1996

Regression estimation for finite population means in the presence of nonresponse

Anthony Baiching An
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Biostatistics Commons, Economic Theory Commons, and the Mathematics Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/11357

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

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

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

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

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

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

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
Regression estimation for finite population means in the presence of nonresponse

by

Anthony Baiching An

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Statistics
Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge Major Work
Signature was redacted for privacy.

For the Major Department
Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1996
# TABLE OF CONTENTS

1 INTRODUCTION ........................................................................................................... 1

2 LITERATURE REVIEW ................................................................................................. 4
  2.1 Estimation of the Population Mean ................................................................. 4
  2.2 Regression Estimator ....................................................................................... 5
  2.3 Weighting Procedures ...................................................................................... 18
  2.4 Nonresponse ..................................................................................................... 27

3 PRELIMINARIES ........................................................................................................... 44
  3.1 Basic Concepts and Definitions ....................................................................... 44
  3.2 Sequences of Finite Populations ...................................................................... 49
  3.3 Sampling with Unequal Probability ................................................................. 53

4 ESTIMATION OF THE POPULATION MEAN IN THE PRESENCE OF NONRESPONSE .......................................................................................................................... 57
  4.1 Nonresponse and Poisson Sampling ................................................................ 57
  4.2 Infinite Sequences of Populations and Samples ........................................... 61
  4.3 The Horvitz-Thompson Estimator with Nonresponse Adjustment ............. 64
  4.4 Regression Estimator for Nonresponse .......................................................... 88
  4.5 Regression Estimator with Nonresponse Adjustment .................................... 93
  4.6 Variance of Regression Estimator ................................................................... 107

5 REGRESSION WEIGHTING METHODS FOR SIPP DATA .............................................. 118
  5.1 The SIPP Data and the Problem ..................................................................... 118
  5.2 Notation and Simple Estimators ..................................................................... 121
  5.3 Estimation of Response Probabilities .............................................................. 125
  5.4 Regression Weighting Schemes ..................................................................... 130
    5.4.1 Regression Weighting by Individuals .................................................... 130
    5.4.2 Three-phase Weighting Using Cluster Totals ......................................... 138
    5.4.3 Minimum Variance Estimators ............................................................... 141
    5.4.4 Superiority of Three-phase Estimator ................................................... 145
  5.5 Application to the SIPP Data ......................................................................... 151
    5.5.1 Regressions Based on Individuals ......................................................... 158
    5.5.2 Regressions Based on Cluster Totals ....................................................... 161
    5.5.3 Response Probability Adjustment for Regression Weights .................. 167

APPENDIX A: NONNEGATIVE REGRESSION ESTIMATION FOR SAMPLE SURVEY DATA ......................................................................................................................... 170
1 INTRODUCTION

In survey sampling, nonresponse is one source of error in the data analysis. Nonresponse introduces bias into the estimation of population characteristics when the respondents and nonrespondents are different. Nonresponse also causes samples to fail to follow the distributions determined by the original sampling design. It is the primary goal of our research to reduce the nonresponse bias in the estimation of population characteristics.

Regression estimation is recognized as a procedure that can be used to reduce the bias from nonresponse by using auxiliary information. In practice, information on the variables of interest is not available for nonrespondents, but information on auxiliary variables may be available for nonrespondents. Therefore, it is desirable to model the response behavior and incorporate the auxiliary information into the estimation, in order to reduce the bias arising from nonresponse. Intuitively, if the auxiliary variables are correlated with the response behavior, then the regression estimators will be more precise when we use the auxiliary information properly.

Cluster sampling is a common sampling design for large complex surveys. An example is the Survey of Income and Program Participation (SIPP), which will be described in Chapter 5. In a survey like the SIPP, nonresponse among members of clusters occurs. Weighting and imputation strategies are used to overcome the imperfection of the samples due to the nonresponse. Both strategies can be applied using regression procedures and the
auxiliary information. For example, one may use regression to generate regression weights to construct regression estimators, or one may fit a regression model and impute the missing values by the predicted values from the regression model.

We use Poisson sampling to model the nonresponse behavior. Poisson sampling is such that elements are included in the sample based on independent Bernoulli trials. If we assume that there is a response probability determining the response behavior of an element, then we can treat the respondents as a sample created by the Poisson sampling mechanism. Poisson sampling is restrictive in that element response is assumed to be independent.

We are interested in obtaining consistent estimators for the population mean. We will consider regression estimators with and without the adjustment for nonresponse. Under the assumption that respondents in each cluster form Poisson samples, we will investigate the conditions under which regression estimators are design consistent. Since in survey practice we do not know the response probabilities, we need to estimate the response probabilities in order to incorporate them into the estimation. We will investigate the consistency of regression estimators using the inverse of estimated response probabilities as weights.

Regression weighting procedures are applied to the Survey of Income and Program Participate (SIPP). The Census Bureau designed the SIPP to provide improved information on income and participation in government programs. We present procedures to create weights for estimating the characteristics of interest. We will compare alternative estimators.
A nonlinear model for the estimation of response probabilities is described and the estimated response probabilities are incorporated into the construction of the regression weights.

The structure of this dissertation is as follows. In Chapter 2, we review literature on estimation of finite population means, regression estimators, weighting procedures, and nonresponse. In Chapter 3, we give some basic definitions and results on unequal probability sampling. In Chapter 4, we investigate the consistency of regression estimators. Horvitz-Thompson estimators with and without the adjustment for response probabilities. Also the consistency of the regression estimators incorporating the estimated response probabilities will be established. The variance expression for an approximation of the error of the regression estimator will be given in Chapter 4 as well. In Chapter 5, we describe the regression weighting procedures using a multi-phase sampling framework. We apply our estimation procedures to the SIPP data. We also describe a model for the estimation of response probabilities for SIPP data. The estimates from different weighting procedures are compared.

We also include Appendix A as a supplement to the dissertation. Appendix A contains a modification on an iterative weight generation algorithm which generates nonnegative regression weights for survey data. This is an extension of the procedure developed by Huang and Fuller (1978).
2 LITERATURE REVIEW

In Section 2.1, we review some results on estimating the mean of a finite population and on estimating the variance of the mean estimators. In Section 2.2, we review regression estimators. Some weighting issues in survey sampling are discussed in Section 2.3, and estimation procedures for nonresponse are reviewed in Section 2.4.

2.1 Estimation of the Population Mean

Suppose we are given a finite population \( \xi_N \) of \( N \) elements, with a characteristic associated with element \( i \) denoted by \( y_i \). We are interested in estimating the population mean

\[
\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i.
\]  

However, if the information on characteristic \( y \) is not easily obtained or if the population size \( N \) is very large, it is infeasible to do a census to establish the population mean. Therefore, we draw a probability sample \( s \) from \( \xi_N \), where the sample size \( n \) is the number of distinct elements in the sample \( s \). The sampling design determines the probability distribution of the estimators we will use. For example, for a simple random sampling design, the common estimator of the population mean \( \bar{Y} \) is the sample mean (Cochran 1977, Ch. 2)

\[
\bar{y} = n^{-1} \sum_{i=1}^{n} y_i,
\]

and for a stratified simple random sampling, the common estimator of the population mean
is

\[ \bar{Y}_{st} = \sum_{h=1}^{L} W_h \bar{Y}_h, \]

where \( h = 1, 2, \ldots, L \) are strata. \( N = \sum_{h=1}^{L} N_h, \ W_h = N^{-1} N_h, \) \( N_h \) is the total elements number in stratum \( h, \bar{Y}_h \) is the sample mean in stratum \( h \) (Cochran 1977, Ch. 5).

Note that estimators such as (2.2) and (2.3) can be written as

\[ \text{estimator} = \sum_{i \in s} w_{si} y_i \] (2.4)

which is a linear combination of the \( \{y_i, i \in s\} \) and \( \{w_{si} : i \in s\} \) are weights. We will consider linear estimators of the form (2.4).

Rao (1994) studied estimating totals and distribution functions using auxiliary information at the estimation stage. He discussed both probability sampling and model-assisted approaches. He considered the asymptotically efficient calibration estimators, where a calibration estimator is an estimator obtained by revising the basic survey weights to satisfy certain consistency constraints. He gave the general set-up for inference from survey data and provided the variance estimators for calibration estimators under stratified simple random sampling and stratified multistage sampling designs.

### 2.2 Regression Estimator

Regression estimators can be used in the presence of auxiliary information. Suppose we observe \((x_i, y_i)\) for each element \( i \) in a sample from a finite population \( \xi_N \), where
\( \mathbf{x}_i = (1, x_{i2}, \ldots, x_{iq}), \quad i = 1, 2, \ldots, n \) and the first element of \( \mathbf{x}_i \) is always one. For convenience, throughout this paper, we will assign the first \( n \) subscripts to sample elements. We denote the population mean for \( \mathbf{X} \) by \( \bar{\mathbf{X}} \) and assume \( \bar{\mathbf{X}} \) is known. We can use \( \bar{\mathbf{X}} \) to construct a regression estimator. For example, for a simple random sample, a regression estimator of \( \bar{Y} \) is

\[ \hat{\mu}_{\text{OLS}} = \bar{\mathbf{X}} \hat{\beta}_{\text{OLS}}, \quad (2.5) \]

where \( \hat{\beta}_{\text{OLS}} \) is the ordinary least squares estimator of the regression coefficient,

\[ \hat{\beta}_{\text{OLS}} = \left( \sum_{i=1}^{n} x_i' x_i \right)^{-1} \left( \sum_{i=1}^{n} x_i' y \right), \quad (2.6) \]

and \( y = (y_1, \ldots, y_n)' \).

Early applications of regression estimators are Wasson (1937), Cochran (1942), and Jensen (1942). Cochran (1977, Ch 7) proved that when the sample size is large, under a simple random sampling design, the regression estimator is as good as the ratio estimator (for the one regressor situation), and as good as the simple sample mean in that the regression estimator has a variance that is never greater than that of the other estimators. Cochran (1977, p. 198) also proved that the bias of the linear regression estimator in (2.5) is \( O(n^{-1}) \) in simple random sampling. In stratified sampling, Cochran (1977, p. 203) discussed which kind of regression coefficients should be used in the regression estimators. For sampling designs selected with unequal probability, Cassel, Särndal, and Wretman (1977, Ch 7) compared a number of regression estimators. They identified two types: mean-of-the-ratios strategies (\( M \)-strategies) and ratio strategies (\( R \)-strategies). An \( M \)-strategy usually
is composed of probability sampling with inclusion probability proportional to sample unit size, and an estimator that combines the ratio \( y_i / x_i \) (for \( i \in \text{sample} \)) into a weighted average. For example, the Horvitz-Thompson (1952) estimator with a fixed sample size, probability sampling design, where the inclusion probability is proportional to the auxiliary variable \( x \), and selection is without replacement, is an \( M \)-strategy. The Horvitz-Thompson estimator is then

\[
\hat{\mu}_{HT} = N^{-1} \sum_{i=1}^{n} \pi_i^{-1} y_i = \bar{X} n^{-1} \sum_{i=1}^{n} y_i / x_i \quad (2.7)
\]

because \( \bar{X} \) is proportional to the selection probabilities \( \pi_i \), and \( \sum_{i=1}^{N} \pi_i = n \). The estimator in (2.7) is constructed as the mean of the ratios \( y_i / x_i \) multiplied by the known population mean \( \bar{X} \) for the \( x \) variable. Some other \( M \)-strategies discussed in Cassel, Särndal and Wretman (1977, p. 153) are those proposed by Hansen-Hurwitz (1943), Raj (1956), Murthy (1957), and Rao-Hartley-Cochran (1962). An \( R \)-strategy usually uses simple random sampling with or without replacement, and the estimator is either the classical ratio estimator

\[
\hat{\mu}_{\text{Ratio}} = \bar{X} \left( n^{-1} \sum_{i=1}^{n} x_i \right)^{-1} \left( n^{-1} \sum_{i=1}^{n} y_i \right) \quad (2.8)
\]

or some modified version thereof. Besides (2.8), other \( R \)-strategies considered in Cassel, Särndal and Wretman (1977, p. 155) are classical ratio estimation for sampling with replacement, strategies proposed by Tin (1965), Hájek (1949), Lahiri (1951), Midzuno (1952),
Sen (1953), and Hartley-Ross (1954). Cassel, Särndal, and Wretman (1977, Ch 7) discuss properties of \( M \)-strategies and \( R \)-strategies under certain sampling designs.

