
<td>3.2</td>
<td>3.2</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>3.1</td>
<td>4.5</td>
<td>2.6</td>
<td>2.6</td>
<td>4.4</td>
<td>2.2</td>
<td>3.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.9</td>
</tr>
<tr>
<td>1600</td>
<td>3.5</td>
<td>2.6</td>
<td>3.4</td>
<td>3.4</td>
<td>2.7</td>
<td>3.2</td>
<td>2.9</td>
<td>3.6</td>
<td>3.6</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>4.4</td>
<td>2.5</td>
<td>2.5</td>
<td>4.0</td>
<td>2.5</td>
<td>3.4</td>
<td>2.3</td>
<td>2.3</td>
<td>3.0</td>
</tr>
</tbody>
</table>
SEP = Simple expansion predictor, DPS = Design-based predictor (DP) based on sample variance and covariances, DPP = DP using sample covariances and population variance of the auxiliary variable, POP = DP based on population variance and covariances, JKF = Ci based on Jackknife variance estimator.
Figure 7.1. Mean squared error by population size and sample size

Sample sizes (from top to bottom): n=25, 50, 100, 200, 400
Symbols: . = SRS, x = DPS, O = DPP, + = POP
Figure 7.2. MSE reduction (%) of a design-based predictor (DPS or DPP )
by population and sample sizes

Estimation method: DPS
Sample size: =25, o=50, x=100, + =200, O=400, D=800

MSE reduction %

Population size

50 100 200 400 800 1600

MSE reduction %

Population size

50 100 200 400 800 1600

Estimation method: DPP
Sample size: =25, o=50, x=100, + =200, O=400, D=800
Figure 7.3. MSE Ratio of design-based predictors to SRS estimator by population and sample sizes

Estimation method: DPS
Sample size: n=25, n=50, n=100, n=200, n=400, n=800

Estimation method: DPP
Sample size: n=25, n=50, n=100, n=200, n=400, n=800
7.2.3. Extension to multinomial auxiliary variable

Let us consider the example of estimating the rate of drug coverage. When a categorical variable has more than two categories, we may extend the above results by replacing the categorical variable with a set of indicator variables. For example, we are interested in estimating the disease prevalence by age categories \((y_s^{(0)} = 1\) is subject \(s\) has drug coverage, \(0\) otherwise) based on a fixed size sample survey. The subjects are classified into three age categories (referred as subpopulations hereafter), e.g., \(\leq 65\), \(>65\) but \(\leq 75\) and \(>75\). The marginal totals of the three age categories \((N_1, N_2, N_3)\) are known. The age categories may be represented using three indicator variables,

\[
y^{(1)} = \begin{cases} 
1 & \text{if } \leq 65, \\
0 & \text{otherwise.}
\end{cases}
\]

\[
y^{(2)} = \begin{cases} 
1 & \text{if } 65 < \text{but } \leq 75, \\
0 & \text{otherwise.}
\end{cases}
\]

\[
y^{(3)} = \begin{cases} 
1 & \text{if } >75, \\
0 & \text{otherwise.}
\end{cases}
\]

To avoid singularity, we use \(y^{(1)}\) and \(y^{(2)}\) in the model. By applying results derived in Section 5.4 of Chapter 5, we have

\[
\hat{r}_0 = \tau_0 - \mathbf{\beta}_0' \left( \overline{Y}^{(*)} - \mathbf{\mu}^{(*)} \right),
\]

where \(\mathbf{\beta}_0 = \Sigma_0^{-1} \mathbf{\sigma}_0\), \(\mathbf{\mu} = \left( \frac{N_1}{N} \quad \frac{N_2}{N} \right)'\) and \(\overline{Y}^{(*)} = \left( \overline{Y}^{(1)} \quad \overline{Y}^{(2)} \right)'\).

Since \(N_1\) and \(N_2\) are known, we have

\[
\Sigma_* = \frac{N_1 N_2}{N(N-1)} \begin{pmatrix} 
N_2^{-1} (N - N_1) & 1 \\
-1 & N_1^{-1} (N - N_2)
\end{pmatrix},
\]

and
\[
\Sigma_\star^{-1} = \frac{N - 1}{N_1 N_2 N_3} \begin{pmatrix}
N_2 (N - N_2) & N_1 N_2 \\
N_1 N_2 & N_1 (N - N_1)
\end{pmatrix}.
\]

It is shown that \( \hat{\sigma}^2_0 = \frac{n_0 (n - n_0)}{n(n-1)} \) and \( \hat{\sigma}_{\star \phi} = \frac{1}{n(n-1)} \left( \frac{n n_{01} - n_0 n_1}{n n_{02} - n_0 n_2} \right) \). Therefore,

\[
\hat{\beta}_{\star \phi} = \frac{N - 1}{n(n-1)} \left( \frac{1}{N_1} \left( \frac{n n_{01} - n_0 n_1}{n n_{02} - n_0 n_2} \right) \right) + \frac{N - 1}{n(n-1)} \frac{1}{N_3} \left( \frac{n_n n_0 - n n_{02}}{n n_{02} - n_0 n_2} \right) \left( \frac{1}{1} \right)
\]

After algebraic simplification, we have

\[
\hat{r}_0 = \frac{n_0}{n} - \frac{N - 1}{n(n-1)} \sum_{k=1}^{3} \left( \frac{n_k}{N_k} - \frac{n}{N} \right) \left( n_{0k} - \frac{n_k}{n} n_0 \right).
\]

Its variance is

\[
\text{var}(\hat{r}_0^2) = \left( 1 - \hat{\rho}^2_{0\phi} \right) \left( \frac{1}{n} - \frac{f}{n} \sigma^2 \right),
\]

where \( \hat{\rho}_{0\phi} = \hat{\sigma}_{\star \phi} \Sigma_\star^{-1} \sigma_{\phi} / \sigma^2_0 \).

When sample estimates of variance and covariance are used

It can be shown that this estimator is similar to poststratified estimators. If the sample estimator of \( \Sigma_\star \) is used, we have \( \hat{\Sigma}_{\star}^{-1} = \frac{n - 1}{n_0 n_1 n_2} \begin{pmatrix} n_1 (n - n_2) & n_1 n_2 \\ n_1 n_2 & n_1 (n - n_1) \end{pmatrix} \),

\( \hat{\sigma}_{\phi} = \frac{1}{n(n-1)} \left( \frac{n n_{01} - n_0 n_1}{n n_{02} - n_0 n_2} \right) \) and thus

\[
\hat{\beta}_{\phi} = \begin{pmatrix}
\frac{1}{n_1} \left( \frac{n_01 - n_1}{n} \right) \\
\frac{1}{n_2} \left( \frac{n_{02} - n_2}{n} \right) - \frac{1}{n_3} \left( \frac{n_{03} - n_3}{n} \right) \end{pmatrix} \left( \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right).
\]
Therefore,

\[
\hat{r}_0 = \frac{n_0}{n} - \frac{1}{n} \sum_{k=1}^{\frac{3}{2}} \frac{N_k}{n_k} \left( \frac{n_k}{N_k} - \frac{n}{N} \right) \left( n_{0k} - \frac{n_k}{n} n_0 \right) = \frac{1}{n} \sum_{k=1}^{\frac{3}{2}} \frac{N_k}{N} \frac{n_{0k}}{n_k} .
\]  

(7.9)

Its variance is

\[
\text{var}(\hat{r}_0) = \frac{1 - f}{n} \frac{1}{n-1} \left( n_0 - \sum_{k=1}^{\frac{3}{2}} \frac{n_{0k}^2}{n_k} \right). 
\]

(7.10)

When \( p > 3 \) categories are present

Results (7.7), (7.8), (7.9) and (7.10) can be readily extended to scenarios when more than three categories are present.

A. When the population variance of the auxiliary variables are used:

\[
\hat{r}_0 = \frac{n_0}{n} - \frac{N-1}{n(n-1)} \sum_{k=1}^{p} \left( \left( \frac{n_k}{N_k} - \frac{n}{N} \right) \left( n_{0k} - \frac{n_k}{n} n_0 \right) \right),
\]

(7.11)

and

\[
\text{var}(\hat{r}_0) = \left( \frac{1 - f}{n} \right) \left( n_0 \left( n - n_0 \right) \right) - \frac{N-1}{n^2(n-1)^2} \sum_{k=1}^{p} \frac{1}{N_k} \left( n_{0k} - n_0 n_k \right)^2 \right). 
\]

(7.12)

B. When the sample variance of the auxiliary variables are used:

\[
\hat{r}_0 = \frac{n_0}{n} - \frac{1}{n} \sum_{k=1}^{p} \frac{N_k}{n_k} \left( \frac{n_k}{N_k} - \frac{n}{N} \right) \left( n_{0k} - \frac{n_k}{n} n_0 \right) = \sum_{k=1}^{p} \frac{N_k}{N} \frac{n_{0k}}{n_k} .
\]

(7.13)

and its variance is

\[
\text{var}(\hat{r}_0) = \frac{1 - f}{n(n-1)} \left( n_0 - \sum_{k=1}^{p} \frac{n_{0k}^2}{n_k} \right) . 
\]

(7.14)

Derivations of the above results are given in Section 7.6.
7.3. **Cases with two categorical variables**

There are generally two scenarios when two categorical auxiliary variables are involved, depending on whether the cell or marginal totals of the auxiliary variables are known. Suppose in a health-related survey, we are interested in estimating the prevalence of skipping medication behavior (binomial), and we have information about the subjects’ gender and age from Center for Medicare and Medicaid Services (CMS) administrative data. After a sample is taken, the gender and age of each sampled subject will be known. There are several scenarios of practical interest, based on the availability of information about subjects’ gender and age.

**Table 7.4. Two categorical auxiliary variable examples**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>65-80</td>
<td>&gt;80</td>
</tr>
<tr>
<td>Female</td>
<td>$T_1$</td>
<td>$T_2$</td>
</tr>
<tr>
<td>Male</td>
<td>$T_3$</td>
<td>$T_4$</td>
</tr>
<tr>
<td>Total</td>
<td>$T_{1}$</td>
<td>$T_{2}$</td>
</tr>
</tbody>
</table>

7.3.1. **Cases with known cell totals**

When all cell totals $(T_1, T_2, T_3, T_4)$ are known, we may create the following set of indicator variables:

$$y^{(i)} = \begin{cases} 1 & \text{Female, Aged 65-80 years,} \\ 0 & \text{Otherwise.} \end{cases}$$
With them, results from Section 5.4 can thus be directly applied to this situation.

To avoid singularity, we may include any three of the four categorical variables into the model. The estimator and its variance will be the same as (7.11) and (7.12), or (7.13) and (7.14), respectively.

7.3.2. Cases with only known marginal totals

When only marginal totals are known, we may use two indicator variables to improve precision of prediction, i.e., one for gender and the other for age category. We define the two indicator variables as

\[
y^{(1)} = \begin{cases} 
1 & \text{Male,} \\
0 & \text{Female,} 
\end{cases}
\]

and

\[
y^{(2)} = \begin{cases} 
1 & \text{Aged 65-80,} \\
0 & \text{Aged >80 years.} 
\end{cases}
\]

We denote the total number of male and female subjects as \(N_m, N_f\), respectively, the number of persons aged 65-80 and >80 years as \(N_a\) and \(N_b\).

We may apply results derived in Section 5.4. of Chapter 5 as follows,
\[ \hat{r}_0 = \bar{r}_0 - \beta_{0*} \left( \bar{Y}^{(*)} - \mu^{(*)} \right), \]  

(7.15)

where \( \beta_{0*} = \Sigma_{0*}^{-1} \sigma_{0*}, \) \( \mu_{0*} = \left( \frac{N_m}{N} \ rac{N_a}{N} \right)' \) and \( \bar{Y}^{(*)} = \left( \bar{Y}^{(1)} \ \bar{Y}^{(2)} \right)' \).

Since only the marginal totals are known, \( \Sigma_{0*} \) is undetermined because the population covariance between \( Y^{(1)} \) and \( Y^{(2)} \) is unknown. We may use the sample variance \( \hat{\Sigma}_{0*} \), the sample covariance between \( Y^{(0)} \) and \( Y^{(1)}, Y^{(2)}, \) \( \hat{\sigma}_{0*} \), and sample variance of \( Y^{(0)}, \) \( \hat{\sigma}_0^2 = n_0 (n - n_0) / n(n - 1) \). Since \( nn_{m,a} - n_m n_a = n_f n_{m,a} - n_m n_{f,a}, \) we have

\[
\hat{\Sigma}_{0*} = \frac{1}{n(n-1)} \begin{pmatrix} n_m n_f & n_{m,a} - n_m n_a \\ n_{m,a} - n_m n_a & n_a n_b \end{pmatrix} = \frac{n_m n_f}{n(n-1)} \begin{pmatrix} 1 & \frac{n_{m,a} - n_m n_a}{n_f} \\ \frac{n_{m,a} - n_m n_a}{n_f} & \frac{n_a n_b}{n_f} \end{pmatrix},
\]

Since \( nn_{0,m} - n_m n_0 = n_f n_{0,m} - n_m n_{0,f} \) and \( nn_{0,a} - n_a n_0 = n_b n_{0,a} - n_a n_{0,b}, \) we have

\[
\hat{\sigma}_{0*} = \frac{1}{n(n-1)} (n_{0,m} - n_m n_0) = \frac{1}{n(n-1)} \begin{pmatrix} n_m n_f & \frac{n_{0,m} - n_0 f}{n_f} \\ \frac{n_{0,m} - n_0 f}{n_f} & \frac{n_b n_{0,a} - n_a n_{0,b}}{n_b} \end{pmatrix}.
\]

For simplicity, we denote \( \Delta r_{m,a} = \left( \frac{n_{m,a}}{n} - \frac{n_{f,a}}{n_f} \right), \Delta r_{0,m} = \left( \frac{n_{0,m}}{n_m} - \frac{n_{0,f}}{n_f} \right), \)

\[
\Delta r_{0,a} = \left( \frac{n_{0,a}}{n_a} - \frac{n_{0,b}}{n_b} \right) \quad \text{and} \quad B = \frac{n_a n_b - n_m n_f (\Delta r_{m,a})^2}{n_a n_b - n_m n_f (\Delta r_{m,a})^2}.
\]

We have
\[ \hat{\Sigma}_{-1} = \frac{n(n-1)}{n_a n_b} B \left( \frac{n_a n_b}{n_m n_f} - \Delta r_{m,a} \right) \] and \( \hat{\sigma} = \frac{1}{n(n-1)} \left( \frac{n_m n_f}{n_a n_b} \Delta r_{0,m} \right) \), and thus

\[ \hat{\beta} = B \left( \Delta r_{0,m} - \Delta r_{m,a} \Delta r_{0,a} \right) \]

Therefore,

\[ \hat{r}_0 = \frac{n_0}{n} + B \left( \frac{N_m - n_m}{N} \right) \Delta r_{0,m} + B \left( \frac{N_a - n_a}{n} \right) \Delta r_{0,a} \]

\[ -B \left( \frac{N_m - n_m}{n} \right) \Delta r_{m,a} \Delta r_{0,a} - B \left( \frac{N_a - n_a}{n} \right) \left( \frac{n_m n_f}{n_a n_b} \right) \Delta r_{m,a} \Delta r_{0,m} \]

The variance of \( \hat{r}_0 \) is

\[ \text{var} \left( \hat{r}_0 \right) = \left( 1 - \frac{n}{n_0} \right)^2 \left( 1 - \frac{f}{n} \sigma^2 \right) \]

where \( \frac{n_0}{n(n-n_0)} B \left( n_m n_f \left( \Delta r_{0,m} \right)^2 + n_a n_b \left( \Delta r_{0,a} \right)^2 - 2 n_m n_f \Delta r_{0,m} \Delta r_{0,a} \right) \).

After algebraic simplification, we have

\[ \text{var} \left( \hat{r}_0 \right) = \frac{1-f}{n} \left( \frac{n_0 (n-n_0)}{n(n-1)} \right) \]

\[ - (1-f)(n-1) B \left( n_m n_f \left( \Delta r_{0,m} \right)^2 + n_a n_b \left( \Delta r_{0,a} \right)^2 - 2 n_m n_f \Delta r_{0,m} \Delta r_{0,a} \right) \]

### 7.4. Cases with single continuous auxiliary variable

Consider a prevalence survey of skipping medications in a low-income senior community, where an individual’s age is a strong predictor. However, the information about an individual’s age may or may not be available before sampling. Assuming such information of the sampled subjects will be available for analysis, the following
scenarios may happen, 1) only the population mean age is known; 2) ages are known for all individuals in the population.

For cases where only the population mean age is known, we may directly apply those results derived in Section 5.3 of Chapter 5 while using sample variance of age and sample covariance between the response variable and age.

For cases where ages are known for all subjects in the population, we may apply those results presented in Section 5.3. of Chapter 5 while using population variance of age and sample covariance between the response variable and age.

When age is known for all subjects in the population, the distribution information such as quartiles, octiles or deciles will also be known and can be used to improve precision of estimation. A question of practical interest is how such information should be used. An analyst may chose to,

*Strategy I:* Use age as a continuous variable, and follow the method of Section 7.4.1;

*Strategy II:* Use distribution properties of age, such as quartiles, octiles or deciles, and follow the strategies of Section 7.2.2.

Intuitively, Strategy II should give better estimators because more information is incorporated into estimation. The immediate question is how much detail of the distribution information should be used. We will answer this question by a series of Monte Carlo simulations. We will compared estimators based on continuous age, its median, quartiles, octiles and double-octiles.
7.4.1. Hypothetical populations

We generated two series of hypothetical populations of sizes 80, 160, 320, 640 and 1280, each series has a hypothetical relationship between smoking behavior and age as illustrated in Figure 7.4. The two series of hypothetical populations were held as fixed finite populations. In populations of Series A, the proportions of smokers increase monotonically with increases of age. In populations of Series B, the relationship between proportion of smokers and age is bell-shaped; the proportions of smokers increases with age until its fifth octile (age 71-75), then gradually declines. Age was set as uniformly distributed within each octiles. The characteristics of the two series of populations are summarized in Table 7.5.

7.4.2. Monte Carlo simulations

In all the simulations, we used the population gender proportions to improve estimation using methods developed in Sections 5.3 and 7.2. The study was carried out as follows. From each fixed finite population, a series of 5,000 SRSWOR samples of size \( n \), where \( n = 20, 40, 80, 160, 320, 640 \) and \( n < N \), were repeatedly drawn. For each sample, we estimated the proportion of smokers with (7.11) and its variance estimate with (7.12). The corresponding 95% confidence intervals (CIs) were constructed based on normal approximation. Age was treated as, 1) a continuous variable (CNT), 2) a binomial variable with cut-off point at its population median (MED), 3) a multinomial variable with cut-off points at its population quartiles (QRT), 4) a multinomial variable with cut-off points at its population octiles (OCT), and 5) a multinomial variable with cut-off points at its population sedeciles (SED, quantiles based on division into 16 groups of equal frequency), respectively.
We compared estimation methods in terms of their average mean squared errors and the overall performance of their 95% confidence intervals based on the estimated variance. We evaluated their confidence intervals in terms of the coverage rates, average length. Simulation results are summarized in Tables 7.6, 7.7, 7.8, 7.9 and 7.10 and presented graphically in Figures 7.5, 7.6, 7.7 and 7.8. Based on these simulations, we made the following observations.

7.4.3. Comparison between the age-adjusted and the simple expansion predictors

When the relationship between smoker proportion and age is monotonic (population series A), as shown in Table 7.6, MSEs were significantly reduced by adjustment for age mean (CNT), median (MED) or quartiles (QTR) even when sample sizes were smaller (n=20, 40, 80). When sample sizes became ≥160, it became more efficient to adjust for octiles or sedeciles than the simple expansion predictors (SEP). Regardless of sample sizes, it was more efficient to adjust for age mean than for age categories, however, the differences were gradually disappearing when sample size became large (n ≥ 320). When sample sizes were small (n≤40), adjusting for age octiles or sedeciles resulted in larger MSE than the SEP.

When the relationship is not monotonic (population series B), as shown in Table 7.7, the estimators adjusted for age mean, median, octiles and sedeciles were less efficient than the SEPs when sample sizes were ≤80. When sample sizes became ≥ 40, estimators adjusted for age quartiles became more efficient; and when sample sizes became ≥ 80, the estimators adjusted for either octiles or sedeciles became more efficient. Finally, when sample sizes became moderate (n≤160), the estimators adjusted for age median, quartiles, octiles or sedeciles became more efficient than the SEP.
Figure 7.4. Hypothetical relationship between smoking and age of population series A and B

Population Series A

Population Series B
Table 7.5. Parameters of hypothetical population Series A and B

### Population Series A

<table>
<thead>
<tr>
<th>Octiles</th>
<th>Range</th>
<th>Population Size (N)</th>
<th>Age mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>160</td>
</tr>
<tr>
<td>1</td>
<td>60.0-62.0</td>
<td>10 .1</td>
<td>20 .1</td>
</tr>
<tr>
<td>2</td>
<td>62.1-64.5</td>
<td>10 .2</td>
<td>20 .2</td>
</tr>
<tr>
<td>3</td>
<td>64.6-67.5</td>
<td>10 .3</td>
<td>20 .3</td>
</tr>
<tr>
<td>4</td>
<td>67.6-71.0</td>
<td>10 .4</td>
<td>20 .4</td>
</tr>
<tr>
<td>5</td>
<td>71.1-75.0</td>
<td>10 .5</td>
<td>20 .5</td>
</tr>
<tr>
<td>6</td>
<td>75.1-80.0</td>
<td>10 .6</td>
<td>20 .6</td>
</tr>
<tr>
<td>7</td>
<td>80.1-85.0</td>
<td>10 .7</td>
<td>20 .7</td>
</tr>
<tr>
<td>8</td>
<td>85.1-100</td>
<td>10 .8</td>
<td>20 .8</td>
</tr>
<tr>
<td>Overall</td>
<td>60.0-62.0</td>
<td>.45</td>
<td>.45</td>
</tr>
<tr>
<td>Age mean</td>
<td></td>
<td>73.0</td>
<td>73.2</td>
</tr>
</tbody>
</table>

### Population Series B

<table>
<thead>
<tr>
<th>Octiles</th>
<th>Range</th>
<th>Population Size (N)</th>
<th>Age mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>160</td>
</tr>
<tr>
<td>1</td>
<td>60.0-62.0</td>
<td>10 .1</td>
<td>20 .1</td>
</tr>
<tr>
<td>2</td>
<td>62.1-64.5</td>
<td>10 .1</td>
<td>20 .1</td>
</tr>
<tr>
<td>3</td>
<td>64.6-67.5</td>
<td>10 .15</td>
<td>20 .15</td>
</tr>
<tr>
<td>4</td>
<td>67.6-71.0</td>
<td>10 .4</td>
<td>20 .4</td>
</tr>
<tr>
<td>5</td>
<td>71.1-75.0</td>
<td>10 .6</td>
<td>20 .6</td>
</tr>
<tr>
<td>6</td>
<td>75.1-80.0</td>
<td>10 .4</td>
<td>20 .4</td>
</tr>
<tr>
<td>7</td>
<td>80.1-85.0</td>
<td>10 .2</td>
<td>20 .2</td>
</tr>
<tr>
<td>8</td>
<td>85.1-100</td>
<td>10 .1</td>
<td>20 .1</td>
</tr>
<tr>
<td>Overall</td>
<td>60.0-62.0</td>
<td>.263</td>
<td>.268</td>
</tr>
<tr>
<td>Age mean</td>
<td></td>
<td>73.0</td>
<td>73.1</td>
</tr>
</tbody>
</table>
Table 7.6. MSE of design-based estimators adjusted to known continuous or categorical auxiliary variables, by population size and sample size (Series A)

<table>
<thead>
<tr>
<th>Sample size and Population size</th>
<th>Estimation method</th>
<th>SEP</th>
<th>CNT</th>
<th>MED</th>
<th>QRT</th>
<th>OCT</th>
<th>SED</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 80</td>
<td></td>
<td>96.3</td>
<td>81.5</td>
<td>86.0</td>
<td>93.0</td>
<td>114.8</td>
<td>143.7</td>
</tr>
<tr>
<td>160</td>
<td></td>
<td>109.7</td>
<td>96.5</td>
<td>97.4</td>
<td>105.6</td>
<td>129.5</td>
<td>172.2</td>
</tr>
<tr>
<td>320</td>
<td></td>
<td>117.7</td>
<td>102.3</td>
<td>104.5</td>
<td>113.5</td>
<td>138.0</td>
<td>188.6</td>
</tr>
<tr>
<td>640</td>
<td></td>
<td>119.3</td>
<td>108.0</td>
<td>106.9</td>
<td>117.7</td>
<td>143.4</td>
<td>195.5</td>
</tr>
<tr>
<td>1280</td>
<td></td>
<td>123.8</td>
<td>106.8</td>
<td>108.4</td>
<td>117.5</td>
<td>143.1</td>
<td>195.8</td>
</tr>
</tbody>
</table>

| 40 80                          |                  | 30.4| 25.0| 27.2| 28.1| 31.6| 36.4|
| 160                            |                  | 45.8| 39.8| 39.7| 41.4| 47.2| 54.9|
| 320                            |                  | 55.3| 45.9| 47.3| 48.2| 54.4| 66.3|
| 640                            |                  | 57.0| 48.1| 48.7| 50.0| 56.7| 70.1|
| 1280                           |                  | 58.0| 48.6| 50.4| 51.3| 57.0| 69.5|

| 80 160                         |                  | 15.3| 12.7| 13.1| 13.1| 13.8| 14.6|
| 320                            |                  | 23.3| 19.1| 19.6| 19.4| 20.6| 23.1|
| 640                            |                  | 26.6| 22.1| 22.8| 22.6| 24.1| 27.3|
| 1280                           |                  | 29.2| 23.8| 24.7| 24.6| 25.9| 28.9|

| 160 320                        |                  | 7.8 | 6.3 | 6.6 | 6.3 | 6.6 | 7.0 |
| 640                            |                  | 11.2| 9.2 | 9.6 | 9.3 | 9.5 | 10.1|
| 1280                           |                  | 13.3| 11.0| 11.4| 11.1| 11.4| 12.0|

| 320 640                        |                  | 3.9 | 3.2 | 3.3 | 3.2 | 3.2 | 3.3 |
| 1280                           |                  | 5.8 | 4.7 | 5.0 | 4.7 | 4.8 | 4.9 |

| 640 1280                       |                  | 1.9 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 |

SEP = Simple expansion predictor, CNT = Design-based predictor (DP) using continuous age, MED = DP using median age, QRT = DP using quartiles of age, OCT= DP using octiles of age, SED=DP using sedeciles (16-quantiles) of age.
Table 7.7. MSE of design-based estimators adjusted to known continuous or categorical auxiliary variables, by population size and sample size (Series B)

<table>
<thead>
<tr>
<th>Sample size and Population size</th>
<th>Estimation method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEP</td>
</tr>
<tr>
<td>20 80</td>
<td>73.6</td>
</tr>
<tr>
<td>160 80</td>
<td>87.5</td>
</tr>
<tr>
<td>320 80</td>
<td>93.1</td>
</tr>
<tr>
<td>640 80</td>
<td>96.8</td>
</tr>
<tr>
<td>1280 80</td>
<td>98.8</td>
</tr>
<tr>
<td>40 80</td>
<td>25.0</td>
</tr>
<tr>
<td>160 80</td>
<td>38.5</td>
</tr>
<tr>
<td>320 80</td>
<td>41.2</td>
</tr>
<tr>
<td>640 80</td>
<td>45.3</td>
</tr>
<tr>
<td>1280 80</td>
<td>45.7</td>
</tr>
<tr>
<td>80 160</td>
<td>12.4</td>
</tr>
<tr>
<td>320 160</td>
<td>18.5</td>
</tr>
<tr>
<td>640 160</td>
<td>21.1</td>
</tr>
<tr>
<td>1280 160</td>
<td>22.3</td>
</tr>
<tr>
<td>160 320</td>
<td>6.2</td>
</tr>
<tr>
<td>640 320</td>
<td>9.0</td>
</tr>
<tr>
<td>1280 320</td>
<td>10.4</td>
</tr>
<tr>
<td>320 640</td>
<td>3.2</td>
</tr>
<tr>
<td>1280 640</td>
<td>4.5</td>
</tr>
<tr>
<td>640 1280</td>
<td>1.5</td>
</tr>
</tbody>
</table>

SEP = Simple expansion predictor, CNT = Design-based predictor (DP) using continuous age, MED = DP using median age, QRT = DP using quartiles of age, OCT= DP using octiles of age, SED=DP using sedeciles (16-quantiles) of age.
7.4.4. MSEs and variances of the estimators when relationship between smoker proportion and age is monotonic (Series A)

When the relationship between smoker proportions and age is monotonic, as shown in Tables 7.6 and Figures 7.5 and 7.6, the estimators adjusted for age categories had larger MSEs than the estimators adjusted for age mean (CNT, thereafter). The more categories the age variable had, the larger their MSEs were. For \( n \leq 80 \), estimators adjusted for age quartiles, octiles and sedeciles had larger MSEs than those adjusted for mean or median age. When the sample became moderate or large \( (n \geq 160) \), the MSEs of the estimators adjusted for age quartiles or octiles became smaller than those of estimators adjusted for median age, and became close to MSEs of the estimators adjusted for mean age.

Figure 7.6 presents ratios of MSE or variance of MED, QTR, OCT and SED estimators to MSE or variance of CNT estimators, respectively. The MED estimators had slightly larger MSEs and variances than CNT estimators, their ratios of MSE and variance were between 1 and 1.05. However, the MSE ratios of QTR, OCT and SED estimators were all larger than 1, regardless of smaller or larger sample sizes. It is obvious that the smaller the sample sizes were, the greater the MSE ratios were. In contrast, the smaller the sample sizes were, the smaller the variance ratios were, which indicated the overall underestimated the variances of estimators using sample-based variance estimators. The smaller the sample size was, the larger the difference was between MSE and variance ratios. This phenomena raises questions about the validity of the 95% confidence intervals constructed using sample-based variance estimators.
Figure 7.5. MSE and variance of design-based predictors adjusted for continuous or categorical age (Series A)

Sample sizes (from top to bottom): 40, 80, 160, 320
Symbols: - =CNT, o = MED, x = QRT, += OCT, D = SED

MSE x 10,000

Variance x 10,000

Population size

Sample sizes (from top to bottom): 40, 80, 160, 320
Symbols: - = SRS, o = CNT, x = MED, += QTR, D = SED

Population size
Figure 7.6. MSE and variance ratios of estimators adjusted for age median, quartiles, octiles and sedeciles to estimators adjusted for age mean (Series A)

Sample Sizes: - = 20, o=40, x=80, +=160, T=320, S=640

**MSE ratio**

- **Estimation method: Median**
  - Population size: 80, 160, 320, 640, 1280
  - Ratios: 1, 1.05, 1.1, 1.15, 1.2

- **Estimation method: Quartiles**
  - Population size: 80, 160, 320, 640, 1280
  - Ratios: 1, 1.05, 1.1, 1.15, 1.2

- **Estimation method: Octiles**
  - Population size: 80, 160, 320, 640, 1280
  - Ratios: 1, 1.25, 1.5, 1.75, 2

- **Estimation method: Sedeciles**
  - Population size: 80, 160, 320, 640, 1280
  - Ratios: 1, 1.25, 1.5, 1.75, 2

**Variance ratio**

- **Estimation method: Median**
  - Population size: 80, 160, 320, 640, 1280
  - Ratios: .85, .9, .95, 1, 1.05

- **Estimation method: Quartiles**
  - Population size: 80, 160, 320, 640, 1280
  - Ratios: .85, .9, .95, 1, 1.05

- **Estimation method: Octiles**
  - Population size: 80, 160, 320, 640, 1280
  - Ratios: .3, .4, .5, .6, .7, .8, .9, 1

- **Estimation method: Sedeciles**
  - Population size: 80, 160, 320, 640, 1280
  - Ratios: .3, .4, .5, .6, .7, .8, .9, 1
7.4.5. MSEs and variances of the estimators when relationship between smoker proportion and age is not monotonic (Series B)

When the relationship between smoker proportions and age is not monotonic, as shown in Tables 7.7 and Figures 7.7 and 7.8, the estimators adjusted for age median (MED) or quartiles (QTR) had smaller MSEs than the estimators adjusted for age mean (CNT), regardless of population or sample sizes. When sample sizes became larger than 80, the estimators adjusted for age octiles (OCT) became more efficient than the CNT estimators. When sample sizes were larger than 160, the estimators adjusted for age sedeciles (SED) became more efficient than the CNT and MED estimators.

When sample sizes were small (n<160) or sampling fractions were high \( f \geq 25\% \), the variance ratios of the QTR to the CNT estimators were much smaller than the MSE ratios. Such effects were more pronounced for the variance ratios of the OCT and SED to the CNT estimators. This indicated overall underestimation of the variability of the estimators using sample-based variance estimators.