Mickey (1959) compared the efficiency of a class of ratio and regression type estimators which are design unbiased for random sampling, using a design without replacement.

Fuller (1975) gave fundamental results for regression estimators in survey sampling. Fuller assumed the finite population to be a simple random sample from an infinite superpopulation and showed that the regression coefficients have asymptotically normal distributions, given mild assumptions.

Assume that a finite population, \( \xi_N \), of size \( N \) is drawn from a superpopulation \( \xi \). Denote the \( N \) values of the \( Y \)-characteristic by

\[
Y_N = (y_1, \ldots, y_N)'
\]

(2.9)

and the auxiliary variables by

\[
X_N = (x'_1, \ldots, x'_N)'
\]

(2.10)

where \( x_i = (1, x_{i1}, \ldots, x_{iq}) \), for \( i = 1, 2, \ldots, N \). Let

\[
Q_N = N^{-1}X_N'X_N
\]

(2.11)

\[
H_N = N^{-1}X_N'Y_N
\]

(2.12)

Thus, the vector of regression coefficients for the population \( \xi_N \) is

\[
\beta_N = Q_N^{-1}H_N = (X_N'X_N)^{-1}X_N'Y_N
\]

(2.13)
Assume the expected values over the superpopulation $\xi$ are

$$(Q, H) = E(Q_N, H_N) \quad (2.14)$$

and let $\beta = Q^{-1}H$.

If a sample of size $n$ is drawn from the finite population, let the data be denoted by

$$Y_n = (y_1, y_2, \ldots, y_n)' \quad (2.15)$$

and let the observations for the auxiliary variables be denoted by

$$X_n = (x'_1, x'_2, \ldots, x'_n)', \quad \text{where } x_i = (1, x_{i1}, \ldots, x_{iq}) \text{ for } i = 1, 2, \ldots, n.$$  

The ordinary least squares estimator is

$$\hat{\beta}_n = Q_n^{-1}H_n = (X'_nX_n)^{-1}X'_nY_n \quad (2.17)$$

where

$$(Q_n, H_n) = n^{-1}(X'_nX_n, X'_nY_n). \quad (2.18)$$

Fuller gave the limit distribution of $n^{1/2}(\hat{\beta}_n - \beta)$ when the superpopulation $\xi$ has finite fourth moments and a positive definite covariance matrix, and the sampling design is simple random sampling without replacement. He showed that

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{L} N(0, (1 - f)Q^{-1}GQ^{-1}) \quad (2.19)$$

as $n \to \infty$ and $n/N \to f$, where the notation $\xrightarrow{L}$ indicates convergence in distribution, and

$$G = E\left\{N^{-1}X'_nD^2X_N\right\}, \quad (2.20)$$

$$D = \text{diag}(Y_N - X_N\beta). \quad (2.21)$$
Under the same conditions, a consistent estimator of $G$ in (2.20) is

$$G = (n - q - 1)^{-1} \sum_{i=1}^{n} \hat{d}_i \hat{d}_i,$$  \hspace{1cm} (2.22)

where $\hat{d}_i = x_i \hat{e}_i$, $\hat{e}_i = y_i - x_i \hat{\beta}_n$.

Royall and Cumberland (1981) studied variance estimators for the linear regression estimator of a finite population mean under superpopulation models such as

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$  \hspace{1cm} (2.23)

$$E(e_i) = 0, \quad E(e_i^2) = \sigma^2, \quad E(e_i e_j) = 0, \quad i \neq j.$$  \hspace{1cm} (2.24)

Under model (2.23) and (2.24), the regression estimator $\mu_{OLS}$ in (2.5) is model unbiased. Here an estimator $\hat{\mu}$ is model unbiased for $\bar{Y}$ if

$$E(\hat{\mu} - \bar{Y} | s) = 0,$$  \hspace{1cm} (2.25)

where $s$ is the selected sample. The model-variance of estimator (2.5) is

$$\text{Var}(\hat{\mu}_{OLS} - \bar{Y} | s) = n^{-1} (1 - f) \sigma^2 \left( 1 + (1 - f)^{-1} n^{-1} \left[ (\bar{x} - \bar{X})^2 \left( \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{-1} \right] \right),$$  \hspace{1cm} (2.26)

where $f = nN^{-1}$, $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$, and $\bar{X} = N^{-1} \sum_{i=1}^{N} x_i$. A common estimator of (2.26) is

$$\hat{V} = n^{-1} (1 - f) \hat{\sigma}^2,$$  \hspace{1cm} (2.27)

where

$$\hat{\sigma}^2 = (n - 2)^{-1} \sum_{i=1}^{n} (y_i - \bar{y} - b (x_i - \bar{x}))^2,$$  \hspace{1cm} (2.28)
The variance estimator $\hat{V}$ in (2.27) is model-unbiased,

$$E \left( \hat{V} \mid s \right) = \text{Var} \left( \hat{\mu}_{\text{OLS}} - \hat{\bar{Y}} \mid s \right),$$

under the model (2.23), (2.24). However, if the model in (2.24) is violated, for example, the variance structure is $E(e_i^2) = \sigma^2 x_i$, then the common variance estimator $\hat{V}$ in (2.27) is model biased.

Other important results using the superpopulation model approach are given in Isaki and Fuller (1982). Assume the superpopulation model

$$y_i = x_i \beta + \epsilon_i,$$

for $i = 1, 2, ..., N$, where $x_i = (x_{i1}, x_{i2}, ..., x_{iq})$, $\epsilon_i$ are random variables such that

$$E(\epsilon_i | i) = 0$$

$$E(\epsilon_i\epsilon_j | i, j) = \begin{cases} \gamma_{ii} \sigma^2 & i = j \\ \rho \gamma_{ij}^{1/2} \gamma_{jj}^{1/2} \sigma^2 & i \neq j \end{cases}$$

where $- (N - 1)^{-1} < \rho < 1$. The parameters $\gamma_{ii} (i = 1, 2, ..., N)$ and the population

mean for $x$, $\bar{X} = N^{-1} \sum_{i=1}^{N} x_i$ are known, and $\sigma^2, \beta, \rho$ are unknown. A sample $s$, of size $n$, is drawn from a finite population $\xi_N$, which is a random sample from $\xi$. If $d$ is a predictor of $\hat{Y}$ from $s$, the anticipated variance, defined as

$$AV \{d - \hat{Y} \} = E \left( (d - \hat{Y})^2 - \left( E(d - \hat{Y}) \right)^2 \right),$$
is introduced as a criterion for evaluating probability designs and predictors. Note that the expected value is taken over both the design and the superpopulation model. We will require the concept of design consistency. An estimator $\hat{\mu}$ of $\bar{Y}$ is design consistent if

$$\lim_{n \to \infty} (\hat{\mu} - \bar{Y}|\xi_N) = 0. \quad (2.35)$$

Isaki and Fuller gave the results:

(i) Under model $(2.32)$, assume that $\gamma_{ii}^{1/2}$ is an element of $x_i$, or $\rho \equiv 0$. For a given sample size $n$, the best linear predictor of $\bar{Y}$, within the class of model unbiased predictors as defined in $(2.25)$, conditional on $X' = (x_1', x_2', \ldots, x_n')$ is

$$\mu_{\text{opt}} = \tilde{y} + (1-f)\tilde{X}_{N-n}\tilde{\beta}; \quad (2.36)$$

where $f = N^{-1}n$, $(\tilde{y}, \tilde{x}) = n^{-1}\sum_{i=1}^{n} (y_i, x_i)$, $\tilde{X} = (1-f)\tilde{X}_{N-n} + f\tilde{x}$, and $\tilde{\beta}$ is the generalized least squares estimator

$$\tilde{\beta} = \left(X_n'\Gamma_n^{-1}X_n\right)^{-1}X_n'\Gamma_n^{-1}y_n, \quad (2.37)$$

with $\Gamma_n = \text{diag}(\gamma_{11}, \gamma_{22}, \ldots, \gamma_{nn})$, $y_n = (y_1, y_2, \ldots, y_n)'$.

(ii) Under the condition in (i), if $\gamma_{ii}$ is in the column space of $X$, then the best model consistent linear predictor of $\bar{Y}$ is

$$\mu_{\phi} = \tilde{X}\tilde{\beta}. \quad (2.38)$$

(iii) Under model $(2.32)$, let $\pi_i$ be the probability that element $i$ is selected in a non-replacement sample of size $n$, and let $\pi_i$ be the first element of $x_i$. Let the regression
coefficients weighted by \( \pi_i^{-2} \) be

\[
\hat{\beta} = (B_1, \ldots, B_q)' = (X_n' \pi_n^{-2} X_n)^{-1} (X_n' \pi_n^{-2} y_n)
\]

(2.39)

where \( \pi_n^{-2} = \text{diag}(\pi_1^{-2}, \ldots, \pi_n^{-2}) \), then the best design consistent linear predictor for unequal probability sampling can be expressed in the form commonly used for the regression estimator in simple random sampling as

\[
\hat{\mu}_{\text{Reg}} = \bar{X}_j \hat{\beta} = \bar{y} + \sum_{j=2}^q (\bar{X}_j - \bar{x}) \hat{B}_j,
\]

(2.40)

where \( (\bar{y}, \bar{x}) = n^{-1} \sum_{i=1}^n (f^{-1} \pi_i)^{-1} (y_i, x_i) \) is the vector of weighted sample means with weights \( (f^{-1} \pi_i)^{-1} \).

Isaki and Fuller (1982) assume that the inclusion probabilities \( \pi_i \) are such that for some constants \( \lambda_1, \lambda_2 \),

\[
0 < \lambda_2 < \pi_i = n \left( \sum_{j=1}^N \gamma_{jj}^{1/2} \right)^{-1} \gamma_{ii}^{1/2} < \lambda_1 < 1.
\]

(2.41)

They also assume that the fourth central moments of the regression coefficients and regressors are of the order \( O(n^{-2}) \). Fuller and Isaki proved that the anticipated variance of the regression predictor in (2.40) is

\[
AV(\hat{\mu}_{\text{Reg}} - \bar{Y}) = (1 - \rho) \sigma^2 N^{-2} \left\{ n^{-1} \left( \sum_{i=1}^N \gamma_{ii}^{1/2} \right)^2 - \sum_{i=1}^N \gamma_{ii} \right\} + O(n^{-3/2}).
\]

(2.42)

Fuller and Isaki (1981) proved that for an unequal-probability-without-replacement sampling design with the inclusion probability \( \pi_i \), the estimator for \( \bar{Y} \) which minimizes the design variance,

\[
V(\hat{\mu} - \bar{Y} | \xi_N) = E \left( (\hat{\mu} - E(\hat{\mu}|\xi_N))^2 | \xi_N \right)
\]

(2.43)
is \( \hat{\mu}_{\text{Reg}} \) in (2.40). If the \( \pi_i \) are proportional to the standard errors of the superpopulation as in (2.41), and if \( \pi_i \) and \( \pi_i^2 \) are in the column space of the matrix of independent variables, then the regression estimator (2.40) is also the best linear design consistent estimator of the finite population mean. That is, (2.40) has the minimum design variance defined in (2.43). If the vector whose elements are one is in the column space of the matrix of independent variables, then the regression predictor is location and scale invariant.

Wright (1983) also studied design consistent and nearly optimal regression estimators under the superpopulation model (2.32) with \( \rho = 0 \) in (2.33). He introduced the \( QR \) class of predictors for \( \bar{Y} \), defined by

\[
T_{QR}(n) = \bar{X} \hat{\beta}_n + N^{-1} \sum_{i=1}^{n} r_i \hat{\epsilon}_i, \tag{2.44}
\]

where

\[
\hat{\beta}_n = \left( X'_n Q_n X_n \right)^{-1} X'_n Q_n Y_n, \tag{2.45}
\]

\( Q_n = \text{diag}(q_1, q_2, \ldots, q_n) \), \( \hat{\epsilon}_i = y_i - x'_i \hat{\beta}_n \), and \( q_i > 0 \), \( r_i \geq 0 \) are additional auxiliary variables. Different choices of \( Q_N = \text{diag}(q_1, \ldots, q_N) \) and \( R_N = \text{diag}(r_1, \ldots, r_N) \) yield familiar estimators. For example, if \( q_i = \gamma_{ii}^{1/2} \), \( r_i = 1 \), then \( T_{QR}(n) \) in (2.44) becomes the best linear model unbiased predictor for \( \bar{Y} \) as described in (2.36). If \( q_i = \pi_i^{-2} \) and \( r_i = 0 \), we get the predictor in (2.40). Wright (1983) pointed out that for a probability sampling design, the only exactly design unbiased linear predictor of \( \bar{Y} \) in the form of \( T = \sum_{i=1}^{n} \lambda_i y_i \),
is the Horvitz-Thompson predictor

\[ T_{HT}(n) = \sum_{i=1}^{n} N^{-1} \pi_i^{-1} y_i. \]  

(2.46)

Here a predictor \( T \) of \( \bar{Y} \) is exactly design unbiased if

\[ E\left(T - \bar{Y} | \xi_N\right) = 0 \]  

(2.47)

for all \( \xi_N \).

Because \( \hat{\beta} \) in the regression estimator is a nonlinear function of \((x, y)\), exact design unbiasedness is hard to achieve for predictors in the QR class. Thus, asymptotically design unbiased estimators that satisfy

\[ \lim_{n \to \infty} E\left(T - \bar{Y} | \xi_N\right) = 0, \]  

(2.48)

and design consistent estimators as defined in (2.35) are considered by Brewer (1979), Isaki and Fuller (1982), Robinson and Särndal (1983), and Särndal (1980). If we assume that the superpopulation has finite fourth moments, then for nonreplacement probability sampling under model (2.32) with \( \rho = 0 \) in (2.33),

\[ \hat{\beta}_n - \beta_N = O_p\left(n^{-1/2}\right), \]  

(2.49)

where \( \beta_N \) is the population weighted regression coefficient vector

\[ \beta_N = (X_N' Q_N \pi_N X_N)^{-1} (X_N' Q_N \pi_N y_N). \]  

(2.50)
where $Q_N = \text{diag}(q_1, ..., q_N)$, $\pi_N = \text{diag}(\pi_1, ..., \pi_N)$, $X'_N = (x'_1, ..., x'_N)$, $y_N = (y_1, ..., y_N)'$. See Fuller (1975), Wright (1983), and Fuller, Loughin, and Baker (1994). Therefore, if we define

$$T_{QR}(n, N) = \bar{X}\beta_N + N^{-1} \sum_{i=1}^{n} r_i a_i,$$  

(2.51)

where $a_i = y_i - x_i\beta_N$, and assume that

$$\lim_{n \to \infty} E \{T_{QR}(n) - T_{QR}(n, N) | \xi_N\} = 0$$  

(2.52)

and

$$E \{T_{QR}(n, N) - \bar{Y} | \xi_N\} = 0$$  

(2.53)

for all $\xi_N$, then

$$\lim_{n \to \infty} E \{T_{QR}(n) - \bar{Y} | \xi_N\} = 0.$$  

(2.54)

That is, the predictor $T_{QR}(n)$ in (2.44) is design consistent as defined in (2.35). Wright (1983) stated the following lemma which gives sufficient conditions for $T_{QR}(n)$ to be asymptotically design unbiased.

**Lemma 2.1** Assume that (2.52) holds, then the following conditions are equivalent and sufficient for $T_{QR}(n)$ to be asymptotically design unbiased as defined by (2.35).

(i) $E \{T_{QR}(n, N) - \bar{Y} | \xi_N\} = 0$.

(ii) The vector $c = (c_1, ..., c_N)'$ belongs to the column space of $X_N$, where $c_i = (1 - \pi_i r_i) (\pi_i q_i)^{-1}$.
For example, for the predictor (2.39) in Isaki and Fuller (1982), with $q_i = \pi_i^{-2}$ and $r_i = 0$ is asymptotically design unbiased if $c_i = \pi_i$ is in the column space of the $X$ matrix.

An interesting result obtained in Wright (1983) is that if $\hat{\beta}_n$ is defined as in (2.45), the generalized regression predictor

$$T_{\text{GREG}}(n) = \bar{X}\hat{\beta}_n + \sum_{i=1}^{n} N^{-1} \pi_i^{-1} (y_i - x_i \hat{\beta}_n).$$

(2.55)

is the only predictor within the class of $T_{QR}(n)$ having the same choice of $q_i$ as $T_{\text{GREG}}(n)$ and satisfying the conditions in Lemma 2.1. That is, if a $QR$ class predictor $T_{QR}(n)$ with $q_i$ and $r_i$ such that the conditions of Lemma 2.1 hold, and the $T_{\text{GREG}}(n)$ uses the same $q_i$ to calculate $\hat{\beta}_n$ in (2.45), then $T_{QR}(n) = T_{\text{GREG}}(n)$ for all $y$ and all samples. In other words, if the conditions in Lemma 2.1 are satisfied, the choice of $r_i$ does not change the predictor except $r_i = \pi_i^{-1}$. This suggests that the focus of study for $QR$-type design consistent predictors may be restricted to the generalized regression predictor (2.55). For a design consistent predictor of $QR$ type, under the model in (2.32) with $\rho = 0$ in (2.33), if

$$\lim_{n \to \infty} nE \left( T_{\text{GREG}}(n) - \hat{T} \right)^2 = 0,$$

(2.56)

for

$$\hat{T} = \bar{X}\hat{\beta} + \sum_{i=1}^{n} (N \pi_i)^{-1} (y_i - x_i \hat{\beta}),$$

(2.57)

then Wright (1983) derived the anticipated variance defined in (2.34) for $T_{QR}(n)$ as

$$AV(T_{QR}(n)) = N^{-2} \sum_{i=1}^{N} \left( \pi_i^{-1} - 1 \right) \gamma_i \sigma^2.$$  

(2.58)
He claimed that given fixed \( n \) for each \( \xi_n \) if \( \pi_i \) is proportional to \( \gamma_i^{-1/2} \), then \( AV(T_{QR}(n)) \) may achieve the lower bound on the right side of (2.59)

\[
AV(T_{QR}(n)) \geq n^{-1} \sigma^2 \left\{ \left( N^{-1} \sum_{i=1}^{N} \gamma_i^{-1/2} \right)^2 - n N^{-2} \left( \sum_{i=1}^{N} \gamma_i^{1/2} \right) \right\}.
\] (2.59)

Therefore, a probability sampling design without replacement with \( \pi_i \propto \gamma_i^{-1/2} \) using the generalized regression predictor in (2.55) will be an asymptotically optimal strategy since it is design consistent and its anticipated variance achieve the lower bound in (2.59).

### 2.3 Weighting Procedures

In sample surveys, we use observations on the elements in the sample to make inferences about the finite population. Estimators of the population mean can usually be expressed as a sum of weighted observations in the sample,

\[
\text{estimator of } \bar{Y} = \sum_{i=1}^{n} \omega_i y_i.
\] (2.60)

For example, the Horvitz-Thompson estimator (2.7) is obtained by using the weights \( \omega_i = \pi_i^{-1} N^{-1} \). The regression estimator, suggested in Mickey (1959) and Fuller, Loughin, and Baker (1994), is obtained by using the vector of weights

\[
\omega = (\omega_1, \ldots, \omega_n) = \bar{X} \left( X_n' \pi_n^{-1} X_n \right)^{-1} \left( X_n' \pi_n^{-1} \right),
\] (2.61)

where \( X \) is defined as in (2.16), \( \pi_n = \text{diag}(\pi_1, \ldots, \pi_n) \), and \( \bar{X} \) is the population mean of \( X \). Other discussions of weighted sums of observations are those in Godambe (1955).
Huang and Fuller (1978), Bethlehem and Keller (1987), Lemaitre and Dufour (1987), Smith (1988), and Fuller, Loughin, and Baker (1994). The advantage of weights such as (2.61) is that these weights, having once been computed, can be used for any characteristic $Y$. Because the weights are constructed by regression, it is possible that the weights of (2.61) have negative values for some observations in the sample. Negative weights may lead to a negative estimate for a nonnegative characteristic. Huang (1978) designed a computer program to produce nonnegative weights. Huang and Fuller (1978) described the iterative weight generation procedure and showed that the modified estimator using the nonnegative weights generated by this procedure has the same limit distribution as that of the ordinary regression estimator.

When we use the generalized regression estimator to construct estimators, it is necessary to choose the weights used in the weighted regression. For example, the inverse of the inclusion probability, $\pi_i^{-1}$, is used as the weight in the weighted regression in (2.61). Many researchers have investigated the issues in weighted regression. Smith (1988) argued that in a model-based framework, probability designs are ignorable, and so probability weights have no obvious role from a Bayesian point of view. Särndal (1980) discussed the $\pi$-inverse weights and best linear unbiased weights for an unequal probability sampling. He claimed that these two schemes are equally efficient as far as the first order efficiency goes. He tried to answer the following question: when we draw a sample with unequal probability and construct design consistent estimators of the finite population mean $\hat{Y}$, using weighted
regression estimators, what weights should we use to estimate the regression coefficients?

Assume the superpopulation model

\[ y_i = x_i \beta + \epsilon_i, \quad (2.62) \]

where

\[
E(\epsilon_i) = 0, \quad \text{Var}(\epsilon_i) = \gamma_{ii}\sigma^2 \quad (2.63)
\]

\[ x_i = (1, x_{i1}, ..., x_{iq}), \]

\[ \beta = (\beta_1, ..., \beta_q)', \]

\[ \gamma_{ii} = \gamma(x_i), \text{ where } \gamma(\cdot) \text{ is a known function, and } \sigma^2 \text{ is unknown.} \]

Assume that \( \bar{X} = N^{-1} \sum_{i=1}^{N} x_i \) is known. Let \( X_n, \pi_n \) be as in (2.61). Särndal (1980) considered the generalized regression estimator

\[
\hat{\mu}_{\text{GREG}} = N^{-1} \sum_{i=1}^{n} \pi_i^{-1} y_i + \left( \bar{X} - N^{-1} \sum_{i=1}^{n} \pi_i^{-1} x_i \right) \hat{\beta}
\]

\[ = N^{-1} 1' \pi_n^{-1} y_n + (\bar{X} - N^{-1} 1' \pi_n^{-1} X_n) \hat{\beta}, \quad (2.64) \]

where \( 1 = (1, ..., 1)', y_n = (y_1, ..., y_n)', \) and \( \hat{\beta} \) is an estimator of \( \beta \) of the form.

\[
\hat{\beta} = (W_n'X_n)^{-1} W_n' y_n. \quad (2.65)
\]

Here \( W_n \) is a \( q \times n \) matrix whose elements may or may not depend on the known quantities \( x_i \) and \( \gamma_{ii} \); that is, they are weights applied to each element in the sample. For example, if
we let

\[ W_n = \pi_n^{-1}X_n, \quad (2.66) \]

then \( \hat{\beta} \) in (2.65) is the familiar generalized regression estimator for the regression coefficients \( \beta \). The generalized regression estimator in (2.64) can be obtained by replacing the \( \beta \) with \( \hat{\beta} \) in

\[ T_{GD} = N^{-1} \sum_{i=1}^{n} \pi_i^{-1}y_i + \left( \bar{X} - N^{-1} \sum_{i=1}^{n} \pi_i^{-1}x_i \right) \beta. \quad (2.67) \]

Estimator \( T_{GD} \) is called generalized difference (GD) estimator, and under models (2.62) and (2.63), \( T_{GD} \) is design unbiased and model unbiased (Cassel, Särndal and Wretman, 1977. p. 95). Although the estimation of \( \beta \) introduces a design bias, Särndal (1980) showed that the bias vanishes asymptotically and the increase in the expected mean square error due to estimating \( \beta \) is small compared to the leading term of the expected mean square error. Särndal (1980) discussed the choice of \( \pi_n \) in (2.64) and \( W_n \) in (2.65), and concluded that in terms of the efficiency, the choice of \( \pi_n \) is crucial, and the choice of \( W_n \) is of secondary importance. If weights \( W_n \) are such that there exists a vector \( c = (c_1, ..., c_q)' \) satisfying

\[ 1'\pi_n^{-1} = c'W'_n \quad (2.68) \]

for any sample, he called these weights "\( \pi \)-inverse weights". If weights \( W_n \) are such that

\[ W_n = V_n^{-1}X_n, \quad (2.69) \]

where \( V_n = \text{diag}(\gamma_{11}\sigma^2, ..., \gamma_{nn}\sigma^2) \), he called these weights "best linear unbiased weights".
The \( \pi \)-inverse weights and the best linear unbiased weights are identical if the model (2.62) is such that there exists a vector \( c \) such that

\[
1' \pi_n^{-1} = c' X_n' V_n^{-1}
\]