7.4.6. Negative variances

The variance estimates were calculated using (7.12). As shown in Table 7.8, the probability of encountering negative variance estimates increased dramatically with the number of categories the age variable had when sample sizes were small (n=20 or 40). Since the sample covariance between response and auxiliary variable(s) were used for estimation, the higher probabilities of encountering negative variance estimates might be attributable to the larger variation of sample covariance when sample sizes were small.
Figure 7.7. MSE and variance of design-based predictors adjusted for continuous or categorical age (Series B)

Sample sizes (from top to bottom): 40, 80, 160, 320
Symbols: - =CNT, o=MED, x=QRT, +=OCT, D=SED

MSE x 10,000

Variance x 10,000

Sample sizes (from top to bottom): 40, 80, 160, 320
Symbols: .=SRS, o=CNT, x=MED, +=QTR, D=SED
Figure 7.8. MSE and variance ratios of estimators adjusted for age median, quartiles, octiles and sedeciles to estimators adjusted for age mean (Series B)

Sample Sizes: \( - = 20, o=40, x=80, +=160, T=320, D=640 \)

**MSE ratio**

- **Estimation method: Median**
- **Estimation method: Quartiles**
- **Estimation method: Octiles**
- **Estimation method: Sedeciles**

---

**Variance ratio**

- **Estimation method: Median**
- **Estimation method: Quartiles**
- **Estimation method: Octiles**
- **Estimation method: Sedeciles**
Table 7.8. Probability of encountering negative variance when sample size is small and age is used as a continuous or categorical auxiliary variable

<table>
<thead>
<tr>
<th>Sample size and Population size</th>
<th>Estimation method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEP</td>
</tr>
<tr>
<td>20</td>
<td>---</td>
</tr>
<tr>
<td>160</td>
<td>---</td>
</tr>
<tr>
<td>320</td>
<td>---</td>
</tr>
<tr>
<td>640</td>
<td>---</td>
</tr>
<tr>
<td>1280</td>
<td>---</td>
</tr>
<tr>
<td>40</td>
<td>---</td>
</tr>
<tr>
<td>160</td>
<td>---</td>
</tr>
<tr>
<td>320</td>
<td>---</td>
</tr>
<tr>
<td>640</td>
<td>---</td>
</tr>
<tr>
<td>1280</td>
<td>---</td>
</tr>
</tbody>
</table>

--- = not apply or negligible (<0.1%). SEP = Simple expansion predictor, CNT = Design-based predictor (DP) using continuous age, MED = DP using median of age, QRT = DP using quartiles of age, OCT = DP using octiles of age, SED = DP sedeciles (16-quantiles) of age. When sample size ≥ 80, the probability of encountering negative variance is negligible (<0.1%).
7.4.7. Performance of 95% confidence intervals

The 95% confidence intervals were constructed using normal approximation with variance estimates based on (7.12). Regardless of the functional form between smoker proportion and age, the 95% confidence intervals (CI) of the estimators adjusted for age mean and quantiles performed worse than simple expansion estimators when sample sizes were very small (n=20). The more categories the variables had, the poorer the coverage rates were. When sample size is 20, the 95% CIs constructed based on age octiles or sedeciles had coverage rates as low as 70% or 30%, respectively. When sample sizes became moderate (n=80), the 95% CIs for estimators adjusted for age mean, median and quartiles had coverage rates close to its nominal level. Not until sample sizes became larger (n=320 or 640), the 95% CIs for estimators adjusted for age octiles or sedeciles had coverage rates close to their nominal level.

Results in Tables 7.9 and 7.10 suggested that the width of the confidence intervals became increasingly narrower when the number of categories of age increased, especially when sample sizes were small. When the relationship between smoker proportions and age was not monotonic and sample size was moderate or large (n≥320), the 95% CIs of the estimators adjusted for age quartiles, octiles or sedeciles were much narrower than those adjusted for age mean or median, yet their coverage rates close to their nominal level.
Table 7.9. Coverage rates and width of 95% confidence interval using age as a continuous or categorical variable, by population size and sample size (Series A)

<table>
<thead>
<tr>
<th>Sample size and Population size</th>
<th>SEP</th>
<th>CNT</th>
<th>MED</th>
<th>QRT</th>
<th>OCT</th>
<th>SED</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 80</td>
<td>92.9</td>
<td>89.0</td>
<td>90.5</td>
<td>85.0</td>
<td>71.9</td>
<td>32.6</td>
</tr>
<tr>
<td></td>
<td>37.9</td>
<td>31.1</td>
<td>32.7</td>
<td>29.9</td>
<td>25.5</td>
<td>16.8</td>
</tr>
<tr>
<td>160</td>
<td>93.8</td>
<td>89.6</td>
<td>90.2</td>
<td>84.1</td>
<td>69.6</td>
<td>26.0</td>
</tr>
<tr>
<td></td>
<td>40.9</td>
<td>34.6</td>
<td>35.1</td>
<td>31.6</td>
<td>26.6</td>
<td>17.1</td>
</tr>
<tr>
<td>320</td>
<td>93.4</td>
<td>89.8</td>
<td>91.1</td>
<td>83.9</td>
<td>69.0</td>
<td>28.3</td>
</tr>
<tr>
<td></td>
<td>42.2</td>
<td>35.5</td>
<td>36.4</td>
<td>32.7</td>
<td>27.2</td>
<td>18.6</td>
</tr>
<tr>
<td>640</td>
<td>93.1</td>
<td>89.6</td>
<td>90.6</td>
<td>83.7</td>
<td>68.4</td>
<td>25.8</td>
</tr>
<tr>
<td></td>
<td>42.9</td>
<td>36.0</td>
<td>36.8</td>
<td>32.9</td>
<td>27.1</td>
<td>18.4</td>
</tr>
<tr>
<td>1280</td>
<td>92.3</td>
<td>89.5</td>
<td>90.5</td>
<td>84.7</td>
<td>69.1</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>43.2</td>
<td>36.1</td>
<td>37.0</td>
<td>33.3</td>
<td>27.3</td>
<td>18.9</td>
</tr>
<tr>
<td>40 80</td>
<td>95.8</td>
<td>92.9</td>
<td>92.9</td>
<td>91.4</td>
<td>87.9</td>
<td>78.3</td>
</tr>
<tr>
<td></td>
<td>21.9</td>
<td>18.8</td>
<td>19.6</td>
<td>18.8</td>
<td>17.9</td>
<td>15.2</td>
</tr>
<tr>
<td>160</td>
<td>96.2</td>
<td>93.5</td>
<td>93.6</td>
<td>91.1</td>
<td>86.8</td>
<td>72.9</td>
</tr>
<tr>
<td></td>
<td>26.8</td>
<td>23.6</td>
<td>23.9</td>
<td>22.6</td>
<td>21.2</td>
<td>16.7</td>
</tr>
<tr>
<td>320</td>
<td>93.4</td>
<td>92.4</td>
<td>92.9</td>
<td>90.5</td>
<td>85.6</td>
<td>72.0</td>
</tr>
<tr>
<td></td>
<td>28.9</td>
<td>25.1</td>
<td>25.7</td>
<td>24.2</td>
<td>22.4</td>
<td>18.6</td>
</tr>
<tr>
<td>640</td>
<td>93.5</td>
<td>93.1</td>
<td>93.4</td>
<td>91.2</td>
<td>85.9</td>
<td>71.0</td>
</tr>
<tr>
<td></td>
<td>29.9</td>
<td>26.0</td>
<td>26.5</td>
<td>24.9</td>
<td>22.8</td>
<td>18.6</td>
</tr>
<tr>
<td>1280</td>
<td>95.0</td>
<td>93.0</td>
<td>93.5</td>
<td>91.1</td>
<td>86.4</td>
<td>71.1</td>
</tr>
<tr>
<td></td>
<td>30.4</td>
<td>26.2</td>
<td>26.9</td>
<td>25.3</td>
<td>23.1</td>
<td>18.7</td>
</tr>
<tr>
<td>80 160</td>
<td>96.2</td>
<td>94.3</td>
<td>94.4</td>
<td>93.4</td>
<td>91.7</td>
<td>87.4</td>
</tr>
<tr>
<td></td>
<td>15.5</td>
<td>13.9</td>
<td>14.0</td>
<td>13.5</td>
<td>13.2</td>
<td>11.9</td>
</tr>
<tr>
<td>320</td>
<td>94.9</td>
<td>93.9</td>
<td>94.7</td>
<td>93.4</td>
<td>91.4</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>18.9</td>
<td>16.7</td>
<td>17.1</td>
<td>16.4</td>
<td>15.9</td>
<td>14.9</td>
</tr>
<tr>
<td>640</td>
<td>94.6</td>
<td>94.6</td>
<td>94.7</td>
<td>93.7</td>
<td>91.9</td>
<td>86.8</td>
</tr>
<tr>
<td></td>
<td>20.4</td>
<td>18.1</td>
<td>18.4</td>
<td>17.7</td>
<td>17.1</td>
<td>15.8</td>
</tr>
<tr>
<td>1280</td>
<td>95.4</td>
<td>94.1</td>
<td>93.9</td>
<td>93.1</td>
<td>90.9</td>
<td>85.7</td>
</tr>
<tr>
<td></td>
<td>21.1</td>
<td>18.6</td>
<td>19.1</td>
<td>18.3</td>
<td>17.6</td>
<td>16.2</td>
</tr>
<tr>
<td>160 320</td>
<td>94.3</td>
<td>94.8</td>
<td>94.5</td>
<td>94.4</td>
<td>93.8</td>
<td>91.7</td>
</tr>
<tr>
<td></td>
<td>10.9</td>
<td>9.7</td>
<td>9.9</td>
<td>9.7</td>
<td>9.5</td>
<td>9.3</td>
</tr>
<tr>
<td>640</td>
<td>95.2</td>
<td>95.1</td>
<td>94.9</td>
<td>94.1</td>
<td>93.5</td>
<td>91.7</td>
</tr>
<tr>
<td></td>
<td>13.4</td>
<td>12.0</td>
<td>12.2</td>
<td>11.8</td>
<td>11.6</td>
<td>11.2</td>
</tr>
<tr>
<td>1280</td>
<td>95.0</td>
<td>94.5</td>
<td>94.5</td>
<td>94.0</td>
<td>93.1</td>
<td>91.6</td>
</tr>
<tr>
<td></td>
<td>14.4</td>
<td>12.8</td>
<td>13.1</td>
<td>12.7</td>
<td>12.4</td>
<td>12.0</td>
</tr>
<tr>
<td>320 640</td>
<td>95.5</td>
<td>94.7</td>
<td>94.8</td>
<td>94.8</td>
<td>94.3</td>
<td>93.6</td>
</tr>
<tr>
<td></td>
<td>7.7</td>
<td>6.9</td>
<td>7.0</td>
<td>6.9</td>
<td>6.8</td>
<td>6.7</td>
</tr>
<tr>
<td>1280</td>
<td>95.1</td>
<td>94.8</td>
<td>95.0</td>
<td>94.8</td>
<td>94.2</td>
<td>93.2</td>
</tr>
<tr>
<td></td>
<td>9.4</td>
<td>8.4</td>
<td>8.6</td>
<td>8.4</td>
<td>8.3</td>
<td>8.1</td>
</tr>
<tr>
<td>640 1280</td>
<td>95.7</td>
<td>94.9</td>
<td>94.8</td>
<td>95.0</td>
<td>94.8</td>
<td>94.3</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>4.9</td>
<td>5.0</td>
<td>4.9</td>
<td>4.8</td>
<td>4.8</td>
</tr>
</tbody>
</table>

SEP = Simple expansion predictor, CNT = Design-based predictor (DP) using continuous age, MED = DP using median of age, QRT = DP using quartiles of age, OCT= DP using octiles of age, SED=DP using sedeciles (16-quantiles) of age.
Table 7.10. Coverage rates and width of 95% confidence interval using age as a continuous or categorical variable, by population and sample sizes (Series B)

<table>
<thead>
<tr>
<th>Sample size and Population size</th>
<th>SEP</th>
<th>CNT</th>
<th>MED</th>
<th>QRT</th>
<th>OCT</th>
<th>SED</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 80</td>
<td>94.7</td>
<td>92.5</td>
<td>91.6</td>
<td>86.7</td>
<td>75.7</td>
<td>36.0</td>
</tr>
<tr>
<td>160</td>
<td>33.3</td>
<td>32.0</td>
<td>31.4</td>
<td>27.9</td>
<td>23.9</td>
<td>15.3</td>
</tr>
<tr>
<td>160 80</td>
<td>92.8</td>
<td>91.6</td>
<td>91.0</td>
<td>86.3</td>
<td>73.5</td>
<td>34.8</td>
</tr>
<tr>
<td>320</td>
<td>36.1</td>
<td>34.6</td>
<td>34.0</td>
<td>30.1</td>
<td>25.2</td>
<td>16.1</td>
</tr>
<tr>
<td>320 80</td>
<td>91.7</td>
<td>90.5</td>
<td>89.7</td>
<td>85.0</td>
<td>71.4</td>
<td>32.9</td>
</tr>
<tr>
<td>640</td>
<td>37.3</td>
<td>35.7</td>
<td>35.1</td>
<td>30.9</td>
<td>25.6</td>
<td>17.0</td>
</tr>
<tr>
<td>640 80</td>
<td>91.4</td>
<td>90.2</td>
<td>90.1</td>
<td>85.2</td>
<td>72.0</td>
<td>32.3</td>
</tr>
<tr>
<td>1280</td>
<td>37.9</td>
<td>36.2</td>
<td>35.6</td>
<td>31.2</td>
<td>25.9</td>
<td>17.0</td>
</tr>
<tr>
<td>1280 80</td>
<td>91.3</td>
<td>90.3</td>
<td>90.2</td>
<td>85.1</td>
<td>71.4</td>
<td>30.3</td>
</tr>
<tr>
<td>40 80</td>
<td>91.6</td>
<td>91.9</td>
<td>92.3</td>
<td>91.4</td>
<td>88.2</td>
<td>79.2</td>
</tr>
<tr>
<td>160</td>
<td>19.4</td>
<td>19.0</td>
<td>18.8</td>
<td>17.5</td>
<td>16.6</td>
<td>14.1</td>
</tr>
<tr>
<td>320</td>
<td>94.7</td>
<td>93.6</td>
<td>93.0</td>
<td>90.8</td>
<td>87.1</td>
<td>76.3</td>
</tr>
<tr>
<td>320 80</td>
<td>23.8</td>
<td>23.3</td>
<td>23.1</td>
<td>21.3</td>
<td>19.8</td>
<td>16.7</td>
</tr>
<tr>
<td>640</td>
<td>94.3</td>
<td>93.6</td>
<td>93.6</td>
<td>91.8</td>
<td>87.1</td>
<td>76.1</td>
</tr>
<tr>
<td>640 80</td>
<td>25.7</td>
<td>25.2</td>
<td>24.9</td>
<td>22.9</td>
<td>21.1</td>
<td>17.6</td>
</tr>
<tr>
<td>1280</td>
<td>92.8</td>
<td>92.6</td>
<td>92.9</td>
<td>90.8</td>
<td>86.3</td>
<td>73.4</td>
</tr>
<tr>
<td>1280 80</td>
<td>26.5</td>
<td>26.0</td>
<td>25.6</td>
<td>23.5</td>
<td>21.6</td>
<td>18.0</td>
</tr>
<tr>
<td>80 160</td>
<td>94.5</td>
<td>94.4</td>
<td>94.2</td>
<td>94.4</td>
<td>92.6</td>
<td>87.6</td>
</tr>
<tr>
<td>160</td>
<td>13.8</td>
<td>13.7</td>
<td>13.5</td>
<td>12.7</td>
<td>12.3</td>
<td>11.6</td>
</tr>
<tr>
<td>320</td>
<td>93.9</td>
<td>93.8</td>
<td>93.8</td>
<td>92.9</td>
<td>91.2</td>
<td>87.8</td>
</tr>
<tr>
<td>320 160</td>
<td>16.8</td>
<td>16.6</td>
<td>16.5</td>
<td>15.5</td>
<td>14.8</td>
<td>13.9</td>
</tr>
<tr>
<td>640</td>
<td>93.9</td>
<td>94.2</td>
<td>94.2</td>
<td>94.1</td>
<td>92.1</td>
<td>88.7</td>
</tr>
<tr>
<td>640 160</td>
<td>18.2</td>
<td>18.0</td>
<td>17.8</td>
<td>16.7</td>
<td>15.9</td>
<td>14.9</td>
</tr>
<tr>
<td>1280</td>
<td>95.1</td>
<td>94.6</td>
<td>94.4</td>
<td>93.7</td>
<td>91.9</td>
<td>87.7</td>
</tr>
<tr>
<td>1280 160</td>
<td>18.8</td>
<td>18.6</td>
<td>18.4</td>
<td>17.2</td>
<td>16.4</td>
<td>15.3</td>
</tr>
<tr>
<td>160 320</td>
<td>95.4</td>
<td>95.2</td>
<td>94.5</td>
<td>94.4</td>
<td>94.0</td>
<td>92.6</td>
</tr>
<tr>
<td>320</td>
<td>9.7</td>
<td>9.7</td>
<td>9.6</td>
<td>9.1</td>
<td>8.9</td>
<td>8.6</td>
</tr>
<tr>
<td>640</td>
<td>94.8</td>
<td>94.6</td>
<td>94.7</td>
<td>94.4</td>
<td>93.7</td>
<td>92.2</td>
</tr>
<tr>
<td>640 320</td>
<td>11.9</td>
<td>11.8</td>
<td>11.7</td>
<td>11.1</td>
<td>10.8</td>
<td>10.5</td>
</tr>
<tr>
<td>1280</td>
<td>95.4</td>
<td>94.6</td>
<td>94.8</td>
<td>94.6</td>
<td>93.8</td>
<td>91.8</td>
</tr>
<tr>
<td>1280 320</td>
<td>12.9</td>
<td>12.8</td>
<td>12.7</td>
<td>11.9</td>
<td>11.6</td>
<td>11.2</td>
</tr>
<tr>
<td>320 640</td>
<td>94.9</td>
<td>94.8</td>
<td>94.6</td>
<td>94.6</td>
<td>94.2</td>
<td>93.5</td>
</tr>
<tr>
<td>640</td>
<td>6.9</td>
<td>6.9</td>
<td>6.8</td>
<td>6.4</td>
<td>6.3</td>
<td>6.2</td>
</tr>
<tr>
<td>640 640</td>
<td>95.5</td>
<td>95.2</td>
<td>95.4</td>
<td>95.2</td>
<td>94.7</td>
<td>93.9</td>
</tr>
<tr>
<td>1280</td>
<td>8.4</td>
<td>8.4</td>
<td>8.3</td>
<td>7.9</td>
<td>7.7</td>
<td>7.6</td>
</tr>
<tr>
<td>1280 640</td>
<td>95.2</td>
<td>95.2</td>
<td>95.2</td>
<td>95.1</td>
<td>94.7</td>
<td>94.4</td>
</tr>
<tr>
<td>640</td>
<td>4.9</td>
<td>4.9</td>
<td>4.8</td>
<td>4.6</td>
<td>4.5</td>
<td>4.5</td>
</tr>
</tbody>
</table>