(2.70) for all samples. Särndal (1980) gave the anticipated variance of the regression estimator (2.64) as

\[
E \left( \hat{\mu}_{GREG} - \tilde{Y} \right)^2 = \sigma^2 E(A(n)) + \sigma^2 E(B(n)) + 2E(D(n)),
\]

(2.71) where \( A(n), B(n), \) and \( D(n) \) are random sample quantities

\[
A(n) = N^{-2} \left\{ \sum_{i=1}^{n} \gamma_{ii} \left( \pi_i^{-2} - 2\pi_i^{-1} \right) \right\} + N^{-2} \sum_{i=1}^{N} \gamma_{ii},
\]

\[
B(n) = N^{-2} \left\{ \left( 1' X_N + 1' \pi_n^{-1} X_n \right) G_n' \left\{ 1' X_N - 1' \pi_n^{-1} X_n \right\} \right\}' V_n \left\{ 1' X_N - 1' \pi_n^{-1} X_n \right\},
\]

\[
D(n) = N^{-2} \left\{ 1' X_N - 1_n \pi_n^{-1} X_n \right\} G_n' V_n \left( I - G_n X_n \right) \pi_n^{-1} 1_n,
\]

(2.72) with \( G_n' = \left( W_n' X_n \right)^{-1} W_n \). For the \( \pi \)-inverse weights and the best linear unbiased weights, \( D(n) = 0 \) for every sample, and (2.71) becomes

\[
E \left( \hat{\mu}_{GREG} - \tilde{Y} \right)^2 = \sigma^2 E(A(n)) + \sigma^2 E(B(n)).
\]

(2.73) Under mild assumptions,

\[
E(A(n)) = O_p\left(n^{-1}\right),
\]

\[
E(B(n)) = O_p\left(n^{-2}\right),
\]

(2.74)
and $E(A(n))$ is minimized by the asymptotically optimum inclusion probabilities

$$
\pi_i = n \gamma_i^{1/2} \left( \sum_{j=1}^{N} \gamma_j^{1/2} \right)^{-1} \propto \gamma_i^{1/2}.
$$

(2.75)

Special superpopulation models were also considered by Särndal.

Särndal, Swensson and Wretman (1989) discussed the weighted residual technique for estimating the variance of the generalized regression estimator of the finite population mean. Various "$g$-weights" were proposed in that paper, where $g$ is a function of the inverse of the inclusion probability, $\pi_i^{-1}$. They attempted to answer the question, "How does one construct a variance estimator for the generalized regression estimator that combines simplicity with generality, that gives valid designed-based confidence intervals and that is at the same time, essentially unbiased with respect to an assumed regression model?" They advocated variance estimators which (i) give valid design-based confidence intervals, (ii) are nearly unbiased under a suitable chosen regression model, and (iii) work well for conditional inference.

Consider the model in (2.62) and the generalized regression estimator in (2.64). Denote the regression residuals by

$$
e_i = y_i - \hat{y}_i = y_i - x_i \hat{\beta}, \quad i = 1, 2, ..., N.
$$

(2.76)
The regression estimator in (2.64) can be written in terms of the predicted values and the residuals

$$\hat{\mu}_{GREG} = N^{-1} \sum_{i=1}^{N} \hat{y}_i + N^{-1} \sum_{i=1}^{N} \pi_i^{-1} t_i e_i,$$  \hspace{1cm} (2.77)

where \( t = (t_1, \ldots, t_N) \) are the indicator variables,

$$t_i = \begin{cases} 
1 & \text{if element } i \text{ is selected,} \\
0 & \text{otherwise.} 
\end{cases} \hspace{1cm} (2.78)$$

The regression coefficient vector \( \hat{\beta} \) in (2.65) can be estimated by

$$\hat{\beta} = \left( \sum_{i=1}^{N} x_i'x_i (\lambda_i \pi_i)^{-1} t_i \right)^{-1} \left( \sum_{i=1}^{N} x_i'y_i (\lambda_i \pi_i)^{-1} t_i \right),$$  \hspace{1cm} (2.79)

which is design consistent for the population regression vector

$$\beta = \left( \sum_{i=1}^{N} \lambda_i^{-1} x_i'x_i \right)^{-1} \left( \sum_{i=1}^{N} \lambda_i^{-1} x_i'y_i \right),$$  \hspace{1cm} (2.80)

under conditions such as those used in Fuller (1975). The weight \( \lambda_i^{-1} \) is one that is deemed appropriate in a census fit and is unrelated to the sampling weight \( \pi_i^{-1} \). For example, \( \lambda_i \) could be the response probability for element \( i \) conditional on element \( i \) being selected. We assume that \( \lambda_i > 0 \) for all \( i = 1, 2, \ldots, N \).

As discussed in Särndal (1980), Isaki and Fuller (1982), Wright (1983), Fuller, Loughin, and Baker (1994), a sufficient condition for \( \hat{\beta} \) in (2.79) to be design consistent is

$$N^{-1} \sum_{i=1}^{N} e_i t_i \pi_i^{-1} = 0 \hspace{1cm} (2.81)$$
for all \( t = (t_1, \ldots, t_N)' \) and \( Y = (y_1, \ldots, y_N)' \). Condition (2.81) will hold if there exists a \( q \)-vector \( \mathbf{c} \) such that

\[
\lambda_i = x_i \mathbf{c}
\]

(2.82)

for \( i = 1, 2, \ldots, N \). We can always satisfy (2.82) by including \( \lambda_i \) as an auxiliary variable. Therefore, we assume (2.82) holds for the rest of the discussion of this section. Under the condition (2.82), the regression estimator in (2.77) can be written in the form

\[
\hat{\mu} = N^{-1} \sum_{i=1}^{N} \bar{y}_i = N^{-1} \sum_{i=1}^{N} t_i \pi^{-1}_i g_i (t) y_i,
\]

(2.83)

where the \( g \)-weight \( g_i (t) \) is a function of the indicator variables \( t \)

\[
g_i (t) = \bar{X} \left[ N^{-1} \sum_{j=1}^{N} x_j' x_j \lambda_j^{-1} \pi_j^{-1} t_j \right]^{-1} x_i' \lambda_i^{-1} t_i.
\]

(2.84)

We will omit the \( t \) in the notation of \( g_i (t) \). The \( g \)-weights have the properties:

(i) They yield the correct population values for auxiliary variables:

\[
N^{-1} \sum_{i=1}^{N} t_i \pi_i^{-1} g_i x_i = \bar{X}.
\]

(2.85)

(ii) For any given \( i \), \( g_i \) is a random variable due to the random sample \( s \). Under regularity conditions, such as those used in Fuller (1975),

\[
\text{plim}_{n \to \infty} (g_i - 1|\xi_N) = 0.
\]

(2.86)

Therefore, for large samples, \( g_i \) may be approximated by unity. Using the \( g \)-weight in (2.84), an estimator of \( V (\hat{\mu}) \) is

\[
\hat{V}_g (\hat{\mu}) = N^{-2} \sum_{i,j} t_i t_j \left( \Delta_{ij} \pi_i^{-1} \right) \left( g_i \pi_i^{-1} e_i \right) \left( g_j \pi_j^{-1} e_j \right),
\]

(2.87)
where

\[ \Delta_{ij} = \pi_{ij} - \pi_i \pi_j, \]
\[ e_i = y_i - x_i' \bar{\beta}, \]

and \( \bar{\beta} \) is as in (2.79). Särndal, Swensson and Wretman (1989) claimed that \( \hat{V}_g(\hat{\mu}) \) is design consistent, i.e.,

\[ \lim_{n \to \infty} \left( \hat{V}_g(\hat{\mu}) - V(\hat{\mu}) | \xi_N \right) = 0. \quad (2.88) \]

Under the regression superpopulation model

\[ y_i = x_i \beta + \epsilon_i, \quad (2.89) \]

where the \( \epsilon_i \) are independent and such that

\[ E(\epsilon_i | i) = 0, \quad V(\epsilon_i | i) = \sigma_i^2 = \sigma^2 \gamma_i = x_i' \sigma^2, \quad i = 1, 2, \ldots \quad (2.90) \]

for some vector \( \sigma \), Särndal, Swensson and Wretman (1989) showed that relative model bias (RMB)

\[ \text{RMB} \left( \hat{V}(\hat{\mu}) | s \right) = (\text{MSE}(\hat{\mu}|s))^{-1} \left[ E \left( \hat{V}(\hat{\mu}) | s \right) - \text{MSE}(\hat{\mu}|s) \right], \quad (2.91) \]

where

\[ \text{MSE}(\hat{\mu}|s) = E \left\{ \left( \hat{\mu} - \bar{Y} \right)^2 | s \right\}, \quad (2.92) \]

may be very small or even exactly zero.
2.4 Nonresponse

Nonresponse is an important potential source of error in surveys. It represents a deviation from the probability sampling design. In many complex surveys, nonresponse is common and if the nonrespondents differ from the respondents, direct estimates based on the respondents will be biased. Therefore, reducing the bias due to nonresponse is a primary goal in the analysis of survey data.

Nonresponse has been studied by numerous researchers under various terms. The term "nonresponse" is used by Kish (1965, p. 532), Cochran (1977, p. 359), Fuller, Loughin, and Baker (1994); "missing data" is used by Zarkovich (1966) and Ford (1976); "sampling mortality" is used by Sudman (1976); and "incomplete samples" is used by Sukhatme and Sukhatme (1970). We will use the term nonresponse to refer to the failure to obtain data from an element in a selected sample. That is, a nonrespondent is eligible for the study but fails to respond.

In a complex survey, nonresponse may occur at different stages of data collection; when locating the sample elements, soliciting the located element to participate in the survey, or collecting the data from sample elements.

Two common strategies for treating nonresponse in survey practice are weighting and imputation. Using the weighting strategy, the original sampling weights for respondents are inflated by dividing by estimates of the response probability. The imputation strategy
replaces the missing values of the nonrespondents by imputed values (Kalton and Kasprzyk, 1982).

The weighting strategy is sometimes called nonresponse adjustment (Hanson, 1978). There are two general models for nonresponse, (i) the deterministic model and (ii) the stochastic model. Both are discussed by Kalsbeek (1980) and Cassel, Särndal, and Wretman (1983).

A deterministic model divides the population of \( N \) elements into two mutually exclusive and exhaustive subsets (strata): (i) \( N_1 \) elements which would respond, if selected in a sample, and (ii) \( N_0 \) elements which would not respond, if selected in a sample. Therefore, \( N = N_0 + N_1 \). The deterministic view is followed in some sampling textbooks, such as Cochran (1977, Ch. 13).

A stochastic model is distinguished from a deterministic model in that the response probabilities, \( p_i \) (\( i = 1, 2, \ldots, N \)) may be any value between zero and one, whereas, in a deterministic model, we have either \( p_i = 0 \) or \( p_i = 1 \). A stochastic model may be a more realistic model for a survey. The response probability is often assumed to be correlated with certain characteristics. For example, response probabilities may have a relationship with age, race, and income for a human population survey.

Some of the literature restricts the response probabilities, \( p_i \). Politz and Simons (1949) assumed a discrete distribution over the \( N \) elements. Deming (1953) views the population
as consisting of six classes, according to the average proportion of interviews that would be completed successfully in eight attempts. Rubin (1977) formulated a measure of the impact of nonresponse using the Bayesian viewpoint. Folsom and Witt (1994) assume that response probabilities follow a logistic regression model.

We use the stochastic model for nonresponse behavior and assume a finite population \( \xi_N \) of size \( N \), and a sample design. Let \( t = (t_1, ..., t_N)' \) be the indicator variables for sample selection defined in (2.78), and let

\[
\mathbf{p} = (\pi_1, \pi_2, ..., \pi_N)' = E(t),
\]  
\[ (2.93) \]

\[
\mathbf{\Pi} = \begin{pmatrix} \pi_1 & \pi_{12} & \cdots & \pi_{1N} \\ \pi_{21} & \pi_2 & \cdots & \pi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{N1} & \pi_{N2} & \cdots & \pi_N \end{pmatrix},
\]  
\[ (2.94) \]

where the joint inclusion probability \( \pi_{k\ell} = E(t_k t_\ell) \), if \( k \neq \ell \). Let the indicator variables for response be

\[
\mathbf{r} = (r_1, ..., r_N)',
\]  
\[ (2.95) \]

where

\[
r_i = \begin{cases} 
1 & \text{if unit } i \text{ is selected } (t_i = 1) \text{ and responds} \\
0 & \text{otherwise}
\end{cases}
\]  
\[ (2.96) \]
From (2.96), the relationship \( r_i = r_i \cdot t_i \) holds, and the response probability is defined as

\[
p_i = E(r_i|t_i = 1) = \pi_i^{-1} E(r_i).
\]

(2.97)

Särndal (1981) and Bethlehem (1988) discussed a modified Horvitz-Thompson estimator (1952) to handle the nonresponse situation using the weighting strategy. They considered the estimator

\[
\tilde{y}_{HT} = N^{-1} \sum_{k=1}^{N} \left[ \pi_k \hat{p}_{HT} \right]^{-1} y_k r_k
\]

(2.98)

where \( \hat{p}_{HT} \) is the Horvitz-Thompson estimator of the mean response probability

\[
\hat{p}_{HT} = N^{-1} \sum_{k=1}^{N} \pi_k r_k.
\]

(2.99)

The estimator \( \tilde{y}_{HT} \) in (2.98) "inflates" the weights by an unbiased estimator \( \hat{p}_{HT} \) of the population mean of response probabilities

\[
\hat{p}_N = N^{-1} \sum_{i=1}^{N} p_i.
\]

(2.100)

The approximate bias of the estimator (2.98) is

\[
\text{Bias} \left( \tilde{y}_{HT} \right) = E \left( \tilde{y}_{HT} \right) - \bar{Y} = \bar{Y}^* - \bar{Y} = \hat{p}_{N}^{-1} C_{pY},
\]

(2.101)

where

\[
\bar{Y}^* = N^{-1} \bar{p}_N^{-1} \sum_{i=1}^{N} p_i y_i.
\]

(2.102)

\[
C_{pY} = N^{-1} \sum_{i=1}^{N} (p_i - \bar{p}_N) \left( y_i - \bar{Y} \right).
\]

(2.103)
From (2.101), if $C_{py}$ is close to zero, the bias will be small. Therefore, if there is a small correlation between the response probability and the target variables, the bias will be small. Särndal (1980) claimed that the estimator defined in (2.98) has properties which make (2.98) preferable to the Horvitz-Thompson estimator without the adjustment of nonresponse.

\[ \tilde{y}_{HT} = N^{-1} \sum_{k=1}^{N} \pi_k^{-1} y_k r_k. \]  

(2.104)

The modified Horvitz-Thompson estimator in (2.98) is an example of reweighting the respondent's measurements without using auxiliary information.

Regression estimators provide an effective way to use auxiliary data to reduce the bias due to the nonresponse. If the control variables used in a regression estimator have a strong relationship with the target variables and the response probabilities, regression estimators will significantly reduce bias (Little and Rubin, 1987, Madow, Nisselson, and Olkin, 1983).

Fuller, Loughin, and Baker (1994) discussed the regression estimator under a superpopulation framework. Assume that a finite population $\xi_N$ is a random sample from a superpopulation $\xi$, and the auxiliary variables are $X = (X_1, \ldots, X_q)$. Assume a nonreplacement sample is drawn from $\xi_N$ with the sample size of $n$. Then a regression estimator of the mean of $Y$ is

\[ \hat{\mu}_{\text{REG}} = \bar{X} \hat{\beta} \]  

(2.105)

where $\bar{X}$ is the known population mean of the control variables and $\hat{\beta}$ is the weighted
regression coefficients with $\pi_i^{-1}$ as weights

$$\hat{\beta} = \left( \sum_{i=1}^{N} x_i' \pi_i^{-1} x_i r_i \right)^{-1} \left( \sum_{i=1}^{N} x_i' \pi_i^{-1} y_i r_i \right)$$  \hfill (2.106)$$

and $(x_i, y_i, \pi_i, p_i)$ is the characteristic associated with the element $i$, $x_i = (1, x_{i2}, \ldots, x_{iq})$. $\pi_i$ is defined in (2.93), $p_i$ is defined in (2.97), and $r_i$ is defined in (2.96), for $i = 1, 2, \ldots, N$. We also assume that $\sum_{i=1}^{N} x_i' \pi_i^{-1} x_i r_i$ is nonsingular. Therefore,

$$E \left\{ N^{-1} \sum_{i=1}^{N} x_i' \pi_i^{-1} x_i r_i | \xi_N \right\} = N^{-1} \sum_{i=1}^{N} x_i' p_i x_i,$$  \hfill (2.107)$$

$$E \left\{ N^{-1} \sum_{i=1}^{N} x_i' \pi_i^{-1} y_i r_i | \xi_N \right\} = N^{-1} \sum_{i=1}^{N} x_i' p_i y_i.$$  \hfill (2.108)$$

Denote the population regression coefficient by

$$\gamma = \left( \sum_{i=1}^{N} x_i' p_i x_i \right)^{-1} \left( \sum_{i=1}^{N} x_i' p_i y_i \right).$$  \hfill (2.109)$$

Then, under the conditions used in Fuller (1975), $\hat{\beta}$ is a consistent estimator of $\gamma$ in the sense that

$$\lim_{n \to \infty, N \to \infty} (\hat{\beta} - \gamma) = 0.$$  \hfill (2.110)$$

Fuller, Loughin, and Baker (1994) gave some sufficient conditions for the regression estimator in (2.105) to be consistent in the sense that

$$\lim_{n \to \infty, N \to \infty} (\hat{\mu}_{\text{REG}} - \bar{Y}) = 0.$$  \hfill (2.111)$$

For a given finite population $\xi_N$, for $i = 1, 2, \ldots, N$, define

$$a_i = y_i - x_i \gamma,$$  \hfill (2.112)$$
and

$$\bar{a} = N^{-1} \sum_{i=1}^{N} a_i = \bar{Y} - \bar{X}\gamma,$$

(2.113)

then

$$\hat{\mu}_{\text{REG}} - \bar{Y} = \bar{X}\hat{\beta} - \bar{Y} = \bar{X}(\hat{\beta} - \gamma) - \bar{a}.$$

(2.114)

By (2.110), \(\text{plim}_{n \to \infty} \bar{a} = 0\) implies \(\hat{\mu}_{\text{REG}}\) is a consistent estimator of \(\bar{Y}\). Some sufficient conditions for consistency were given by Fuller, Loughin, and Baker (1994):

1. The superpopulation \(\xi\), from which the finite population \(\xi_N\) is assumed to be drawn satisfies a linear model

$$y_i = x_i\beta + e_i,$$

(2.115)

where \(E(e_i) = 0\), for all \(i = 1, 2, ..., N\), and

$$\beta = \left( \sum_{i=1}^{N} x_i'x_i \right)^{-1} \left( \sum_{i=1}^{N} x_i'y_i \right).$$

(2.116)

If \(p_i = 1\), and the vector 1 is in the column space of \((x_1', x_2', ..., x_N')'\), then \(\gamma = \beta\), and \(\bar{a} = 0\).

2. If there exists a vector \(c\), such that

$$x_i'c_i = p_i^{-1},$$

(2.117)

then \(\bar{a} = 0\).

The condition (2.117) is satisfied if we know the response probability \(p_i\) and include \(p_i^{-1}\) as one of the control variables. However, \(p_i\) is not known in most survey practice. One approach is to define dummy variables as some of the control variables which divide the
whole population into subgroups in such a way that elements in each of the subgroups have the same response probabilities $p_i$. This situation is sometimes described by saying that elements are missing at random in each group. These subgroups are called adjustment cells by Little (1986). Fuller, Loughin, and Baker (1994) pointed out that it is impossible to use the sample to verify if the chosen set of control variables have removed the bias due to the nonresponse even if we believe that the chosen set of control variables are correlated with the target variables and the response probabilities.

When $\hat{\beta}$ in (2.106) is used to estimate $\gamma$ in (2.109), the error can be approximated by

$$\hat{\beta} - \gamma = G^{-1} \sum_{i=1}^{N} x'_i \pi_i^{-1} a_i r_i,$$  \hspace{1cm} (2.118)

where $a_i$ is as in (2.112), $r_i$ are the response indicator variables as in (2.96), and

$$G = \sum_{i=1}^{N} x'_i p_i x_i.$$ \hspace{1cm} (2.119)

Under conditions used in Fuller (1975), a consistent estimator of $G$ is

$$\hat{G} = \sum_{i=1}^{N} T_i x'_i \pi_i^{-1} x_i.$$ \hspace{1cm} (2.120)

Thus, the variance of the regression estimator in (2.105) is approximately the variance of

$$\hat{A} = \sum_{i=1}^{N} x'_i \pi_i^{-1} a_i r_i.$$ \hspace{1cm} (2.121)

Fuller, Loughin, and Baker (1994) also gave a variance estimator for the regression estimator for an unequal probability stratified two-stage sample. We summarize the result in Lemma 2.2.
Lemma 2.2 Let a finite population sequence and sample sequence be as described in Fuller (1975). For each given finite population $\xi_N$, a stratified two-stage sample is drawn from $\xi_N$. The data vector associated with element $(i, j, k)$ in the sample is denoted by $(x_{ijk}, y_{ijk}, \pi_{ijk}, g_{ijk})$. where $i = 1, 2, \ldots, L$ indicates strata, $j = 1, 2, \ldots, n_i$ indicates primary sampling units in stratum $i$, and $k = 1, 2, \ldots, m_{ij}$ indicates the elements of the $j$-th primary sampling unit in stratum $i$. $y_{ijk}$ is the observation for variable $Y$ which is of interest. $x_{ijk}$ is the auxiliary information which is a $q$ dimension row vector with known population mean

$$\bar{X} = N^{-1} \sum_{i,j,k} x_{ijk}, \quad (2.122)$$

$N$ is the total number of elements in the finite population $\xi_N$. $\pi_{ijk}$ is the inclusion probability, and $g_{ijk}$ is a known function of the $x_{ijk}$. Assume the matrix

$$\sum_{i,j,k} x'_{ijk} g_{ijk}^* x_{ijk}$$

is nonsingular, where

$$g_{ijk}^* = \pi_{ijk}^{-1} g_{ijk} r_{ijk}, \quad (2.123)$$

$$r_{ijk} = \begin{cases} 1 & \text{if } (i, j, k) \text{ responds when selected} \\ 0 & \text{otherwise}, \end{cases} \quad (2.124)$$

and define the regression weights by

$$w_{gijk} = \bar{X} \left[ \sum_{i,j,k} x'_{ijk} g_{ijk}^* x_{ijk} \right]^{-1} \left[ x'_{ijk} g_{ijk}^* \right]. \quad (2.125)$$
Let
\[ \hat{\beta}_g = \left[ \sum_{i,j,k} x'_{ijk} g_{ijk} x_{ijk} \right]^{-1} \left[ \sum_{i,j,k} x'_{ijk} g_{ijk} y_{ijk} \right]. \] (2.126)

and assume
\[ \lim_{N \to \infty} \hat{\beta}_g = \gamma_g. \] (2.127)
\[ \lim_{N \to \infty} N^{-1} \sum_{i,j,k} (y_{ijk} - x_{ijk}\gamma_g) = 0. \] (2.128)

Let the regression estimator be
\[ \hat{\mu}_{g\text{REG}} = \bar{X} \hat{\beta}_g \] (2.129)

and let
\[ \hat{V} (\hat{\mu}_{g\text{REG}}) = (n - q)^{-1} (n - 1) \sum_{i=1}^{L} \left\{ (1 - f_i) (n_i - 1)^{-1} n_i \sum_{j=1}^{m_{ij}} (d_{ij} - d_{ij})^2 \right\}, \] (2.130)

where \( n \) is the total number of sample elements,
\[ n = \sum_{i=1}^{L} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} r_{ijk}. \] (2.131)

\( f_i \) is the sampling rate for the \( i \)-th stratum,
\[ f_i = n_i^{-1} N_i, \] (2.132)

\( N_i \) is the total number of clusters in the \( i \)-th stratum, and
\[ d_{ij} = \sum_{k=1}^{m_{ij}} w_{gijk} (y_{ijk} - x_{ijk}\hat{\beta}_g), \] (2.133)
\[ d_{ri} = n_i^{-1} \sum_{j=1}^{n_i} d_{ij}, \quad (2.134) \]

\( n_i > 0 \) for all stratum \( i \).