SRS = Simple expansion predictor, CNT = Design-based predictor (DP) using continuous age, MED = DP using median of age, QRT = DP using quartiles of age, OCT= DP using octiles of age, SED=DP using sedeciles (16-quantiles) of age.
In conclusion, the estimators adjusted for age categories were more efficient than the estimators adjusted for age mean when the functional form between smoker proportions and age is not monotonic. When sample size became large, such efficiency became more apparent. However, we caution the use of categorical auxiliary variable to improve rate estimation when sample size is small due to the higher probability of encountering negative variances and overall underestimate of the variances. For the simulations with sample size less than 40, using more than 2 categories resulted in loss of efficiency. When sample size is greater than 80 and less than 160, age quartiles may be used. When the sample size is moderate (n=320 or 640), octiles may be used. As a rough guild based on the limited simulations, the average number per group in the sample should not be less than 20.

7.5. Derivations of results (7.1) and (7.2) in Section 7.2.

By applying results derived in Section 5.4 of Chapter 5, we have

\[
\hat{r}_0 = \bar{r}_0 - \beta_0' \left( \bar{Y}^{(*)} - \mu^{(*)} \right), \tag{7.16}
\]

where \( \beta_{0,*} = \Sigma^{-1} \sigma_{y0} \), \( \mu_* = \left( \frac{N_1}{N} \frac{N_2}{N} \right)' \) and \( \bar{Y}^{(*)} = \left( \bar{Y}^{(1)} \bar{Y}^{(2)} \right)'\).

Since \( N_1 \) and \( N_2 \) are known, we have

\[
\Sigma_* = \frac{N_1 N_2}{N(N-1)} \begin{pmatrix} N_2^{-1}(N-N_1) & -1 \\ -1 & N_1^{-1}(N-N_2) \end{pmatrix},
\]

\[
\Sigma^{-1}_* = \frac{N-1}{N_1 N_2 N_3} \begin{pmatrix} N_2 (N-N_2) & N_1 N_2 \\ N_1 N_2 & N_1(N-N_1) \end{pmatrix}.
\]

Since \( \hat{\sigma}^2_0 = \frac{n_0(n-n_0)}{n(n-1)} \) and \( \hat{\sigma}_{y0} = \frac{1}{n(n-1)} \begin{pmatrix} m_{01}n_1 - n_0 n_1 \\ m_{02}n_2 - n_0 n_2 \end{pmatrix} \), we have
\[
\hat{\beta}_{0,\bullet} = \frac{1}{n(n-1)} N_{1} N_{2} N_{3} \left( \begin{array}{cc}
N_{2} (N - N_{2}) & N_{1} N_{2} \\
N_{1} N_{2} & N_{1} (N - N_{1})
\end{array} \right) \left( \begin{array}{c}
n_{n_{01}} - n_{0} n_{1} \\
n_{n_{02}} - n_{0} n_{2}
\end{array} \right)
\]
\[
= \frac{N - 1}{n(n-1)} \left( \begin{array}{c}
N_{1} N_{2} (n_{n_{01}} - n_{0} n_{1}) + N_{1} N_{2} (n_{n_{02}} - n_{0} n_{2}) \\
N_{1} N_{2} (n_{n_{01}} - n_{0} n_{1}) + N_{1} (N - N_{1}) (n_{n_{02}} - n_{0} n_{2})
\end{array} \right)
\]
\[
= \frac{N - 1}{n(n-1)} \left( \begin{array}{c}
\frac{1}{N_{1}} (n_{n_{01}} - n_{0} n_{1}) \\
\frac{1}{N_{2}} (n_{n_{02}} - n_{0} n_{2})
\end{array} \right) + \frac{N - 1}{n(n-1)} \left( \begin{array}{c}
n_{n_{01}} - n_{n_{02}} \end{array} \right) \left\{ \begin{array}{c}1 \\
0
\end{array} \right\}.
\]

Consequently,
\[
\hat{r}_{0} = \frac{n_{0}}{n} - \frac{N - 1}{n(n-1)} \left( \frac{n_{n_{01}} - n_{0} n_{1}}{N_{1}} - \frac{n_{n_{02}} - n_{0} n_{2}}{N_{2}} \right) \left( \begin{array}{c}
n_{1} \\\nn_{2}
\end{array} \right) - \frac{N - 1}{n(n-1)} \frac{n_{n_{01}} - n_{n_{02}}}{N_{3}} (1 \left( \begin{array}{c}1 \\
n_{1} \\\nn_{2}
\end{array} \right)
\]
\[
+ \frac{N - 1}{n(n-1)} \left( \frac{n_{n_{01}} - n_{0} n_{1}}{N_{1}} - \frac{n_{n_{02}} - n_{0} n_{2}}{N_{2}} \right) \left( \begin{array}{c}
N_{1} \\\nN_{2}
\end{array} \right) + \frac{N - 1}{n(n-1)} \frac{n_{n_{01}} - n_{n_{02}}}{N_{3}} (1 \left( \begin{array}{c}N_{1} \\
N_{2}
\end{array} \right)
\]

After algebraic simplification, we have
\[
\hat{r}_{0} = \frac{n_{0}}{n} - \frac{N - 1}{n(n-1)} \sum_{k=1}^{3} \left( \frac{n_{k}}{N_{k}} - \frac{n}{N} \right) \left( n_{0k} - \frac{n_{k}}{n} \frac{n_{0}}{n} \right) \right\}.
\] (7.17)

This is a “post-stratified” estimator. To obtain its variance, we first evaluate its \(\hat{\rho}_{0,\bullet}^{2}\)
\[
\hat{\rho}_{0,\bullet}^{2} = \frac{\sigma_{0}^{2}}{\sigma_{0}^{2}} / \\sigma_{\theta}^{2} = \sigma_{\theta}^{2}
\]
\[
= \frac{1}{n(n-1)} \frac{1}{n_{0} (n - n_{0}) N_{1} N_{2} N_{3}}
\times \left( n_{n_{01}} - n_{0} n_{1} \right) \left( n_{n_{02}} - n_{0} n_{2} \right) \left( \begin{array}{cc}
N_{2} (N - N_{2}) & N_{1} N_{2} \\
N_{1} N_{2} & N_{1} (N - N_{1})
\end{array} \right) \left( \begin{array}{c}
n_{n_{01}} - n_{0} n_{1} \\
n_{n_{02}} - n_{0} n_{2}
\end{array} \right)
\]

147
\[
\hat{\rho}_0^2 = \frac{1}{n(n-1)} \frac{1}{n(n-n_0)} N_1 N_2 N_3 \frac{N-1}{N_1 N_2 N_3} \times \left\{ \frac{N_1 N_2 (nn_{01} - n_0 n_1)^2}{+2N N_2 (nn_{01} - n_0 n_1) (nn_{02} - n_0 n_2)} + \frac{N_1 N_3 (nn_{02} - n_0 n_2)^2}{+N_1 N_3 (nn_{02} - n_0 n_2)^2 + N_2 N_3 (nn_{01} - n_0 n_1)^2} \right\}
\]

\[
= \frac{1}{n(n-1)} \frac{1}{n(n-n_0)} N_1 N_2 N_3 \frac{N-1}{N_1 N_2 N_3} \times \left\{ \frac{N_1 N_2 (nn_{01} - n_0 n_1 + nn_{02} - n_0 n_2)^2}{+N_1 N_3 (nn_{02} - n_0 n_2)^2 + N_2 N_3 (nn_{01} - n_0 n_1)^2} \right\}
\]

\[
= \left( \frac{N-1}{n-1} \right) \left( \frac{1}{nn_0 (n-n_0)} \right) \sum_{k=1}^{3} \frac{1}{N_k} (nn_{0k} - n_0 n_k)^2.
\]

Therefore, its variance is

\[
\text{var}(\hat{\rho}_0) = \left( \frac{1-f}{n} \right) n_0 (n-n_0) - \left( \frac{1-f}{n} \right) \frac{N-1}{n^2 (n-1)^2} \sum_{k=1}^{3} \frac{1}{N_k} (nn_{0k} - n_0 n_k)^2 \quad (7.18)
\]

Results (7.17) and (7.18) can be readily extended to scenarios where \( p > 3 \) categories are present:

\[
\hat{\rho}_0 = \frac{n_0}{n} - \frac{N-1}{n(n-1)} \sum_{k=1}^{3} \left\{ \frac{n_k}{N_k} - \frac{n}{N} \right\} \left( nn_{0k} - n_0 n_k \right) \quad (7.19)
\]

and

\[
\text{var}(\hat{\rho}_0) = \left( \frac{1-f}{n} \right) \left( \frac{n_0 (n-n_0)}{n(n-1)} - \frac{N-1}{n^2 (n-1)^2} \right) \sum_{k=1}^{3} \frac{1}{N_k} (nn_{0k} - n_0 n_k)^2. \quad (7.20)
\]

Further, if the sample estimator of \( \Sigma_0 \) is used, we have

\[
\hat{\Sigma}_0^{-1} = \frac{n-1}{n n_1 n_2 n_3} \begin{pmatrix} n_2 (n-n_2) & n_1 n_2 \\ n_1 n_2 & n_1 (n-n_1) \end{pmatrix}, \quad \text{and} \quad \hat{\sigma}_{\omega} = \frac{1}{n(n-1)} \left( nn_{01} - n_0 n_1 \right) \text{, and then}
\]

\[
\hat{\mathbf{b}}_0 = \begin{pmatrix} \frac{1}{n_1} \left( n_{01} - n \right) \\ \frac{1}{n_2} \left( n_{02} - n \right) \end{pmatrix} - \frac{1}{n_1} \left( n_{03} - n \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (7.21)
\]
\[\hat{r}_0 = \bar{r}_0 - \beta_*^\prime \overline{Y}^{(n)} + \beta_*^\prime \mu^{(c)}\]

\[= \frac{n_0}{n} - \frac{1}{n} \left( \frac{1}{n_1} \left( n_{01} - \frac{n_{11}}{n} \right) + \frac{1}{n_2} \left( n_{02} - \frac{n_{12}}{n} \right) \right) \left( \frac{n_{11}}{n_1} \right) + \frac{1}{n} \left( \frac{1}{n_3} \left( n_{03} - \frac{n_{13}}{n} \right) \right) \left( \frac{n_{13}}{n_3} \right) \]

\[= \frac{n_0}{n} \left( \frac{1}{n_1} \left( n_{01} - \frac{n_{11}}{n} \right) + \frac{1}{n_2} \left( n_{02} - \frac{n_{12}}{n} \right) \right) \left( \frac{n_{11}}{n_1} \right) + \frac{1}{n} \left( \frac{1}{n_3} \left( n_{03} - \frac{n_{13}}{n} \right) \right) \left( \frac{n_{13}}{n_3} \right) \]

After algebraic simplification, we have

\[\hat{r}_0 = \frac{n_0}{n} - \frac{1}{n} \sum_{k=0}^{3} \left( \frac{n_k}{n} \right) \left( \frac{n_{0k} - \frac{n_k}{n}}{n_0} \right) \left( n_{0k} - \frac{n_k}{n} \right). \quad (7.22)\]

To obtain its variance, we first calculate \(\hat{\rho}_0^2\).

\[\hat{\rho}_0^2 = \frac{\sigma_0^2}{\sigma_{\hat{r}_0}^2} \]

\[= \frac{1}{n_1 n_2 n_3} \left( \frac{n_{01} - n_0 n_1}{n-n_0} \right) \left( \frac{n_{02} - n_0 n_2}{n-n_0} \right) \left( \frac{n_{03} - n_0 n_3}{n-n_0} \right) \]

This can be simplified as follows.

\[\hat{\rho}_0^2 = \frac{1}{n_0 \left( n-n_0 \right)} \frac{1}{n_1 n_2 n_3} \left( n_{01} - n_0 n_1 \right) + \frac{1}{n_0 \left( n-n_0 \right)} \frac{1}{n_2 n_3} \left( n_{02} - n_0 n_2 \right) + \frac{1}{n_0 \left( n-n_0 \right)} \frac{1}{n_1 n_3} \left( n_{03} - n_0 n_3 \right) \]

\[= \frac{1}{n_0 \left( n-n_0 \right)} \left( n_{01} - n_0 n_1 \right) + \frac{1}{n_2 n_3} \left( n_{02} - n_0 n_2 \right) + \frac{1}{n_1 n_3} \left( n_{03} - n_0 n_3 \right) \]
\[
\hat{\rho}_0^2 = \frac{1}{n_0(n-n_0)} \frac{1}{n_1n_2n_3} \left\{ n \left( n_3n_{01} - n_2n_{03} \right) \left( n_{01} - n_0n_1 \right) + n_1 \left( n_3n_{02} - n_2n_{03} \right) \left( n_{02} - n_0n_2 \right) \right\}
\]
\[
= \frac{1}{n_0(n-n_0)} \frac{1}{n_1n_2n_3} \left\{ n^2n_3n_{01}^2 + n_1n_2n_{02}n_{03}^2 + n_1n_2n_{03}^2 - n_2n_3n_{02}^2 \right\}
\]
\[
= \frac{1}{n_0(n-n_0)} \left\{ \frac{n}{n_3n_{01}^2} + \frac{n}{n_2n_{02}} + \frac{n}{n_3n_{03}^2} - \frac{n_2n_3n_{02}^2}{n} \right\}
\]
\[
= \frac{1}{n_0(n-n_0)} \left\{ \sum_{k=1}^{3} \frac{n^2}{n_k} - n^2 \right\}
\]

Therefore, the variance of estimator (7.22) is

\[
\text{var}(\hat{\rho}_0) = \frac{1-f}{n(n-1)} \left( n_0 - \sum_{k=1}^{3} \frac{n^2_{0k}}{n_k} \right). \tag{7.23}
\]

Analogously, these can be extend to situations where the subjects are classified into more than three groups (i.e., \( p > 3 \)).

\[
\hat{\rho} = \frac{n}{n} - \frac{1}{n} \sum_{k=0}^{p} \frac{N_k}{n_k} \left( \frac{n_k}{N_k} - \frac{n}{N} \right) \left( n_{0k} - \frac{n_k}{n} \right). \tag{7.24}
\]

and its variance is

\[
\text{var}(\hat{\rho}) = \frac{1-f}{n(n-1)} \left( n_0 - \sum_{k=1}^{p} \frac{n^2_{0k}}{n_k} \right). \tag{7.25}
\]

### 7.6. Derivations of results (7.3) and (7.4) in Section 7.2

By applying results derived in Section 5.4 of Chapter 5, we have

\[
\hat{\rho}_0 = \bar{y} - \beta_* \left( \bar{y}^{(i)} - \mu^{(i)} \right), \tag{7.26}
\]

where \( \beta_* = \Sigma^{-1} \sigma_* \), \( \mu_* = \left( \frac{N_1}{N} \quad \frac{N_2}{N} \right)' \) and \( \bar{y}^{(i)} = \left( \bar{y}^{(i)} \quad \bar{y}^{(2)} \right)' \).