Assume that the superpopulation \( \xi \) has finite fourth moments, then

\[ \text{plim} \ n \left( \hat{V} (\hat{\mu}_{\text{REG}}) - V (\hat{\mu}_{\text{REG}} | \xi_N) \right) = 0, \quad (2.135) \]

where

\[ V (\hat{\mu}_{\text{REG}} | \xi_N) = E \left\{ (\hat{\mu}_{\text{REG}} - \bar{Y})^2 | \xi_N \right\}. \quad (2.136) \]

\[ \text{plim} \ n \left( \hat{V} (\hat{\mu}_{\text{REG}}) - V (\hat{\mu}_{\text{REG}} | \xi_N) \right) = 0, \quad (2.135) \]

where

\[ V (\hat{\mu}_{\text{REG}} | \xi_N) = E \left\{ (\hat{\mu}_{\text{REG}} - \bar{Y})^2 | \xi_N \right\}. \quad (2.136) \]

Folsom and Witt (1994) studied the nonresponse adjustment methods using data from the Survey of Income and Program Participation (SIPP). They modeled the response probability using a logistic regression model.

\[ p_i = \left\{ 1 + \exp \left( -x_i \theta \right) \right\}^{-1}, \quad (2.137) \]

where \( \theta \) is the logistic parameter vector and \( x_i \) is a vector of characteristics.

Suppose there is a set of initial weights, \( \{ w_i \} \), for each element in the sample. For example, \( \{ w_i \} \) can be initial regression weights such that

\[ N^{-1} \sum_{i=1}^{n} w_i x_i = \bar{X}, \quad (2.138) \]
where \( n \) is the sample size and \( \bar{X} \) is the population mean for \( X \) variables. Folsom and Witt (1994) estimated the logistic regression coefficients \( \theta \) in (2.137) subject to the "generalized raking constraints"

\[
\sum_{i=1}^{n} r_{i} w_{i} \hat{p}_{i}^{-1} x_{i} = \sum_{i=1}^{n} w_{i} x_{i},
\]

(2.139)

where \( \hat{p}_{i} \) is estimated by the logistic model (2.137), and \( r_{i} \) is the response indicator variable for the element \( i \) in the sample as defined in (2.96). Therefore, for the subsample consisting of respondents, indexed by \( 1, 2, \ldots, n^{*} \), we have

\[
\sum_{i=1}^{n^{*}} w_{i}^{*} x_{i} = \sum_{i=1}^{n} w_{i} x_{i},
\]

(2.140)

where \( w_{i}^{*} = r_{i} w_{i} \hat{p}_{i}^{-1} \). In particular, the sum of the weights is preserved when the nonresponse is present. That is, if \( x_{i} \) contains one as the first element,

\[
\sum_{i=1}^{n^{*}} w_{i}^{*} = \sum_{i=1}^{n} w_{i}.
\]

(2.141)

They also proved that the estimator defined by reweighting respondents

\[
\hat{\mu}_{\text{rewighted}} = \left( \sum_{i=1}^{n^{*}} w_{i}^{*} \right)^{-1} \left( \sum_{i=1}^{n^{*}} w_{i}^{*} y_{i} \right)
\]

(2.142)

using the adjusted weights \( w_{i}^{*} \) in (2.140) is the same as imputation where the nonrespondent data are replaced by a linear predictor \( \hat{y}_{i} = x_{i} \hat{\beta} \),

\[
\hat{\mu}_{\text{imputation}} = \left( \sum_{i=1}^{n} w_{i} \right)^{-1} \left( \sum_{i=1}^{n} w_{i} \{ r_{i} y_{i} + (1 - r_{i}) \hat{y}_{i} \} \right)
\]

(2.143)

and the estimated regression coefficient \( \hat{\beta} \) is

\[
\hat{\beta} = \left( \sum_{i=1}^{n} w_{i} \hat{p}_{i}^{-1} (1 - \hat{p}_{i}) x_{i} x_{i}^{'} \right)^{-1} \left( \sum_{i=1}^{n} w_{i} \hat{p}_{i}^{-1} (1 - \hat{p}_{i}) x_{i} y_{i} \right).
\]

(2.144)
They showed that when the prediction model does not hold,

\[ E(y_i - x_i \beta) \neq 0, \]

(2.145)

where \( \beta \) is the regression coefficient which is estimated by \( \hat{\beta} \) in (2.144), but the logistic response probability model (2.137) holds, \( \hat{\mu}_{reweighted} \) is consistent. The equivalence of \( \hat{\mu}_{reweighted} \) and \( \hat{\mu}_{imputation} \) suggests that one can reduce the variance of the estimator by including auxiliary variables, which are strongly correlated with \( y_i \), but not particularly good predictors of the response propensity.

Cassel, Särndal, and Wretman (1983) discussed uses of statistical models in connection with nonresponse. They compared traditional approaches to Bayesian approaches for the nonresponse problem, where traditional approaches refer to the framework in which bias and variance are defined with respect to the sampling distribution generated by the sampling design. In Bayesian analysis of nonresponse, the randomization distributions are of little interest. Cassel, Särndal, and Wretman proposed an analysis that combined randomized sampling design and a model for the response probability. A linear model, such that the variable of interest, \( Y \), has a linear relationship with the auxiliary vector \( x \), was used. They discussed estimators both with and without the adjustment for the response probabilities under the linear model assumption. They also gave several examples on how to estimate the nonresponse probabilities for special sampling designs.

Kott (1994) discussed two distinct models for handling nonresponse in survey sampling: a response model and a parametric model. In a response model, the response behavior is
modeled as a random process, an additional phase of sampling. In a parametric model, the survey data are themselves modeled. He proposed to use these two types of models simultaneously in the estimation of a population mean so that they can provide some protection against the potential for failure in the other model. A regression estimator using response probabilities was presented in which Kott assumes that the response probabilities are known. Also, a direct expansion estimator with imputed missing values was discussed. He proved that these estimators are unbiased under a parametric model and are design unbiased under a response model. Also, the variance estimation for these estimators are discussed under both types of models.

Kalton (1986) studied weighting adjustment and imputation methods for handling wave nonresponse in panel surveys. He pointed out that weighting adjustment is routinely used to compensate for total nonresponse and imputations are used for item nonresponse. He suggested that a number of factors, such as the number of waves of missing data, the type of analysis to be conducted, the availability of auxiliary variables with high predictive power for the missing values, and the work involved in implementing the procedures should be considered for handling wave nonresponse. Kalton and Kish (1984) and Little (1986) discussed more about various imputation methods to handle the nonresponse.

Regression weighting methods for multiphase samples is analogous to regression weighting for nonresponse problems. Kish (1965, pp. 440-450) and Cochran (1977, Ch. 12) gave the basic theory of two-phase sampling. Särndal and Swensson (1987) discussed a number
of two-phase estimators. Breidt and Fuller (1993) discussed the multiple phase sampling method to produce estimators reducing the variance of the estimator relative to single phase sampling. Assume a phase I sample of \( n_I \) element is drawn and observations are obtained for \( x \) variables

\[ x_i = (x_{i1}, \ldots, x_{ip}) \]  

(2.146)

for \( i = 1, 2, \ldots, n_I \). A phase II sample of \( n_{II} \) elements is drawn from the phase I sample and \( y \) variables

\[ y_i = (y_{i1}, \ldots, y_{iq}) \]

(2.147)

observed for \( i = 1, 2, \ldots, n_{II} \). For three-phase sampling, a phase III subsample of \( n_{III} \) elements is drawn from the phase II sample and \( Z \) variables,

\[ z_i = (z_{i1}, \ldots, z_{ir}) \]

(2.148)

observed for \( i = 1, 2, \ldots, n_{III} \). We estimate the mean of \( X \) variables in phase I sample, and use the estimate in the estimation of the mean of \( Y \) variables in phase II sample. Finally, the mean of \( Z \) variables are estimated in phase III sample using estimates for \( X \) and \( Y \) from phase I and phase II samples respectively. Breidt and Fuller (1993) discussed the three-phase regression estimator under a simple random sampling design

\[ \hat{\mu}_{z\text{-three-phase}} = \bar{z}_{III} - (\hat{\mu}_x - \bar{x}_{III}, \hat{\mu}_y - \bar{y}_{III}) \hat{\beta}_{x,xy}, \]

(2.149)

where

\[ (\bar{x}_{III}, \bar{y}_{III}, \bar{z}_{III}) = n_{III}^{-1} \sum_{i=1}^{n_{III}} (x_i, y_i, z_i), \]

(2.150)
\[(\bar{x}_{II}, \bar{y}_{II}) = n_{II}^{-1} \sum_{i=1}^{n_{II}} (x_i, y_i), \quad (2.151)\]

\[\hat{\mu}_x = \bar{x}_I = n_I^{-1} \sum_{i=1}^{n_I} x_i, \quad (2.152)\]

\[\hat{\mu}_y = \bar{y}_{II} + (\hat{\mu}_X - \bar{x}_{II}) \hat{\beta}_{y,x}, \quad (2.153)\]

and \(\hat{\beta}_{y,x}\) is submatrix of the estimated coefficient matrix for the regression of \(Y\) on \(X\) omitting the intercept, and \(\hat{\beta}_{z:xy}\) is the submatrix of the estimated regression coefficient matrix for the regression \(z\) on \(x\) and \(y\), omitting the intercept. They compared the sampling error using the decomposition

\[\hat{\mu}_{z:\text{three-phase}} - \mu_z = \text{Phase I error} + \text{Phase II error} + \text{Phase III error} + \text{estimation error}, \quad (2.154)\]

where \(\mu_z\) is the true population mean for variable \(z\), and estimation error refers to the error due to estimating the regression coefficients, and phase I, II, and III errors are uncorrelated, mean zero random variables. Therefore, when the estimation error in the regression coefficients is negligible, the three-phase estimator is an approximately design unbiased estimator and the usual three-phase estimator will be optimal. However, when there are a large number of regression coefficients to be estimated, then a two-phase estimator which requires less numbers of regression coefficients to be estimated can have smaller variance than the usual three-phase estimator.

The use of the multiphase sampling method to handle the nonresponse problem is discussed by Srinath (1971), Drew and Fuller (1981), Singh and Sedransk (1978, 1985), and
Särndal and Swensson (1987). We group auxiliary information into $x$ and $y$ variables, and assume the respondents consist of a phase II or a phase III sample. The selected original sample is the phase I sample. The difference between the multiphase sampling and nonresponse is that multiphase sampling obeys a known sampling design, while the nonresponse does not obey a known probability distribution in most cases.
3 PRELIMINARIES

In this chapter, we will give some basic concepts and results related to the estimation of the finite population mean. Some preliminary results are quoted for further discussion.

3.1 Basic Concepts and Definitions

We assume that a finite population with \( N \) elements, denoted by \( \xi_N \), is a random sample drawn from a superpopulation \( \xi \). In describing the probability model, each element of the finite population is denoted by the pair \((i, y_i)\), where \( i \) is the index of the element, and \( y_i \) is the outcome associated with the index \( i \). Thus,

\[
\xi_N = \{(1, y_1), \ldots, (N, y_N)\}.
\]

A sample \( s \) from \( \xi_N \) is a subset of the finite population \( \xi_N \).

\[
s = \{(i_1, y_{i_1}), \ldots, (i_n, y_{i_n}) | (i_j, y_{i_j}) \in \xi_N \text{ for } j = 1, 2, \ldots, n\}.
\]

Elements in \( s \) may be repeated when \( s \) is a replacement sample.