Since \( N_1 \) and \( N_2 \) are known, we have

\[150\]
\[ \Sigma_\epsilon = \frac{N_1N_2}{N(N-1)} \begin{pmatrix} N_2^{-1}(N-N_1) & -1 \\ -1 & N_1^{-1}(N-N_2) \end{pmatrix}, \]

and

\[ \Sigma^{-1}_\epsilon = \frac{N-1}{N_1N_2N_3} \begin{pmatrix} N_2(N-N_2) & N_1N_2 \\ N_1N_2 & N_1(N-N_1) \end{pmatrix}. \]

It is shown that \( \hat{\sigma}_0^2 = \frac{d_0(n-d_0)}{n(n-1)} \) and \( \hat{\sigma}_0 = \frac{1}{n(n-1)} \left( nn_{01} - n_0n_1 \right) \). Therefore,

\[ \hat{\beta}_0 = \frac{1}{n(n-1)} \frac{N-1}{N_1N_2N_3} \begin{pmatrix} N_2(N-N_2) & N_1N_2 \\ N_1N_2 & N_1(N-N_1) \end{pmatrix} \begin{pmatrix} nn_{01} - n_0n_1 \\ nn_{02} - n_0n_2 \end{pmatrix} \]

\[ = \frac{N-1}{n(n-1)} \frac{1}{N_1N_2N_3} \begin{pmatrix} N_2(N-N_2)(nn_{01} - n_0n_1) + N_1N_2(nn_{02} - n_0n_2) \\ N_1N_2(nn_{01} - n_0n_1) + N_1(N-N_1)(nn_{02} - n_0n_2) \end{pmatrix} \]

\[ = \frac{N-1}{n(n-1)} \frac{1}{N_1} \begin{pmatrix} \frac{1}{N_1}(nn_{01} - n_0n_1) \\ \frac{1}{N_2}(nn_{02} - n_0n_2) \end{pmatrix} + \frac{N-1}{n(n-1)} \frac{1}{N_3} \begin{pmatrix} n_3n_0 - nn_{02} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \]

\[ \hat{\epsilon}_0 = \bar{r}_0 - \hat{\beta}_0 \bar{X}^{(s)} + \hat{\beta}_0 \bar{\mu}^{(s)} \]

\[ = \frac{n_0}{n} - \frac{N-1}{n(n-1)} \begin{pmatrix} nn_{01} - n_0n_1 \\ nn_{02} - n_0n_2 \end{pmatrix} \begin{pmatrix} \frac{n_1}{n} \\ \frac{n_2}{n} \end{pmatrix} - \frac{N-1}{n(n-1)} \frac{1}{N_3} \begin{pmatrix} n_3n_0 - nn_{02} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \]

\[ + \frac{N-1}{n(n-1)} \begin{pmatrix} nn_{01} - n_0n_1 \\ nn_{02} - n_0n_2 \end{pmatrix} \begin{pmatrix} \frac{N_1}{N} \\ \frac{N_2}{N} \end{pmatrix} + \frac{N-1}{n(n-1)} \frac{n_3n_0 - nn_{02}}{N_3} \begin{pmatrix} N_1 \\ N_2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \]

After algebraic simplification, we have

\[ \hat{\epsilon}_0 = \frac{n_0}{n} - \frac{N-1}{n(n-1)} \sum_{k=1}^{3} \left( \frac{n_k}{N} - \frac{n}{N} \right) \left( n_{0k} - \frac{n}{n} \right). \] (7.27)

This is a “post-stratified” estimator.
Further, if the sample estimator of $\Sigma_*$ is used, we have

$$\hat{\Sigma}_*^{-1} = \frac{n-1}{nn_3n_2n_1} \left( \begin{array}{ccc} n_2(n-n_2) & n_1n_2 & n_1(n-n_1) \\ n_2 & n_1(n-n_1) & n_1(n-n_1) \\ n_2 & n_1(n-n_1) & n_1(n-n_1) \end{array} \right), \quad \text{and} \quad \hat{\sigma}_0 = \frac{1}{n(n-1)} \left( \begin{array}{c} mn_{o1} - n_0n_1 \\ mn_{o2} - n_0n_2 \end{array} \right)$$

$$\hat{\beta}_{o*} = \frac{1}{nn_3n_2n_1} \left( \begin{array}{ccc} n_2(n-n_2) & n_2n_1 & n_2(n-n_2) \\ n_2 & n_1(n-n_1) & n_1(n-n_1) \\ n_2 & n_1(n-n_1) & n_1(n-n_1) \end{array} \right) \left( \begin{array}{c} mn_{o1} - n_0n_1 \\ mn_{o2} - n_0n_2 \end{array} \right)$$

$$= \frac{1}{nn_3n_2n_1} \left( \begin{array}{c} n_2(n-n_2)(mn_{o1} - n_0n_1) + n_2n_1(mn_{o2} - n_0n_2) \\ n_2(n-n_2)(mn_{o1} - n_0n_1) + n_2n_1(mn_{o2} - n_0n_2) \\ n_2(n-n_2)(mn_{o1} - n_0n_1) + n_2n_1(mn_{o2} - n_0n_2) \end{array} \right)$$

$$= \frac{1}{nn_3n_2n_1} \left( \begin{array}{c} n_1(n-n_1)(mn_{o1} - n_0n_1 + mn_{o2} - n_0n_2) + n_3(mn_{o1} - n_0n_1) \\ n_1(n-n_1)(mn_{o1} - n_0n_1 + mn_{o2} - n_0n_2) + n_3(mn_{o1} - n_0n_1) \\ n_1(n-n_1)(mn_{o1} - n_0n_1 + mn_{o2} - n_0n_2) + n_3(mn_{o1} - n_0n_1) \end{array} \right)$$

$$= \left( \begin{array}{c} \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) \\ \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) \end{array} \right) - \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right) \left( \begin{array}{c} 1 \\ 1 \end{array} \right)$$

$$\hat{r}_0 = \bar{r}_0 - \beta'_{o*} \bar{Y}^{(*)} + \beta'_{o*} \mu^{(*)}$$

$$= \frac{n_0}{n} - \left( \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) \right) \left( \begin{array}{c} n_1 \\ n_2 \end{array} \right) + \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right) \left( \begin{array}{c} 1 \\ 1 \end{array} \right)$$

$$+ \left( \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) \right) \left( \begin{array}{c} n_{o1} \ N_1 \\ n_{o2} \ N_2 \end{array} \right) - \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right) \left( \begin{array}{c} n_{o1} \\ n_{o2} \end{array} \right)$$

$$= \frac{n_0}{n} - \left( \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) \right) \left( \begin{array}{c} n_{o1} \ N_1 \\ n_{o2} \ N_2 \end{array} \right) - \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right) \left( \begin{array}{c} n_{o1} \\ n_{o2} \end{array} \right)$$

$$+ \left( \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) \right) \left( \begin{array}{c} n_{o1} \ N_1 \\ n_{o2} \ N_2 \end{array} \right) - \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right) \left( \begin{array}{c} n_{o1} \\ n_{o2} \end{array} \right)$$

$$= \frac{n_0}{n} - \left( \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) + \left( n_{o3} - \frac{n_3}{n}n_0 \right) + \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right) \right)$$

$$+ \left( \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) + \left( n_{o3} - \frac{n_3}{n}n_0 \right) + \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right) \right)$$

$$+ \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) + \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right)$$

$$= \frac{n_0}{n} - \left( \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) + \left( n_{o3} - \frac{n_3}{n}n_0 \right) + \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right) \right)$$

$$+ \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) + \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right)$$

$$+ \frac{1}{n_1} \left(n_{o1} - \frac{n_1}{n}n_0 \right) + \frac{1}{n_2} \left(n_{o2} - \frac{n_2}{n}n_0 \right) + \frac{1}{n_3} \left(n_{o3} - \frac{n_3}{n}n_0 \right)$$
\[ \hat{r}_0 = \frac{n_0}{n} - \frac{1}{n} \sum_{k=0}^{3} \left( n_{ok} - \frac{n_k}{n} n_0 \right) + \frac{1}{N} \sum_{k=0}^{3} \left( \frac{N_k}{n_k} \left( n_{ok} - \frac{n_k}{n} n_0 \right) \right) \]

\[ = \frac{n_0}{n} - \frac{1}{n} \sum_{k=0}^{3} \frac{N_k}{n_k} \left( \frac{n_k}{N_k} - \frac{n}{N} \right) \left( n_{ok} - \frac{n_k}{n} n_0 \right) \]

After algebraic simplification, we have

\[ \hat{r}_0 = \frac{n_0}{n} - \frac{1}{n} \sum_{k=0}^{3} \frac{N_k}{n_k} \left( \frac{n_k}{N_k} - \frac{n}{N} \right) \left( n_{ok} - \frac{n_k}{n} n_0 \right). \]
CHAPTER 8

DIRECT RATE STANDARDIZATION USING
RANDOM PERMUTATION MODELS

8.1. Motivations

Results derived in Chapters 4 and 5 can be also used for “rate standardization” - a common procedure in Epidemiology and vital statistics. In this chapter, we will show that direct rate standardization is a special case of the estimators derived in Chapter 5 when rates are assumed to be equal across populations. We may view direct rate standardization as a prediction problem, where one is trying to predict the population death rates of the pooled population based on the two “samples”. Both populations, observed either completely or partially, can be considered as two parts resulting from a random partition of the pooled population. Therefore, the previously described estimation strategies can be applied with minimal modification.

Without losing generality, we use gender-adjusted death rate as an example. Results can be extended to situation where auxiliary variable is multinomial such as age- or age and gender specific categories. We will consider a simple case of internal direct standardization when two populations under comparison are observed completely.

8.2. Rate estimation with internal direct adjustment

A common approach to rate standardization is to “pool” the populations under study, and use the “combined population structure” for adjustment. For example, if two populations A and B are under comparison, let us denote their sizes and gender ratio (\% 

154
of males) are \( N_A, g_A \) and \( N_B, g_B \), respectively. Then the pooled population \( AB \) will have size \( N = N_A + N_B \) and gender ratio \( g_{AB} = \left( N_A g_A + N_B g_B \right) / \left( N_A + N_B \right) \). The rate standardization can be made using \( g_{AB} \). This method was first proposed by Neison (1844), who suggested comparing the mortality of two communities by having the population of one community “actually transferred” to the other community and subjecting it to “exactly the same rate of mortality as that prevailed” in other community (Neison 1844).

Let us consider a simple case. Let \( y_j^{(0)} \) represent survival status of a subject \((y = 1 \text{ for death, and } 0 \text{ for alive})\), and \( y_j^{(1)} \) represent a subject’s gender \((1 \text{ for male and } 0 \text{ for female})\). Suppose that we are interested in estimating and comparing death rates of population A and B \((i.e., \bar{y}_A^{(0)} \text{ and } \bar{y}_B^{(0)})\). Further, suppose that population A consists of \( N_{Am} \) males and \( N_{Af} \) females, and \( N_{Am} + N_{Af} = N_A \); population B consists of \( N_{Bm} \) males and \( N_{Bf} \) females, and \( N_{Bm} + N_{Bf} = N_B \). The pooled population \( AB \) thus consists of \( N_m = N_{Am} + N_{Bm} \) males and \( N_f = N_{Af} + N_{Bf} \), and \( N_m + N_f = N \) subjects. Under Neison’s approach, populations A and B are considered as random partitions of the pooled population \( AB \). Based on parts A and B, we will obtain a set of two predicted total numbers of deaths (or death rates), i.e., \( \hat{T}_A^{(0)} \text{ and } \hat{T}_B^{(0)} \) as well as \( \hat{r}_A \text{ and } \hat{r}_B \). The variance-covariance matrices of \( \hat{T}_A^{(0)} \text{ and } \hat{T}_B^{(0)} \) \((or \hat{r}_A \text{ and } \hat{r}_B)\) can be also obtained. Further, we will be able to estimate difference between \( \hat{r}_A \) and \( \hat{r}_B \) and the rate ratio of \( \hat{r}_A \) to \( \hat{r}_B \). Since the predicted total numbers are predictors of the “same” population, they are comparable.
Suppose that two populations are completely observed, and the two observed populations are viewed as a realization of random partition of the pooled population into two parts that have the sizes $N_A$ and $N_B$. Let $\left( Y^{(0)}_r \mid Y^{(1)}_r \right)'$ represent a random permutation of the vector of response and auxiliary variables of the pooled population, $z = \left( \begin{array}{c} Y^{(0)}_A \\ Y^{(1)}_B \\ Y^{(1)}_A \\ Y^{(1)}_B \end{array} \right)'$. Using the method outlined in Chapter 5, we transform $Y^{(1)}$ to $Y^{(1)*} = \left( I_y - \frac{1}{N} J_N \right) Y^{(1)}$. We then partition $\left( Y^{(0)}_r \mid Y^{(1)*}_r \right)'$ into parts A and B that have the same dimension as populations A and B,

$$Z = \begin{pmatrix} Y^{(0)*}_A & Y^{(1)*}_A \\ Y^{(0)*}_B & Y^{(1)*}_B \end{pmatrix}'.$$

(8.1)

The total number of deaths in the combined population is defined as

$$T^{(0)} = \left( I_A' \mid I_B' \right) \begin{pmatrix} Z_A \\ Z_B \end{pmatrix},$$

where $I_A = (1 \ 0)' \otimes 1_{N_A}$, $I_B = (1 \ 0)' \otimes 1_{N_B}$, $Z_A = \left( Y^{(0)*}_A \ Y^{(1)*}_A \right)'$ and $Z_B = \left( Y^{(0)*}_B \ Y^{(1)*}_B \right)'$.

Let assume that we are interested in deriving estimator of the overall total number of deaths $T^{(0)}$ based on populations A and B, i.e., $\hat{T}^{(0)}_A$ and $\hat{T}^{(0)}_B$, respectively. They can be defined using (8.1) as,

$$\begin{pmatrix} \hat{T}^{(0)}_A \\ \hat{T}^{(0)}_B \end{pmatrix} = \begin{pmatrix} I_A' + w_A' \\ 0 \\ 0 \\ I_B' + w_B' \end{pmatrix} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix}.$$

(8.2)
The prediction error is defined as,
\[
\begin{pmatrix} \hat{T}_A^{(0)} - T^{(0)} \\ \hat{T}_B^{(0)} - T^{(0)} \end{pmatrix} = \begin{pmatrix} I'_A + w'_A \\ 0 \\ 0 \end{pmatrix} \frac{Z_A}{Z_B} - \begin{pmatrix} I'_A \\ I'_B \end{pmatrix} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} = \begin{pmatrix} w'_A - I'_B \\ -I'_A \end{pmatrix} \frac{Z_A}{Z_B}.
\]

The unbiasedness of \( \hat{T}_A^{(0)} \) and \( \hat{T}_B^{(0)} \) for \( T^{(0)} \) implies that,
\[
E \left( \begin{pmatrix} \hat{T}_A^{(0)} - T^{(0)} \\ \hat{T}_B^{(0)} - T^{(0)} \end{pmatrix} \right) = E \begin{pmatrix} w'_AZ_A - I'_BZ_B \\ w'_BZ_B - I'_AZ_A \end{pmatrix} = 0.
\]

Since \( E(Z_A) = (I_2 \otimes 1_{N_A})\left( \mu^{(0)} 0 \right)^\prime \) and \( E(Z_B) = (I_2 \otimes 1_{N_B})\left( \mu^{(0)} 0 \right)^\prime \), this constraint is equivalent to
\[
\begin{pmatrix} (1'_{N_A} 0)w_A - N_B \\ (1'_{N_B} 0)w_B - N_A \end{pmatrix} = \begin{pmatrix} (1'_{N_A} 0) \\ 0 \end{pmatrix} = \begin{pmatrix} 1'_{N_B} 0 \end{pmatrix} = 0,
\]
where \( w = (w'_A w'_B)' \).