If sample \( s \) contains \( n \) distinct elements, to simplify notation, we use the first \( n \) subscripts to define the sample

\[
s = \{(1, y_1), (2, y_2), \ldots, (n, y_n)\} \subset \xi_N.
\]

although the \( n \) elements of \( s \) are not necessarily the first \( n \) elements of \( \xi_N \). We assume the characteristic \( Y \) is of interest in our study. In this chapter, we focus on estimation of the
finite population mean of \( Y \), defined by

\[
\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i.
\]

(3.4)

Let \( \varphi_{\xi_N}^* \) be the set of all possible samples from a given finite population \( \xi_N \),

\[
\varphi_{\xi_N}^* = \{ s : s \subset \xi_N \}.
\]

(3.5)

A sampling design \( p \) is a probability measure for \( \varphi_{\xi_N}^* \), given \( \xi_N \). Therefore, \( p(s) \geq 0 \) for all \( s \in \varphi_{\xi_N}^* \), and

\[
\sum_{s \in \varphi_{\xi_N}^*} p(s) = 1. \tag{3.6}
\]

If a sampling design \( p \) does not depend on the \( y \)-values associated with the elements selected in the sample, we call \( p \) a noninformative design. Given a finite population \( \xi_N \) and a sampling design \( p \) over the set of all possible samples \( \varphi_{\xi_N}^* \), we give the definition of the inclusion probability.

**Definition 3.1** The inclusion probability for element \( i \) in \( \xi_N \) is

\[
\pi_i = \sum_{\{s \in \varphi_{\xi_N}^*: i \in s\}} p(s), \tag{3.7}
\]

where \( p \) is the sampling design.

**Example 3.1** For a nonreplacement simple random sample of size \( n \), drawn from a finite population \( \xi_N \) of size \( N \),

\[
p(s) = \binom{N}{n}^{-1}
\]
for all \( s \in \varphi_{\xi N} \), and the inclusion probability

\[
\pi_i = \sum_{s \in \varphi_{\xi N} : i \in s} p(s) = \binom{N}{n}^{-1} \binom{N-1}{n-1} = N^{-1}n,
\]

for \( i = 1, 2, \ldots, N \).

We assume that our sampling designs are noninformative designs, i.e., the probability of a sample being selected does not depend on the \( Y \)-values associated with the labels of samples.

We assume that each element of the finite population is a random selection from a superpopulation \( \xi \). Thus, \((y_1, y_2, \ldots, y_N)\) is a random vector. We assume

\[
E \{ y_i | i \} = \mu_i,
\]

and

\[
E \{ y_i y_j | i, j \} = \sigma^2 \gamma_{ij} + \mu_i \mu_j,
\]

where \( \gamma_{ij} > 0 \).

An estimator of the finite population mean \( \bar{Y} \), denoted by \( \hat{\mu} \) is a function of the sample size \( s \). The estimator \( \hat{\mu} \) depends on \( y_1, y_2, \ldots, y_N \) only through those \( y_i \) for which \( i \in s \).
Definition 3.2 An estimator $\hat{\mu}$ is a design-unbiased estimator of $\bar{Y}$ if and only if for a given finite population $\xi_N$, and design $p$.

$$E\{\hat{\mu}|\xi_N\} = \bar{Y}. \quad (3.9)$$

The estimator $\hat{\mu}$ is a model-unbiased estimator for $\bar{Y}$ if and only if for any given sample $s$.

$$E\{(\hat{\mu} - \bar{Y})|s\} = 0. \quad (3.10)$$

The estimator $\hat{\mu}$ is called a design-model-unbiased estimator of $\bar{Y}$ if and only if

$$E\{\hat{\mu} - \bar{Y}\} = 0. \quad (3.11)$$

Among various possible criteria for judging the estimators of $\bar{Y}$, one of the most commonly used is the mean square error (MSE)

$$MSE(\hat{\mu}) = E(\hat{\mu} - \bar{Y})^2. \quad (3.12)$$

Also of interest is the mean square error conditional on the finite population $\xi_N$

$$MSE(\hat{\mu}|\xi_N) = E\{(\hat{\mu} - \bar{Y})^2|\xi_N\}. \quad (3.13)$$

See Cochran (1946), Godambe (1955), Isaki and Fuller (1982) for discussions of MSE for survey samples.
Since \( \bar{Y} \) is the characteristic of interest, it is reasonable to choose an estimator \( \hat{\mu} \) which minimizes the MSE(\( \hat{\mu} \)) in (3.12). Cassel, Särndal and Wretman (1977, p. 94) state Lemma 3.1.

**Lemma 3.1** Let \( \hat{\mu} \) be an estimator of \( \bar{Y} \). For any superpopulation \( \xi \) and for any noninformative design \( p \),

\[
E \left\{ (\hat{\mu} - \bar{Y})^2 \right\} = E \left\{ V(\hat{\mu}|s) \right\} + 2E \left\{ (\bar{Y} - \mu) E(\hat{\mu} - \mu|\xi_N) \right\}. \tag{3.14}
\]

where

\[
V(\hat{\mu}|s) = E \left\{ ([\hat{\mu} - E(\hat{\mu}|s)]^2 | s) \right\}, \tag{3.15}
\]

\[
B(\hat{\mu}|s) = E \left\{ \hat{\mu} - \bar{Y} | s \right\} \tag{3.16}
\]

are called the model-variance of \( \hat{\mu} \) and the model-bias of \( \hat{\mu} \) conditional on the given sample, respectively, and

\[
\mu = E \{ \bar{Y} \} = N^{-1} \sum_{i=1}^{N} E \{ y_i \}
\]

denotes the average of the expected values of the \( y \)-values. Then

(i) If \( \hat{\mu} \) is design-unbiased,

\[
E \left\{ (\hat{\mu} - \bar{Y})^2 \right\} = E \left\{ V(\hat{\mu}|s) \right\} + E \left\{ B(\hat{\mu}|s)^2 \right\} - V(\bar{Y}). \tag{3.17}
\]
(ii) If \( \hat{\mu} \) is design-unbiased, as well as model-unbiased,

\[
E \left\{ (\hat{\mu} - \bar{Y})^2 \right\} = E \{ V(\hat{\mu}|s) \} - V(\bar{Y}). \tag{3.18}
\]

Lemma 3.1 gives a useful partition of MSE(\( \hat{\mu} \)) and simplifies the computation of mean square error for design-unbiased and model-unbiased estimators.

### 3.2 Sequences of Finite Populations

In this section, we discuss the estimation of the finite population mean under the framework of a sequence of finite populations. A finite population, \( \xi_N \) is assumed to be randomly drawn from a superpopulation \( \xi \). Let \( t = (t_1, ..., t_N)' \) be the vector of indicator random variables for a sample from \( \xi_N \), where

\[
t_i = \begin{cases} 
1 & \text{if element } i \text{ is selected in the sample } s \\
0 & \text{otherwise.}
\end{cases} \tag{3.19}
\]

Thus, a sample \( s \) from \( \xi_N \) can also be written as

\[
s = \{(i, y_i) \in \xi_N : t_i = 1\}. \tag{3.20}
\]

The inclusion probability of Definition 3.1 is \( \pi_i = E(t_i) \) for \( i = 1, 2, ..., N \). In order to investigate the large sample properties of estimators, we define a finite population sequence and an associated sequence of samples. Let \( \{(y_i, w_i) : i = 1, 2, ...\} \) be a sequence of
elements from a superpopulation $\xi$, where $(y_i, w_i)$ are characteristics associated with the $i$-th element, and $w_i > 0$ for all $i$. Let $\{\xi_{N(t)}, t = 1, 2, \ldots\}$ denote a sequence of finite populations randomly drawn from $\xi$. The size of population $\xi_{N(t)}$ is $N(t), 0 < N(1) < N(2) < \ldots < N(t) < \ldots$. That is, for $t = 1, 2, \ldots$

$$\xi_{N(t)} = \{(y_i, w_i) : i = 1, 2, \ldots, N(t)\}, \quad (3.21)$$

and the finite populations are nested. $\xi_N(1) \subset \xi_N(2) \subset \ldots \subset \xi_N(t) \subset \ldots$. Let a sample $s(t)$, with $n(t)$ distinct elements, be drawn from the finite population $\xi_{N(t)}$. We restrict $n(1) < n(2) < n(3) < \ldots$ and $n(t) < N(t)$ for all $t$. Although the sequence $\{\xi_{N(t)} : t = 1, 2, \ldots\}$ is nested, $\{s(t) : t = 1, 2, \ldots\}$ is not necessarily nested. The characteristic $\{w_i\}$ is used to define the inclusion probability. That is, for a given finite population $\xi_{N(t)}$, assume $0 < \pi(i) < 1$, and

$$\pi(i) = \frac{1}{\sum_{k=1}^{N(t)} w_k} w_i$$

for $t = 1, 2, \ldots$, where $\pi(i)$ is the inclusion probability for element $i$ in the finite population $\xi_{N(t)}$. We also assume that there is a limit for the sample fraction,

$$f = \lim_{t \to \infty} \frac{N^{-1}(t)}{\pi(i)}. \quad (3.23)$$

If no confusion results, we will drop the subscript $t$ which identifies the finite population.

We define design consistency for an estimator $\hat{\mu}$ for $\hat{\gamma}$.

**Definition 3.3** Given the finite population sequence in (3.21), let $\{\hat{\mu}(t)\}$ be a sequence of estimators of the finite population mean, where the finite population mean is

$$\hat{\gamma}_t = \frac{1}{N(t)} \sum_{i=1}^{N(t)} y_i.$$

$$\lim_{t \to \infty} \hat{\gamma}_t = \hat{\gamma}.$$
If for all $\epsilon > 0$,
\[
\lim_{t \to \infty} \Pr \left\{ \left| \hat{\mu}(t) - \bar{Y}(t) \right| > \epsilon \xi_{N(t)} \right\} = 0, \tag{3.25}
\]
$\hat{\mu}(t)$ is called design consistent for $\bar{Y}$.

We also write (3.25) as
\[
\plim_{t \to \infty} (\hat{\mu}(t) - \bar{Y}(t)\xi_{N(t)}) = 0. \tag{3.26}
\]

Now we give some notation for order in probability.

**Definition 3.4** Let $\{Z(t), \ t = 1, 2, \ldots\}$ be a sequence of random variables, and $\{a_t, \ t = 1, 2, \ldots\}$ be a sequence of positive real numbers. Let $Z$ be a random variable. Then we write
\[
Z_t - Z = o_p(a_t), \tag{3.27}
\]
if and only if
\[
\plim_{t \to \infty} \left\{ a_t^{-1} (Z_t - Z) \right\} = 0. \tag{3.28}
\]

We write
\[
Z_t - Z = O_p(a_t),
\]
if and only if for any $\epsilon > 0$, there exist $M_t > 0$ and $t_0$ such that when $t > t_0$,
\[
\Pr \left( \left| Z_t - Z \right| > M_t a_t \right) < \epsilon. \tag{3.29}
\]

Lemma 3.2 and Lemma 3.3 give the Taylor expansion which will be used later.

**Lemma 3.2** Let $\{Z_t, \ t = 1, 2, \ldots\}$ be a sequence of random $q$ dimensional vectors, where $Z_t = (z_{t1}, \ldots, z_{tq})'$. Let $\{a_t, \ t = 1, 2, \ldots\}$ be a sequence of positive numbers, such that
\( \alpha_t \to 0 \) as \( t \to \infty \). If \( f(z) \) is a real valued function defined on \( q \)-dimensional Euclidean space with continuous partial derivatives of order \( s \) at the point \( \Theta = (\theta_1, \theta_2, ..., \theta_q)' \), and if

\[
Z_t - \Theta = O_p(\alpha_t),
\]

then

\[
f(Z_t) - f(\Theta) = \sum_{i=1}^{q} f^{(1)}_{i}(\Theta) (z_{it} - \theta_{i}) + \sum_{i_1, i_2=1}^{q} \frac{1}{2!} f^{(2)}_{i_1 i_2}(\Theta) (z_{i_1 t} - \theta_{i_1}) (z_{i_2 t} - \theta_{i_2})
\]

\[
+ ... + \sum_{i_1, ..., i_s=1}^{q} \frac{1}{s!} f^{(s)}_{i_1...i_s}(\Theta) (z_{i_1 t} - \theta_{i_1}) ... (z_{i_s t} - \theta_{i_s}) + O_p(\alpha_t^{s+1}),
\]

where

\[
f^{(i)}_{i_1...i_s}(z) = \frac{\partial^{(i)} f(z)}{\partial z_{i_1} \partial z_{i_2} ... \partial z_{i_s}}
\]

for \( i = 1, 2, ..., s \), denotes the \( i \)-th order partial derivative of \( f \) with respect to \( z_{i_1}, z_{i_2}, ..., z_{i_s} \).