From Chapter 2, we have \( \text{cov} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} = \begin{pmatrix} V_A & V_{A,B} \\ V_{B,A} & V_B \end{pmatrix} \), where \( V_A = \Sigma \otimes P_{N_A,N} \), \( V_B = \Sigma \otimes P_{N_B,N} \) and \( V_{A,B} = V'_{B,A} = \Sigma \otimes \left( -\frac{1}{N} J_{N_A \times N_B} \right) \). The variance-covariance matrix of \( \begin{pmatrix} \hat{T}_A^{(0)} \\ \hat{T}_B^{(0)} \end{pmatrix} \) can be represented as
\[
\text{cov} \begin{pmatrix} \hat{T}_A^{(0)} \\ \hat{T}_B^{(0)} \end{pmatrix} = \begin{pmatrix} w'_AV_Aw_A & w'_AV_{A,B}w_B \\ w'_BV_{B,A}w_A & w'_BW_B \end{pmatrix} - 2\begin{pmatrix} I'_BV_{B,A}w_A \\ I'_AV_{A,B}w_B \end{pmatrix} + \begin{pmatrix} I'_BV_{B,A}I_A \\ I'_AV_{A,B}I_B \end{pmatrix}.
\]

Furthermore, we can show that
\[
\text{trace} \left\{ \text{cov} \begin{pmatrix} \hat{T}_A^{(0)} \\ \hat{T}_B^{(0)} \end{pmatrix} \right\} = w' \begin{pmatrix} V_A & 0 \\ 0 & V_B \end{pmatrix} w - 2(I'_B I'_A) \begin{pmatrix} V_{BA} & 0 \\ 0 & V_{AB} \end{pmatrix} w + (I'_A V_A I_A + I'_B V_B I_B).
\]

(8.4)
Following the methods of Chapters 4 and 5, we derive the estimators \( \hat{T}_A^{(0)} \) and 
\( \hat{T}_B^{(0)} \) by minimizing the trace of \( \text{cov}(\hat{T}_A^{(0)}, T_B^{(0)}) \) subject to (8.3), which is equivalent to minimizing the following optimization function

\[
\Phi(w) = w^T \begin{pmatrix} V_A & 0 \\ 0 & V_B \end{pmatrix} w - 2 \begin{pmatrix} I_B' & I_A' \end{pmatrix} \begin{pmatrix} V_{BA} & 0 \\ 0 & V_{AB} \end{pmatrix} w
\]

\[
+ 2 \lambda \left\{ \begin{pmatrix} 1'_{N_A} & 0 \\ 0 & 1'_{N_B} \end{pmatrix} w - \begin{pmatrix} N_B \\ N_A \end{pmatrix} \right\}.
\]

Differentiating \( \Phi(w) \) with respect to \( w \) and \( \lambda \), and setting the derivatives to zero results in the following estimation equations,

\[
\begin{pmatrix}
\begin{pmatrix} V_A & 0 \\ 0 & V_B \end{pmatrix} - 2 \begin{pmatrix} I_B' & I_A' \end{pmatrix} \begin{pmatrix} V_{BA} & 0 \\ 0 & V_{AB} \end{pmatrix} \\
\begin{pmatrix} 1'_{N_A} & 0 \\ 0 & 1'_{N_B} \end{pmatrix} - \begin{pmatrix} N_B \\ N_A \end{pmatrix}
\end{pmatrix}
\begin{pmatrix}
\hat{\lambda} \\
\hat{w}
\end{pmatrix} = \begin{pmatrix}
\begin{pmatrix} V_{A,B} & 0 \\ 0 & V_{B,A} \end{pmatrix} \\
\begin{pmatrix} N_B \\ N_A \end{pmatrix}
\end{pmatrix}.
\]

Solving (8.5) yields the following unique solution,

\[
\hat{w} = \begin{pmatrix} \hat{w}_A \\ \hat{w}_B \end{pmatrix} = -\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1_{N_A} \\ 1_{N_B} \end{pmatrix} + \frac{N}{N_A} \begin{pmatrix} 1 \\ \beta_0 \end{pmatrix} \otimes \begin{pmatrix} 1_{N_A} \\ 1_{N_B} \end{pmatrix} + \frac{N}{N_B} \begin{pmatrix} 1 \\ \beta_0 \end{pmatrix} \otimes \begin{pmatrix} 1_{N_A} \\ 1_{N_B} \end{pmatrix}.
\]

where \( \beta_0 = \sigma_{01}/\sigma_1^2 \). Using (8.6) and (8.2), it is convenient to show that

\[
\begin{pmatrix}
\hat{T}_A^{(0)} \\
\hat{T}_B^{(0)}
\end{pmatrix} = \frac{N}{N} \begin{pmatrix} \bar{Y}_A^{(0)} - \beta_0 (\mu^{(i)} - \bar{Y}_A^{(i)}) \\ \bar{Y}_B^{(0)} - \beta_0 (\mu^{(i)} - \bar{Y}_B^{(i)}) \end{pmatrix}.
\]
where \( \mu^{(1)} \) is the population mean of the auxiliary variable of the pooled population,
\( \bar{Y}_A^{(0)} \) and \( \bar{Y}_B^{(1)} \), \( \bar{Y}_A^{(0)} \) and \( \bar{Y}_B^{(1)} \) are the means of the response and auxiliary variables of
subpopulation A and B, respectively. The variance-covariance matrix of \( \left( \hat{T}_A^{(0)}, \hat{T}_B^{(0)} \right) \) is
\[
\text{cov} \left( \begin{array}{c}
\hat{T}_A^{(0)} \\
\hat{T}_B^{(0)}
\end{array} \right) = N \begin{pmatrix}
\frac{N_B}{N_A} & -1 \\
-1 & \frac{N_A}{N_B}
\end{pmatrix} \left( \sigma_0^2 + \beta^2 \sigma_1^2 - 2 \beta_0 \sigma_0 \right). 
\] (8.8)

**Application to gender-adjusted death rates**

By definition,
\( \mu^{(1)} = \frac{N_m}{N} = \frac{N_{mA} + N_{mB}}{N_A + N_B}, \sigma_{01} = \frac{N_m N_f}{N(N-1)} \left( \frac{N_{0m} - N_{0f}}{N_m} \right), \)

\( \sigma_1^2 = \frac{N_m N_f}{N(N-1)}, \sigma_0^2 = \frac{N_0(N - N_0)}{N(N-1)}, \) and thus \( \beta_{01} = \frac{N_{0m} - N_{0f}}{N_m} \). When applied to
estimating gender-adjusted population totals of deaths, (8.7) can be simplified as
\[
\left( \begin{array}{c}
\hat{T}_A^{(0)} \\
\hat{T}_B^{(0)}
\end{array} \right) = N \begin{pmatrix}
\frac{N_{0A} - N_B}{N_A} & \frac{N_{mB} - N_{mA}}{N_B} \\
\frac{N_{0B} - N_B}{N_B} & \frac{N_{mA} - N_{mB}}{N_A}
\end{pmatrix} \left( \frac{N_{0m} - N_{0f}}{N_m} \right). 
\] (8.9)

Its variance-variance matrix is
\[
\text{cov} \left( \begin{array}{c}
\hat{T}_A^{(0)} \\
\hat{T}_B^{(0)}
\end{array} \right) = N \begin{pmatrix}
\frac{N_B}{N_A} & -1 \\
-1 & \frac{N_A}{N_B}
\end{pmatrix} \left( \frac{N_0(N - N_0)}{N(N-1)} - \frac{N_m N_f}{N(N-1)} \right) \left( \frac{N_{0m} - N_{0f}}{N_m} \right). 
\] (8.10)

Similarly, the standardized death rates and the variance-covariances are
\[
\begin{pmatrix}
\hat{r}_A^{(0)} \\
\hat{r}_B^{(0)}
\end{pmatrix} = \frac{\left(\frac{N_{0A}}{N_A} - \frac{N_{0B}}{N_B}\right) \left(\frac{N_{mB}}{N_B} - \frac{N_{mA}}{N_A}\right) \left(\frac{N_{0m} - N_{0f}}{N_m - N_f}\right)}{\left(\frac{N_{0B}}{N_B} - \frac{N_{0A}}{N_A}\right) \left(\frac{N_{mA}}{N_A} - \frac{N_{mB}}{N_B}\right) \left(\frac{N_{0m} - N_{0f}}{N_m - N_f}\right)},
\]
(8.11)
\[
\text{cov}\left(\hat{r}_A^{(0)}, \hat{r}_B^{(0)}\right) = \frac{1}{N} \left[\frac{N_0(N - N_0)}{N(N - 1)} - \frac{N_mN_f}{N(N - 1)} \left(\frac{N_{0m} - N_{0f}}{N_m - N_f}\right)^2\right] \left(\frac{N_B/N_A}{-1} \frac{1}{N_{A}/N_B}\right),
\]
(8.12)
respectively.

Detailed derivations are given in Section 8.6.

8.3. Rate comparison with internal direct adjustment

With the results established in the previous sections, rates can be compared using either their differences or their ratios.

8.3.1. Rate differences

When all units of the subpopulations are observed, one may evaluate the rate difference by applying result (8.11). We define the rate difference between A and B as,
\[
\widehat{DR}_{AB} = \hat{r}_A^{(0)} - \hat{r}_B^{(0)} = \frac{N_{0A}}{N_A} - \frac{N_{0B}}{N_B} + \left(\frac{N_{mA}}{N_A} - \frac{N_{mB}}{N_B}\right) \left(\frac{N_{0m} - N_{0f}}{N_m - N_f}\right).
\]
(8.13)

Applying result (8.12), the variance of \(\widehat{DR}_{AB}\) is evaluated as
\[
\text{var}\left(\widehat{DR}_{AB}\right) = \begin{pmatrix} 1 & -1 \end{pmatrix} \text{cov}\left(\hat{r}_A^{(0)}, \hat{r}_B^{(0)}\right) \begin{pmatrix} 1 \\ -1 \end{pmatrix}
\]
\[
= \frac{N}{N_A N_B} \left(\sigma_0^2 + \beta_0^2 \sigma_1^2 - 2 \beta_0 \sigma_{01}\right)
\]
\[
= \frac{1}{N - 1} \left(\frac{N_0(N - N_0)}{N_A N_B} - \frac{1}{N - 1} \left(\frac{N_mN_f}{N_A N_B}\right) \left(\frac{N_{0m} - N_{0f}}{N_m - N_f}\right)^2\right).
\]
Loosely based on (8.13) being a function of sample means, a test statistic may be formulated based on the above results. One may test if $DR_{AB}$ is equal to zero based on normal approximation. Notice that, when the two populations under comparison are completely observed, $\text{var}(DR_{AB})$ are known constant.

8.3.2. Rate ratio

Similarly, the rate ratio of $\hat{p}_A^{(0)}$ to $\hat{p}_B^{(0)}$ is

$$RR_{AB} = \frac{\hat{p}_A^{(0)}}{\hat{p}_B^{(0)}} = \left( \frac{N_B}{N_A} \right) \left( \frac{N_{0_A}}{N_{0_B}} \right) - \left( \frac{N_{m_B}}{N_{m_A}} \right) \left( \frac{N_{0_m}}{N_{0_f}} \right).$$

(8.14)

Result (8.14) can be also expressed alternatively as,

$$RR_{AB} = \left( \frac{N_B}{N_A} \right) \left\{ \frac{N_{0_A}}{N_{0_B}} \right\} - 1.$$

Its variance can be evaluated using the first-order Taylor linearization of $RR_{AB} = r_A/r_B$, expanding around $r = (r_A, r_B)'$, i.e.,

$$\hat{RR}_{AB} = \hat{R}_{A}^{2} / \hat{R}_{B} = RR_{AB} + \Delta_{r=r}^{'} (\hat{r} - r) + \text{Remainder},$$

where $\Delta_{r} = \frac{\partial}{\partial r} RR_{AB} \bigg|_{r} = \left( \frac{1}{r_B} - \frac{r_A}{r_B^2} \right)$. When the remainder term is ignored, we have an expression for the linearized error,

$$\varepsilon_{L} = \hat{RR}_{AB} - RR_{AB} \approx \Delta_{r=r}^{'} (\hat{r} - r).$$

(8.15)
Since \( E\left[ \Delta_{r|ri}^\prime \right] = \Delta_{r|ri}^\prime E(\mathbf{r} - \mathbf{r}) = 0 \), \( \widehat{RR}_{AB} \) is approximately unbiased for \( RR_{AB} \). The linearized approximate variance is thus

\[
AV\left( \hat{R}_{0,1} \right) \cong \Delta_{r}^\prime \text{cov}(\hat{\mathbf{r}})\Delta_{r}. \tag{8.16}
\]

A sample estimator can be obtained by substituting a consistent estimator of \( \text{cov}(\hat{\mathbf{r}}) \) and replace \( \mathbf{r} \) in \( \Delta_{r} \) by \( \hat{\mathbf{r}} \), with the formula:

\[
\hat{\Delta}_{r} = \left( \frac{1}{\hat{r}_{A}^{2}} - \frac{\hat{r}_{r}}{r_{r}^{2}} \right), \text{ such that}
\]

\[
\hat{V}(\widehat{RR}_{AB}) \cong \left( \frac{1}{\hat{r}_{A}^{2}} - \frac{\hat{r}_{r}}{r_{r}^{2}} \right)\text{cov}(\hat{\mathbf{r}})\left( \frac{1}{\hat{r}_{A}^{2}} - \frac{\hat{r}_{r}}{r_{r}^{2}} \right)^\prime. \tag{8.17}
\]

Equation (8.17) can be further simplified as

\[
\hat{V}(\widehat{RR}_{AB}) \cong \frac{1}{N} \frac{N_{B}}{N_{A}} \left( \frac{N_{0} N_{A}^{2} + N_{A} N_{0A}^{2}}{N_{0B} N_{0A}} \right)^{2} \left\{ \frac{N_{0} (N - N_{0})}{N (N - 1)} - \frac{N_{m} N_{f}}{N (N - 1)} \frac{N_{0m} - N_{0f}}{N_{m} - N_{f}} \right\}.
\]

If we assume that sample sizes are large enough so that asymptotic normality applies, a 95% confidence interval of \( \widehat{RR}_{AB} \) can be constructed using \( \hat{V}(\widehat{RR}_{AB}) \). A test statistic can be formulated to test whether \( \widehat{RR}_{AB} \) is equal to 1, that is, \( \widehat{R}_{A}^{RR} = R_{B} \).

### 8.4. External direct adjustment

External direct adjustment method usually use an “outside” reference population, for example, US Population based on recent census, or the immediately larger population which includes the populations under comparison. For example, if we are interested in estimating and comparing the directly adjusted death rates of Franklin County (A) and Worcester County (B) in Massachusetts, we may choose the Commonwealth of Massachusetts as the reference population.
Suppose we are interested in the adjusted rates for population A and B. Similarly, populations A and B can be considered as two parts resulting from random permutation of the larger reference population (the Commonwealth of Massachusetts). The estimation process and the consequent estimators are invariant to those presented in Section 8.2.

8.5. Discussion

In this Chapter, we have developed a method of internal direct rate standardization based on random permutation models where we assume equal population rates. Unlike those direct rate standardization method commonly used in Epidemiology and vital statistics, this new method addresses the finiteness of the populations involved and yields a variance-covariance matrix of the standardized rates of the populations under study. In conventional approaches (Kahn and Sempo 1989), the covariances between the rates of any two populations (or subclasses) are ignored. With the variance-covariance, it is possible to conduct hypothesis testing based on asymptotic normal approximation. When both populations under comparison are observed completely, it is also possible to use non-parametric method to evaluate the distribution of the two adjusted rates, their differences or ratios.

We limited our discussion to gender-adjusted rates. Extension of the above results to age- or gender- and age-adjusted rates, rate differences and ratios are analogue to cases where an auxiliary variable is multinomial. Applying the results developed in Chapter 5 to internal direct standardization involving multiple auxiliary variables is an interest field of research in the future.
8.6. Derivations of the results in Section 8.2.

As shown in (8.1), the estimators for \( \begin{pmatrix} T_A^{(0)} & T_B^{(0)} \end{pmatrix}' \) is defined as,

\[
\begin{pmatrix} \hat{T}_A^{(0)} \\ \hat{T}_B^{(0)} \end{pmatrix} = \begin{pmatrix} \frac{A_B T}{A_B T} & 0 \\ 0 & \frac{B_B T}{B_B T} \end{pmatrix} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix}.
\]

The corresponding prediction error is,

\[
\begin{pmatrix} \hat{T}_A^{(0)} \\ \hat{T}_B^{(0)} \end{pmatrix} - \begin{pmatrix} T_A^{(0)} \\ T_B^{(0)} \end{pmatrix} = \begin{pmatrix} \frac{A_B T}{A_B T} & 0 \\ 0 & \frac{B_B T}{B_B T} \end{pmatrix} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} - \begin{pmatrix} \frac{A_B T}{A_B T} & 0 \\ 0 & \frac{B_B T}{B_B T} \end{pmatrix} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} = \begin{pmatrix} \frac{A_B T}{A_B T} - \frac{B_B T}{B_B T} \\ -\frac{A_B T}{A_B T} \end{pmatrix} \begin{pmatrix} Z_A \\ Z_B \end{pmatrix}.
\]

The unbiasedness of \( T_A^{(0)} \) and \( T_B^{(0)} \) for \( T^{(0)} \) implies that,

\[
E \left( \frac{T_A^{(0)} - T^{(0)}}{T_B^{(0)} - T^{(0)}} \right) = E \left( \frac{w' A - l' B}{w' B - l' A} \right) = 0.
\]

Since \( E(Z_A) = \left( 1_2 \otimes 1_{N_A} \right) \left( \mu^{(0)} \right)' 0 \) and \( E(Z_B) = \left( 1_2 \otimes 1_{N_B} \right) \left( \mu^{(0)} \right)' 0 \), this constraint is equivalent to

\[
\begin{pmatrix} 1'_{N_A} & 0 \end{pmatrix} A - B \begin{pmatrix} 1'_{N_B} & 0 \end{pmatrix} = (1'_{N_A} & 0 \begin{pmatrix} 0 & (1'_{N_B}) \\ 0 & 1 \end{pmatrix} - (1'_{N_A} & 0 \begin{pmatrix} 0 & 1 \end{pmatrix} = 0.
\]

where \( w = \left( w'_A \quad w'_B \right)' \).

Now, we apply results derived in Section 4.5.2 of Chapter 4 and Chapter 5. We optimize the estimators in terms of the trace of \( \text{cov} \left( \begin{pmatrix} T_A^{(0)} & T_B^{(0)} \end{pmatrix} \right) \). From Chapter 2, we have \( \text{cov} \left( \begin{pmatrix} Z_A \\ Z_B \end{pmatrix} = \begin{pmatrix} V_A & V_{A,B} \\ V_{B,A} & V_B \end{pmatrix} \right) \), where \( V_A = \Sigma \otimes P_{N_A,N} \), \( V_B = \Sigma \otimes P_{N_B,N} \) and
$$\mathbf{V}_{A,B} = \mathbf{V}_{B,A} = \Sigma \otimes \left( -\frac{1}{N} \mathbf{J}_{N_A \times N_B} \right).$$ 
Therefore, \( \text{cov}(\mathbf{T}_A^{(0)}, \mathbf{T}_B^{(0)}) \) can be represented as,

$$\text{cov} \left( \mathbf{T}_A^{(0)}, \mathbf{T}_B^{(0)} \right) = \begin{pmatrix} \mathbf{w}_A' & -\mathbf{l}'_B \\ -\mathbf{l}'_A & \mathbf{w}_B' \end{pmatrix} \begin{pmatrix} \mathbf{V}_A & \mathbf{V}_{A,B} \\ \mathbf{V}_{B,A} & \mathbf{V}_B \end{pmatrix} \begin{pmatrix} \mathbf{w}_A & -\mathbf{l}_A \\ -\mathbf{l}_B & \mathbf{w}_B \end{pmatrix},$$

$$= \left\{ \begin{pmatrix} \mathbf{w}_A' & 0 \\ 0 & \mathbf{w}_B' \end{pmatrix} - \begin{pmatrix} 0 & \mathbf{l}'_B \\ -\mathbf{l}'_A & 0 \end{pmatrix} \right\} \begin{pmatrix} \mathbf{V}_A & \mathbf{V}_{A,B} \\ \mathbf{V}_{B,A} & \mathbf{V}_B \end{pmatrix} \left\{ \begin{pmatrix} \mathbf{w}_A & 0 \\ 0 & \mathbf{w}_B \end{pmatrix} - \begin{pmatrix} 0 & \mathbf{l}_A \\ -\mathbf{l}_B & 0 \end{pmatrix} \right\}$$

$$= \begin{pmatrix} \mathbf{w}_A' & 0 \\ 0 & \mathbf{w}_B' \end{pmatrix} \begin{pmatrix} \mathbf{V}_A & \mathbf{V}_{A,B} \\ \mathbf{V}_{B,A} & \mathbf{V}_B \end{pmatrix} \begin{pmatrix} \mathbf{w}_A & 0 \\ 0 & \mathbf{w}_B \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{l}'_B \\ -\mathbf{l}'_A & 0 \end{pmatrix} \begin{pmatrix} \mathbf{V}_A & \mathbf{V}_{A,B} \\ \mathbf{V}_{B,A} & \mathbf{V}_B \end{pmatrix} \begin{pmatrix} \mathbf{w}_A & 0 \\ 0 & \mathbf{w}_B \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{w}_A' \mathbf{V}_A \mathbf{w}_A \\ \mathbf{w}_B' \mathbf{V}_{B,A} \mathbf{w}_B \end{pmatrix} \begin{pmatrix} \mathbf{w}_A' \mathbf{V}_{A,B} \mathbf{w}_B \\ \mathbf{w}_B' \mathbf{V}_{B,B} \mathbf{w}_B \end{pmatrix} - 2 \begin{pmatrix} \mathbf{l}'_B \mathbf{V}_{B,A} \mathbf{w}_A \\ \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{w}_B \end{pmatrix} \begin{pmatrix} \mathbf{w}_A' \mathbf{V}_{A,B} \mathbf{w}_B \\ \mathbf{w}_B' \mathbf{V}_{B,B} \mathbf{w}_B \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{I}_A \\ \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{I}_A \end{pmatrix} \begin{pmatrix} \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{I}_A \\ \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{I}_A \end{pmatrix}$$

We derive estimators of \( \left( \mathbf{T}_A^{(0)}, \mathbf{T}_B^{(0)} \right) \) by minimizing trace of \( \text{cov}(\mathbf{T}_A^{(0)}, \mathbf{T}_B^{(0)}) \).

The trace of \( \text{cov}(\mathbf{T}_A^{(0)}, \mathbf{T}_B^{(0)}) \) is

$$\text{trace} \left( \text{cov} \left( \mathbf{T}_A^{(0)}, \mathbf{T}_B^{(0)} \right) \right) = \text{trace} \left( \begin{pmatrix} \mathbf{w}_A' \mathbf{V}_A \mathbf{w}_A \\ \mathbf{w}_B' \mathbf{V}_{B,A} \mathbf{w}_A \end{pmatrix} \begin{pmatrix} \mathbf{w}_A' \mathbf{V}_{A,B} \mathbf{w}_B \\ \mathbf{w}_B' \mathbf{V}_{B,B} \mathbf{w}_B \end{pmatrix} \right) - 2 \text{trace} \left( \begin{pmatrix} \mathbf{l}'_B \mathbf{V}_{B,A} \mathbf{w}_A \\ \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{w}_B \end{pmatrix} \begin{pmatrix} \mathbf{w}_A' \mathbf{V}_{A,B} \mathbf{w}_B \\ \mathbf{w}_B' \mathbf{V}_{B,B} \mathbf{w}_B \end{pmatrix} \right)$$

$$+ \text{trace} \left( \begin{pmatrix} \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{I}_A \\ \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{I}_A \end{pmatrix} \begin{pmatrix} \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{I}_A \\ \mathbf{l}'_B \mathbf{V}_{B,B} \mathbf{I}_A \end{pmatrix} \right)$$

$$= \mathbf{w}' \begin{pmatrix} \mathbf{V}_A & 0 \\ 0 & \mathbf{V}_B \end{pmatrix} \mathbf{w} - 2 \left( \mathbf{l}'_B \mathbf{l}'_A \begin{pmatrix} \mathbf{V}_{B,A} & 0 \\ 0 & \mathbf{V}_{A,B} \end{pmatrix} \mathbf{w} \right)$$

$$+ \mathbf{w}' \begin{pmatrix} \mathbf{V}_A & 0 \\ 0 & \mathbf{V}_B \end{pmatrix} \mathbf{w} = 2 \mathbf{w}' \begin{pmatrix} \mathbf{V}_A & 0 \\ 0 & \mathbf{V}_B \end{pmatrix} \mathbf{w}$$

The optimization function is thus defined as

$$\Phi(\mathbf{w}) = \mathbf{w}' \begin{pmatrix} \mathbf{V}_A & 0 \\ 0 & \mathbf{V}_B \end{pmatrix} \mathbf{w} - 2 \left( \mathbf{l}'_B \mathbf{l}'_A \begin{pmatrix} \mathbf{V}_{B,A} & 0 \\ 0 & \mathbf{V}_{A,B} \end{pmatrix} \mathbf{w} \right)$$

$$+ 2\lambda' \begin{pmatrix} \mathbf{1}_{N_A} & 0 \\ 0 & \mathbf{1}_{N_B} \end{pmatrix} \mathbf{w} - \begin{pmatrix} \mathbf{N}_B \\ \mathbf{N}_A \end{pmatrix}.$$
Differentiating $\Phi(w)$ with respect to $w$ and $\lambda$, we have

$$
\frac{\partial}{\partial w} \Phi(w) = 2\begin{pmatrix} V_A & 0 \\ 0 & V_B \end{pmatrix} w - 2\begin{pmatrix} V_{A,B} & 0 \\ 0 & V_{B,A} \end{pmatrix} \begin{pmatrix} I_B \\ I_A \end{pmatrix} + 2 \begin{pmatrix} 1_{N_d} \\ 0 \end{pmatrix} 0 \\
0 & \begin{pmatrix} 1_{N_B} \\ 0 \end{pmatrix} \lambda
$$

$$
\frac{\partial}{\partial \lambda} \Phi(w) = 2\begin{pmatrix} 1'_{N_d} & 0 \\ 0 & 1'_{N_B} \end{pmatrix} w - \begin{pmatrix} N_B \\ N_A \end{pmatrix}
$$

Setting both derivatives to zero results in the following estimating equations

$$
\begin{pmatrix} V_A & 0 \\ 0 & V_B \end{pmatrix} \begin{pmatrix} 1_{N_d} \\ 0 \end{pmatrix} 0 \\
0 & \begin{pmatrix} 1_{N_B} \\ 0 \end{pmatrix} 0
\end{pmatrix} \begin{pmatrix} \hat{w} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} V_{A,B} & 0 \\ 0 & V_{B,A} \end{pmatrix} \begin{pmatrix} I_B \\ I_A \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1_{N_d} \\ 0 \end{pmatrix}. \tag{8.19}
$$

Solving (8.19) yields the following unique solution,

$$
\hat{w} = \left( V_A \begin{pmatrix} V_{A,B} & 0 \\ 0 & V_{B,A} \end{pmatrix} \begin{pmatrix} I_B \\ I_A \end{pmatrix} - \begin{pmatrix} V_A & 0 \\ 0 & V_B \end{pmatrix} \begin{pmatrix} 1_{N_d} \\ 0 \end{pmatrix} 0 \right)^{-1} \begin{pmatrix} 1_{N_d} \\ 0 \end{pmatrix}
$$

\(\hat{\lambda} = \left( \begin{pmatrix} 1'_{N_d} & 0 \\ 0 & 1'_{N_B} \end{pmatrix} \begin{pmatrix} V_A & 0 \\ 0 & V_B \end{pmatrix} \begin{pmatrix} 1_{N_d} \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1_{N_B} \\ 0 \end{pmatrix} \right)^{-1}

\times \left( \begin{pmatrix} 1'_{N_d} & 0 \\ 0 & 1'_{N_B} \end{pmatrix} \begin{pmatrix} V_A & 0 \\ 0 & V_B \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} V_{A,B} & 0 \\ 0 & V_{B,A} \end{pmatrix} \begin{pmatrix} I_B \\ I_A \end{pmatrix} - \begin{pmatrix} N_B \\ N_A \end{pmatrix} \right). \tag{8.19}
$$

After algebraic simplification, we have
\[
\hat{\lambda} = \begin{pmatrix}
(1'_{N_A} \ 0) V^{-1}_A \begin{pmatrix}
1'_{N_A} \\
0
\end{pmatrix} & \begin{pmatrix}
1'_{N_A} \\
0
\end{pmatrix} V^{-1} V_B \begin{pmatrix}
1'_{N_A} \\
0
\end{pmatrix} I_B - N_B \\
(1'_{N_B} \ 0) V^{-1}_B \begin{pmatrix}
1'_{N_B} \\
0
\end{pmatrix} & \begin{pmatrix}
1'_{N_B} \\
0
\end{pmatrix} V^{-1} V_B \begin{pmatrix}
1'_{N_B} \\
0
\end{pmatrix} I_A - N_A
\end{pmatrix}
\]

and

\[
\hat{w} = \begin{pmatrix}
V^{-1}_A V_{A,B}^I_B \\
V^{-1}_B V_{B,A}^I_A
\end{pmatrix} - \begin{pmatrix}
V^{-1}_A \begin{pmatrix}
1'_{N_A} \\
0
\end{pmatrix} & \begin{pmatrix}
1'_{N_A} \\
0
\end{pmatrix} V^{-1} V_B \begin{pmatrix}
1'_{N_A} \\
0
\end{pmatrix} I_B - N_B \\
V^{-1}_B \begin{pmatrix}
1'_{N_B} \\
0
\end{pmatrix} & \begin{pmatrix}
1'_{N_B} \\
0
\end{pmatrix} V^{-1} V_B \begin{pmatrix}
1'_{N_B} \\
0
\end{pmatrix} I_A - N_A
\end{pmatrix}
\]

Since \( V^{-1}_A = \Sigma^{-1} \otimes P_{N_A,N}^{-1} \), \( V_B = \Sigma^{-1} \otimes P_{N_B,N}^{-1} \), \( V_{A,B} = V'_{A,B} = \Sigma \otimes \left( -\frac{1}{N} J_{N_A \times N_B} \right) \), we have

\[
V^{-1}_A V_{A,B}^I_B = I_2 \otimes \left( -\frac{1}{N-N_A} J_{N_A \times N_B} \right), \quad V^{-1}_B V_{B,A}^I_A = I_2 \otimes \left( -\frac{1}{N-N_B} J_{N_B \times N_A} \right),
\]

\[
V^{-1}_B V_{B,A}^I_A = -(1 \ 0)' \otimes 1_{N_B}, \quad V^{-1}_A V_{A,B}^I_B = -(1 \ 0)' \otimes 1_{N_A}.
\]

Consequently,

\[
(1 \ 0) \Sigma^{-1} \begin{pmatrix}
1 \\
0
\end{pmatrix} = \frac{\sigma_2^2}{\sigma_0^2 \sigma_1^2 - \sigma_{01}^2}, \quad \text{and} \quad (1 \ 0) \Sigma^{-1} \begin{pmatrix}
1 \\
0
\end{pmatrix}^{-1} = \frac{\sigma_0^2 \sigma_1^2 - \sigma_{01}^2}{\sigma_1^2}.
\]

Further,

\[
(1'_{N_A} \ 0) V^{-1}_A \begin{pmatrix}
1'_{N_A} \\
0
\end{pmatrix} = \frac{NN_A}{N-N_A} \frac{\sigma_1^2}{\sigma_0^2 \sigma_1^2 - \sigma_{01}^2}, \quad (1'_{N_B} \ 0) V^{-1}_B \begin{pmatrix}
1'_{N_B} \\
0
\end{pmatrix} = \frac{NN_B}{N-N_B} \frac{\sigma_1^2}{\sigma_0^2 \sigma_1^2 - \sigma_{01}^2}.
\]

Therefore,

\[
\hat{w} = \begin{pmatrix}
(1 \otimes 1_{N_A}) + \frac{N}{N_A} \begin{pmatrix}
1 \\
-\beta_{01}
\end{pmatrix} \otimes 1_{N_A} \\
(1 \otimes 1_{N_B}) + \frac{N}{N_B} \begin{pmatrix}
1 \\
-\beta_{01}
\end{pmatrix} \otimes 1_{N_B}
\end{pmatrix} = \begin{pmatrix}
\frac{N}{N_A} \begin{pmatrix}
-\frac{N_B}{N} \\
-\beta_{01}
\end{pmatrix} \otimes 1_{N_A} \\
\frac{N}{N_B} \begin{pmatrix}
-\frac{N_A}{N} \\
-\beta_{01}
\end{pmatrix} \otimes 1_{N_B}
\end{pmatrix},
\]

where \( \beta_{01} = \sigma_{01} \sigma_1^2 \). Consequently,
\[
\hat{w}_A = -\left(1 \otimes 1_{N_A}\right) + \frac{N}{N_A} \left(1 - \beta_{01}\right) \otimes 1_{N_A} = \frac{N}{N_A} \left(-\frac{N_B}{N} \left(-\beta_{01}\right)\right) \otimes 1_{N_A},
\]

\[
\hat{w}_B = -\left(1 \otimes 1_{N_B}\right) + \frac{N}{N_B} \left(1 - \beta_{01}\right) \otimes 1_{N_B} = \frac{N}{N_B} \left(-\frac{N_A}{N} \left(-\beta_{01}\right)\right) \otimes 1_{N_B}.
\]

The estimators for \((T_A^{(0)} \ T_B^{(0)})\) can be represented as

\[
\begin{pmatrix}
\hat{t}_A^{(0)} \\
\hat{t}_B^{(0)}
\end{pmatrix} = \begin{pmatrix}
I_A' + w_A' & 0 \\
0 & I_B' + w_B'
\end{pmatrix} \begin{pmatrix}
Z_A \\
Z_B
\end{pmatrix} = \begin{pmatrix}
\frac{N}{N_A} \left(1 - \beta_{01}\right) \otimes 1_{N_A} \\
\frac{N}{N_B} \left(1 - \beta_{01}\right) \otimes 1_{N_B}
\end{pmatrix} \begin{pmatrix}
Y_A^{(0)} \\
Y_B^{(0)}
\end{pmatrix} = \begin{pmatrix}
\frac{N}{N_A} \left(1 - \beta_{01}\right) \otimes 1_{N_A} \\
\frac{N}{N_B} \left(1 - \beta_{01}\right) \otimes 1_{N_B}
\end{pmatrix} \begin{pmatrix}
Y_A^{(0)} \\
Y_B^{(0)}
\end{pmatrix},
\]

or simply,

\[
\begin{pmatrix}
\hat{t}_A^{(0)} \\
\hat{t}_B^{(0)}
\end{pmatrix} = \begin{pmatrix}
\frac{N}{N_A} \left(1 - \beta_{01}\right) \otimes 1_{N_A} \\
\frac{N}{N_B} \left(1 - \beta_{01}\right) \otimes 1_{N_B}
\end{pmatrix} \begin{pmatrix}
\tilde{Y}_A^{(0)} - \beta_{01} \tilde{Y}_A^{(1)} \\
\tilde{Y}_B^{(0)} - \beta_{01} \tilde{Y}_B^{(1)}
\end{pmatrix},
\]

\[\text{(8.21)}\]

where \(\mu^{(l)}\) is the population mean of the auxiliary variable of the pooled population, \(\tilde{Y}_A^{(0)}\) and \(\tilde{Y}_A^{(1)}\), \(\tilde{Y}_B^{(0)}\) and \(\tilde{Y}_B^{(1)}\) are the means of the response and auxiliary variables of subpopulation A and B, respectively.

The variance-covariance matrix of \((T_A^{(0)} \ T_B^{(0)})\) is evaluated as follows.
\[
\text{cov} \left( \hat{t}^{(0)}_A, \hat{t}^{(0)}_B \right) = \left( \begin{bmatrix} t'_A + w'_A & 0 \\ 0 & t'_B + w'_B \end{bmatrix} \right) \left( \begin{bmatrix} V_A & V_{A,B} \\ V_{B,A} & V_B \end{bmatrix} \right) \left( \begin{bmatrix} t_A + w_A & 0 \\ 0 & t_B + w_B \end{bmatrix} \right) = \left( \begin{bmatrix} (t'_A + w'_A)^T V_A (t_A + w_A) & (t'_A + w'_A)^T V_{A,B} (t_B + w_B) \\ (t'_B + w'_B)^T V_{B,A} (t_A + w_A) & (t'_B + w'_B)^T V_B (t_B + w_B) \end{bmatrix} \right). \]

Since \( t'_A + w'_A = \frac{N}{N_A} ((1 - \beta_{01}) \otimes 1'_{N_A}) \) and \( t'_B + w'_B = \frac{N}{N_B} ((1 - \beta_{01}) \otimes 1'_{N_B}) \), we have

\[
(t'_A + w'_A)^T V_A (t_A + w_A) = \left( \frac{N}{N_A} \right)^2 \left( (1 - \beta_{01}) \otimes 1'_{N_A} \right) \left( \Sigma \otimes \mathbf{P}_{N_A,N} \right) \left( \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \otimes \mathbf{1}_{N_A} \right) = \left( \frac{N}{N_A} \right)^2 \left( (1 - \beta_{01}) \Sigma \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \right) \otimes \left( 1'_{N_A} \mathbf{P}_{N_A,N} \mathbf{1}_{N_A} \right) = \left( \frac{N}{N_A} \right)^2 \left( 1 - \beta_{01} \right) \left( \begin{bmatrix} \sigma^2_0 \\ \sigma^2_1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \right) = N \frac{N_B}{N_A} \left( \sigma^2_0 + \beta_{01} \sigma^2_1 - 2 \beta_{01} \sigma_{01} \right)
\]

\[
(t'_B + w'_B)^T V_B (t_B + w_B) = \left( \frac{N}{N_B} \right)^2 \left( (1 - \beta_{01}) \otimes 1'_{N_B} \right) \left( \Sigma \otimes \mathbf{P}_{N_B,N} \right) \left( \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \otimes \mathbf{1}_{N_B} \right) = \left( \frac{N}{N_B} \right)^2 \left( (1 - \beta_{01}) \Sigma \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \right) \otimes \left( 1'_{N_B} \mathbf{P}_{N_B,N} \mathbf{1}_{N_B} \right) = \left( \frac{N}{N_B} \right)^2 \left( 1 - \beta_{01} \right) \left( \begin{bmatrix} \sigma^2_0 \\ \sigma^2_1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \right) = N \frac{N_A}{N_B} \left( \sigma^2_0 + \beta_{01} \sigma^2_1 - 2 \beta_{01} \sigma_{01} \right)
\]

\[
(t'_A + w'_A)^T V_{A,B} (t_B + w_B) = (t'_B + w'_B)^T V_{B,A} (t_A + w_A) \]

\[
= \left( \frac{N^2}{N_A N_B} \right) \left( \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \otimes 1'_{N_A} \right) \left( \Sigma \otimes \left( -\frac{\mathbf{J}_{N_A \times N_B}}{N} \right) \right) \left( \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \otimes 1'_{N_B} \right) = \left( \frac{N^2}{N_A N_B} \right) \left( 1 - \beta_{01} \right) \Sigma \begin{bmatrix} 1 \\ -\beta_{01} \end{bmatrix} \otimes \left( -\frac{1}{N} 1'_{N_A} \mathbf{J}_{N_A \times N_B} \mathbf{1}_{N_B} \right) = -N \left( \sigma^2_0 + \beta_{01} \sigma^2_1 - 2 \beta_{01} \sigma_{01} \right)
\]

169
Consequently,

$$\text{cov}\left(\hat{T}^{(0)}_A, \hat{T}^{(0)}_B\right) = N \begin{pmatrix} \frac{N_B}{N} & -1 \\ \frac{N_A}{N} & -1 \end{pmatrix} \left(\sigma_0^2 + \beta_{01}\sigma_1^2 - 2\beta_{01}\sigma_{01}\right), \quad (8.22)$$

By definition, $$\mu^{(1)} = \frac{N_m}{N} = \frac{N_m + N_mB}{N_A + N_B}, \quad \sigma_0^2 = \frac{N_0(N - N_0)}{N(N - 1)}, \quad \sigma_1^2 = \frac{N_mN_f}{N(N - 1)},$$

$$\sigma_{01} = \frac{N_mN_f}{N(N - 1)} \left(\frac{N_{0m}}{N_m} - \frac{N_{0f}}{N_f}\right),$$

and thus $$\beta_{01} = \frac{N_{0m}}{N_m} - \frac{N_{0f}}{N_f}.$$ Equation (8.21) can be further simplified as follows,

$$\left(\begin{array}{c} \hat{T}^{(0)}_A \\ \hat{T}^{(0)}_B \end{array}\right) = \left(\begin{array}{c} N \left(\frac{N_{0A}}{N_A} - \beta_{01}\left(\frac{N_{mA} + N_{mB} - N_{mA}}{N_A + N_B}\right)\right) \\ N \left(\frac{N_{0B}}{N_B} - \beta_{01}\left(\frac{N_{mA} + N_{mB} - N_{mB}}{N_A + N_B}\right)\right) \end{array}\right) = \left(\begin{array}{c} N \left(\frac{N_{0A}}{N_A} - \beta_{01}\frac{N_B}{N}\left(\frac{N_{mA} - N_{mA}}{N_A}\right)\right) \\ N \left(\frac{N_{0B}}{N_B} - \beta_{01}\frac{N_A}{N}\left(\frac{N_{mB} - N_{mA}}{N_B}\right)\right) \end{array}\right)

= \left(\begin{array}{c} N \left(\frac{N_{0A}}{N_A} - \frac{N_B}{N}\left(\frac{N_{mB} - N_{mA}}{N_A}\right)\right) \\ N \left(\frac{N_{0B}}{N_B} - \frac{N_A}{N}\left(\frac{N_{mA} - N_{mB}}{N_B}\right)\right) \end{array}\right)

\left(\begin{array}{c} N_{0m} - N_{0f} \\ N_{m} - N_{f} \end{array}\right).$$

It is also shown that

$$\sigma_0^2 + \beta_{01}\sigma_1^2 - 2\beta_{01}\sigma_{01} = \frac{N_0(N - N_0)}{N(N - 1)} - \frac{N_mN_f}{N(N - 1)} \left(\frac{N_{0m}}{N_m} - \frac{N_{0f}}{N_f}\right)^2. \quad (8.23)$$

Therefore,

$$\text{cov}\left(\hat{T}^{(0)}_A, \hat{T}^{(0)}_B\right) = N \begin{pmatrix} \frac{N_B}{N} & -1 \\ \frac{N_A}{N} & -1 \end{pmatrix} \left\{\frac{N_0(N - N_0)}{N(N - 1)} - \frac{N_mN_f}{N(N - 1)} \left(\frac{N_{0m}}{N_m} - \frac{N_{0f}}{N_f}\right)^2\right\}.$$
Through similar derivation, the estimators and the variance-covariance matrix for gender-adjusted rates of populations A and B are shown to be

\[
\begin{align*}
\begin{pmatrix}
\hat{p}_A^{(0)} \\
\hat{p}_B^{(0)}
\end{pmatrix}
&= \frac{N_{0A} - N_B (N_{mB} - N_{mA})}{N - N_B (N_{mA} - N_{mB})} \begin{pmatrix}
\frac{N_{0m} - N_{0f}}{N_m - N_f}
\end{pmatrix} \\
&- \frac{N_{0B} - N_A (N_{mB} - N_{mA})}{N - N_A (N_{mA} - N_{mB})} \begin{pmatrix}
\frac{N_{0m} - N_{0f}}{N_m - N_f}
\end{pmatrix},
\end{align*}
\]

\[
\text{cov}\left(\begin{pmatrix}
\hat{p}_A^{(0)} \\
\hat{p}_B^{(0)}
\end{pmatrix}\right) = \frac{1}{N} \left\{ \frac{N_B}{N_A} \begin{pmatrix}
-1
\end{pmatrix} \begin{pmatrix}
N_0 (N - N_0) \\
N (N - 1) (N_0 - N_0)
\end{pmatrix} - \frac{N_m N_f}{N (N - 1)} \begin{pmatrix}
N_0m - N_0f
\end{pmatrix}^2 \right\},
\]

respectively.
This dissertation proposed alternative methods of deriving estimators using random permutation model (Stanek, Singer and Lencina 2003) for settings of simple random sampling without replacement (SRSWOR) and stratified simple random sampling without replacement (STSRS). We integrated techniques from several areas in survey sampling. We used a finite random permutation as a probability link between samples and the population. This framework does not require assumptions about parametric distribution, and the randomness is attributable only to random sampling (permutation). We represented the joint permutation of response and auxiliary variables with a seemingly unrelated regression setup, and thus the structural relationship between the permuted population and the population was represented explicitly in matrix format. This setup provides a convenient mathematical vehicle for deriving estimators when multiple variables are involved. Such a setup enabled us to derive linear unbiased estimators under design-based framework (Stanek, Singer and Lencina 2003) using estimation techniques that are commonly applied in prediction-based approaches, such as Royall’s general prediction theorem (Royall 1976; Valliant, Dorfman and Royall 2000). When multiple estimators are simultaneous estimated, this approach is capable of retaining the covariance between the estimators, which is a feature particularly useful in rate standardization and small-area estimates. Finally, we used finite estimating equations approach (Binder and Patak 1994) to optimize the
estimators by minimizing either generalized mean squared error or the trace of variance-covariance matrix of the target parameters subject to various constraints.

With this proposed method, we obtained explicit expression for the joint permutation of response and multiple auxiliary variables, and formulae for the population mean vector and their variance-covariance matrix. These results are similar to those commonly seen in literature regarding seemingly unrelated regression (Zellner 1962; Zellner 1963; Revankar 1974; Srivastava and Giles 1987; Gao and Huang 2000; Liu 2000), but our results incorporate finiteness and rely on neither parametric nor superpopulation model assumptions.

We showed that sampling can be represented as a partition of the joint permutation into the sample and the remainder parts, and the vector of population means can be represented as a linear function of the random variables arising from the permutation. We derived linear unbiased minimum variance predictors of population means by predicting the remainder parts based on the sample, which may be called as “design-based predictors”. Using this method, we derived, in Chapter 4, two classes of estimators for population totals when auxiliary information was not available. The two classes were defined by two sets of weights that differed in whether identical weights were applied to each variable. The estimators were optimized by minimizing either the trace of $\text{cov}(\hat{T})$ or M-estimation criteria. Under SRSWOR and unbiasedness constraints, it is shown that the estimators are identical to simple expansion estimators. It is apparent that neither the method of defining estimators nor the optimization criteria had impact on the estimation.
In Chapter 5, we derived estimators adjusted for known auxiliary information through reparameterization. This approach yielded a set of weights that are not sample-dependent, and the estimators are unbiased by definition. In contrast, the usual calibration technique of incorporating auxiliary information leads to sample-dependent weights, and the corresponding estimators do not have an unbiasedness property (Deville and Särndal 1992).

Although the estimators derived in this dissertation have a similar functional form to those derived using design-based regression model (Cochran 1977), model-assisted (Särndal, Swensson and Wretman 1992), model-based (Bolfarine and Zacks 1992; Valliant, Dorfman and Royall 2000) and calibration approaches (Deville and Särndal 1992; Brewer 1999), they depend on neither the superpopulation model nor regression model assumptions. Further, the characteristics of design-based predictors are best illustrated in contrast to usual calibration estimators:

1) Both approaches incorporate known auxiliary information and satisfy constraints for the known auxiliary quantities.

2) Design-based predictors are optimized in terms of its mean squared error or its variance, a direct and natural measure of precision; while calibration estimators are optimized in terms of the distance between adjusted and naïve weights, with no assurance that MSE or variance is minimum.

3) Unbiasedness constraints can be explicitly prescribed in design-based prediction method but not in the usual calibration method.
4) Design-based prediction approach considers weights for auxiliary variables irrelevant to the estimation of parameters, while calibration approach is based on argument that identical weights should be used for response and auxiliary variables.

5) Unlike the usual calibration technique, the design-based prediction approach leads to a set of coefficients (weights) that do not depend on the sample.

6) Another advantage of design-based predictors is that it always has a unique solution when known population variance (or variance-covariance matrix) of auxiliary variable(s) is used for estimation. In contrast, calibration estimator is computed based on sample variance (or variance-covariance matrix) of the auxiliary variable(s), there is no unique solution when sample variance-covariance matrix of the auxiliary variables is singular.

In Chapter 6, we extended the results of Chapter 5 to settings of STSRS. By treating the sampling process in each stratum as independent permutation processes, extension of the results to STSRS is straightforward.

We applied these results to rate estimation in Chapter 7. When the response variable is binomial and the auxiliary variable is categorical, we showed that the derived estimators are post-stratified estimators when sample variance-covariances are used for estimation, and equal to those derived using model-assisted (Särndal, Swensson and Wretman 1992) or calibration (Deville and Särndal 1992) approaches. Through a series Monte Carlo simulations, we illustrated that using known population variance-covariance matrix in estimation has its competitive edge, which always yields a unique solution with equal or smaller MSE to estimators based on sample auxiliary variance-covariances.
We showed, in Chapter 8, that rate standardization based on a random permutation model may account for the finiteness of the involved populations and yields a variance-covariance matrix of the corresponding standardized rates. Unlike the conventional rate standardization methods that assume the rate estimations are independent from each other (Kahn and Sempos 1989), we obtained the covariance of the estimated (standardized) rates under a model that assume equal rates in each population.

In conclusion, we demonstrated in this dissertation a design-based prediction approach of incorporating auxiliary information through reparameterization under random permutation framework. We showed that this method is a useful alternative approach of deriving linear unbiased estimators.

This dissertation research has its limitations. We considered only linear unbiased minimum variance estimators for (stratified) simple random sampling without replacement and use of auxiliary variables at the estimation stage. The usefulness of the design-based prediction approach can be extended with further research on its application to other estimation/prediction problems, such as small-domain or –area estimation, unit non-response adjustment, and estimation of non-linear function. It is interesting to examine whether relaxation of unbiasedness constraints or use of other optimal criteria such as log-determinants or M-estimation will result in improvement or loss of efficiency in estimation. Another challenge to this design-based prediction approach is to develop estimation/prediction method for more complex sampling design, such as multiple stage cluster random sampling or other type of unequal probability sampling.
BIBLIOGRAPHY