**Lemma 3.3** If we replace (3.30) in Lemma 3.2 by

\[
Z_t - \Theta = o_p(\alpha_t),
\]

then we may replace \( O_p(\alpha_t^{s+1}) \) in (3.31) by \( o_p(\alpha_t^{s+1}) \).
3.3 Sampling with Unequal Probability

We give some basic results for sampling with unequal probabilities.

**Definition 3.5** The second-order inclusion probability, $\pi_{ij}$, of units $i$ and $j$ is the probability of selecting both units $i$ and $j$.

$$\pi_{ij} = \begin{cases} 
E\{t_it_j\} & \text{for } i \neq j \\
\pi_i & \text{if } i = j.
\end{cases} \quad (3.34)$$

Some basic properties of inclusion probabilities are summarized in Lemma 3.4 and Lemma 3.5, given by Hanurav (1966), due to Godambe (1955) and Yates and Grundy (1953).

**Lemma 3.4** Let $\{\pi_i\}$ and $\{\pi_{ij}\}$ be defined as in (3.7) and (3.34). Given a sampling design $p$ such that the sample size of every sample $s$ is $n$, then

$$\sum_{i=1}^{N} \pi_i = n \quad (3.35)$$

$$\sum_{i \neq j=1}^{N} \pi_{ij} = n(n-1), \quad (3.36)$$

and for $i = 1, \ldots, N,$

$$\sum_{j=1 \atop j \neq i}^{N} \pi_{ij} = (n-1)\pi_i. \quad (3.37)$$
Lemma 3.5 Let \( t \) be the vector of indicator variables defined in (3.19), then

\[
E(t|\xi_N) = \pi. \tag{3.38}
\]

where

\[
\pi = (\pi_1, \pi_2, \ldots, \pi_N)' \tag{3.39}
\]

is the vector of inclusion probabilities. And

\[
\text{Cov}(t_i, t_j) = \begin{cases} 
\pi_{ij} - \pi_i \pi_j & \text{if } i \neq j \\
\pi_i (1 - \pi_i) & \text{if } i = j,
\end{cases} \tag{3.40}
\]

is the \((i, j)\)-th element of the \( N \times N \) covariance matrix of \( t \), denoted by

\[
\Sigma = \{\text{Cov}(t_i, t_j)\}. \tag{3.41}
\]

In the discussion of the rest of this section, we will assume that there are constants \( k_L > 0, k_U > 0, \) and \( \lambda_1 > 0 \) such that

\[
0 < k_L \leq \pi_i^{-1} \pi_j^{-1} \pi_{ij} \leq k_U \quad \text{for } i \neq j = 1, \ldots, N \tag{3.42}
\]

and

\[
0 < \lambda_1 < \pi_i < 1 \quad \text{for } i = 1, 2, \ldots, N. \tag{3.43}
\]

Isaki and Fuller (1982) gave the following Lemma which states sufficient conditions for the Horvitz-Thompson estimator to be design consistent.
Lemma 3.6 Let the sequence of finite populations \( \{ \xi_{N(t)} \} \) and samples \( \{ s_{(t)} \} \) be as described in Section 3.2. Let the Horvitz-Thompson estimator be

\[
\hat{\mu}_{HT} = N_{(t)}^{-1} \sum_{i=1}^{n_{(t)}} \pi_{i(t)}^{-1} y_i. \tag{3.44}
\]

Define

\[
g_{ij(t)} = \begin{cases} 
\pi_{i(t)} \pi_{j(t)} - \pi_{ij(t)} & \text{if } \pi_{i(t)} \pi_{j(t)} \geq \pi_{ij(t)}, i \neq j \\
0 & \text{otherwise},
\end{cases} \tag{3.45}
\]

and assume that

\[
N_{(t)}^{-2} \sum_{i \neq j=1}^{N_{(t)}} g_{ij(t)} = O \left( n_{(t)}^{-2r^2} \right) \tag{3.46}
\]

and

\[
N_{(t)}^{-1} \sum_{i=1}^{n_{(t)}} \left( \pi_{i(t)}^{-1} y_i - f_{(t)}^{-1} \bar{Y}_{(t)} \right)^{2k} < M < \infty \tag{3.47}
\]

for \( \delta > 0, r^{-1} + k^{-1} = 1 \). Then

\[
\hat{\mu}_{HT(t)} - \bar{Y}_{(t)} = O_p \left( n_{(t)}^{-\delta} \right). \tag{3.48}
\]

Assume that if \( r = 1 \), then \( \left| \pi_{i(t)}^{-1} y_i \right| \) is bounded, and if \( k = 1 \), then \( g_{ij(t)} n_{(t)}^{2\delta} \) is bounded.

Lemma 3.7, given by Isaki and Fuller (1982), states that under mild conditions, it is possible to construct a sequence of designs such that the variance in the error of the Horvitz-Thompson estimator in (3.44) is \( O \left( n_{(t)}^{-1} \right) \).
Lemma 3.7 Let the sequence of finite populations and samples $\tilde{\mu}_{HT}$ be as defined in Lemma 3.6. Assume

\[
\lim_{t \to \infty} \tilde{Y}_{(t)} = \bar{\mu}_Y. \tag{3.49}
\]

\[
\lim_{t \to \infty} N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} (y_i - \bar{Y}_{(t)})^2 = \sigma_y^2 \tag{3.50}
\]

and for each finite population, $\pi = (\pi_{1(t)}, \pi_{2(t)}, \ldots, \pi_{N(t)})'$ is such that

\[
0 < \lambda_1 < \pi_{i(t)} < \lambda_2 < 1 \tag{3.51}
\]

for $i = 1, 2, \ldots, N_{(t)}$ and some $\lambda_1$ and $\lambda_2$. Then there is a sequence of designs with inclusion probabilities $\pi$, such that

\[
E \left\{ \left( \tilde{\mu}_{HT(t)} - \bar{Y}_{(t)} \right)^2 \right\} = O \left( n_{(t)}^{-1} \right). \tag{3.52}
\]

Since only noninformative designs will be considered, Lemma 3.7 provides a very useful tool in examining the large sample properties of estimators. For example, if $y_i \equiv 1$, (3.49) and (3.50) hold trivially. Therefore by Lemma 3.7,

\[
E \left\{ \left( N_{(t)}^{-1} \sum_{i=1}^{n} \pi_{i(t)}^{-1} - 1 \right)^2 \right\} = O \left( n_{(t)}^{-1} \right). \tag{3.53}
\]

Results such as in (3.53) will be used in the later discussion.
4 ESTIMATION OF THE POPULATION MEAN IN THE PRESENCE OF NONRESPONSE

In this chapter, we discuss estimators of the finite population mean constructed in the presence of nonresponse. In order to investigate the large sample properties of the mean estimators, we will assume a sequence of finite populations \( \{ \xi_N \} \) and a sequence of samples \( \{ s_n \} \). The sequences are described in Section 4.2. We will model the nonresponse behavior by the Poisson sampling mechanism. Horvitz-Thompson estimators and regression estimators that include the inverse of the response probability as an explanatory variable will be considered for cluster sampling. A necessary and sufficient condition for the consistency of regression estimators will be established. The estimation of the response probabilities will also be considered.

4.1 Nonresponse and Poisson Sampling

For a given finite population \( \xi_N \), let the vector of indicator-variables for inclusion be

\[
t = (t_1, \ldots, t_N)',
\]

where

\[
t_i = \begin{cases} 
1 & \text{if element } i \text{ is selected} \\
0 & \text{otherwise}, 
\end{cases}
\]

and let the vector of indicator-variables for response be

\[
r = (r_1, \ldots, r_N)',
\]
where for each element $i$,

$$r_i = \begin{cases} 
1 & \text{if element } i \text{ is selected and responds} \\
0 & \text{otherwise.} 
\end{cases} \quad (4.3)$$

That is, if we denote a sample drawn from $\xi_N$ by $s$, $r_i$ is one if $i \in s$ and element $i$ responds. It is understood that $r_i = 1$ implies that $t_i = 1$.

Define the response probability for element $i$ by

$$p_i = P(r_i = 1| t_i = 1), \quad (4.4)$$

the probability that $i$ responds conditional on $i$ being selected.

Given the inclusion probability and response probability, we define the observation probability.

**Definition 4.1** The probability of that an element is selected in a sample and responds is called the observation probability. $\blacksquare$

The observation probability will be denoted by $\pi^*$ if it is not specified otherwise. We have the relationship,

$$\pi_i^* = E(r_i|\xi_N)$$

$$= Pr(r_i = 1|t_i = 1)Pr(t_i = 1|\xi_N)$$

$$= p_i \pi_i. \quad (4.5)$$

for $i = 1, 2, \ldots, N$. 
Given a selected sample, one model for response behavior is the Poisson sampling mechanism. Poisson sampling was studied by Hájek (1981, Ch. 6), Brewer, Early and Hanif (1984), and Milbrodt (1987). Poisson sampling is a sampling procedure in which an element \( i \) is selected by a Bernoulli trial with success probability \( \pi_i \), and the trials for different elements are independent. We call a sample drawn by Poisson sampling a Poisson sample. Therefore, if a Poisson sample is indicated by the vector \( t = (t_1, \ldots, t_N)' \) in (4.1), the sample size

\[
n_{\text{Poisson}} = \sum_{i=1}^{N} t_i,
\]

is a random variable.

The following Lemma summarizes some basic characteristics of Poisson sampling as stated in Hájek (1981, Ch. 6).

**Lemma 4.1** Let a sample be drawn from a finite population \( \xi_N \), by Poisson sampling with the probability of selection for element \( i \) equal to \( \pi_i \) for \( i = 1, 2, \ldots, N \). Let \( n_{\text{Poisson}} \) be the sample size. Then

\[
E(n_{\text{Poisson}}) = \sum_{i=1}^{N} \pi_i
\]

\[
V(n_{\text{Poisson}}) = \sum_{i=1}^{N} \pi_i (1 - \pi_i)
\]

and the joint inclusion probability for element \( i \) and \( j \) is:

\[
\pi_{ij} = \pi_i \pi_j
\]

for \( i \neq j = 1, 2, \ldots, N \). \( \blacksquare \)
We shall use Poisson sampling to model the nonresponse. Poisson sampling is a rather restrictive model because it assumes the probability that element $i$ responds, does not depend on the probability that element $j$ responds. A number of our results on unbiasedness do not require independence, but independence is used in deriving the variance expressions.

Assume a finite population $\xi_N$ contains $N_c$ primary sampling units, called clusters, where the $i$-th cluster contains $m_i$ elements. A probability sample $s$, which contains $n_c$ clusters, is selected from the finite population $\xi_N$. As in (4.1), we denote the cluster sample by the vector of indicator variables $t = (t_1, \ldots, t_{n_c})'$ where

$$t_i = \begin{cases} 1 & \text{if cluster } i \text{ is selected} \\ 0 & \text{otherwise.} \end{cases}$$

(4.9)

The vector of inclusion probabilities is,

$$\pi = (\pi_1, \ldots, \pi_{n_c})' = E(t|\xi_N).$$

(4.10)

and the joint inclusion probabilities are

$$\pi_{ij} = E(t_it_j|\xi_N),$$

(4.11)

for $i, j = 1, 2, \ldots, N_c$. When a cluster is selected, all members in that cluster are contacted to be measured. However, not all members in a selected cluster may be respondents for the survey. We assume that within each selected cluster, respondents consist of a Poisson sample. That is, if cluster $i$ is selected, the $j$th element in cluster $i$ will respond if the result
of a Bernoulli trial is a success, with success probability $p_{ij}$. Let

$$r_{ij} = \begin{cases} 
1 & \text{if member } j \text{ in cluster } i \text{ responds when cluster } i \text{ is selected} \\
0 & \text{otherwise} 
\end{cases} \quad (4.12)$$

for $i = 1, 2, \ldots, N_c$, and $j = 1, 2, \ldots, m_i$. The response probability of (4.4) defined with the cluster and element subscript is

$$p_{ij} = P(r_{ij} = 1 | t_i = 1). \quad (4.13)$$

We also assume that the response behavior of members in cluster $i$ is independent of that of members in cluster $i$. Thus, by Lemma 4.1,

$$E(r_{ij}r_{i'j'} | t_i, t_{i'} = 1) = \begin{cases} 
p_{ij} & \text{if } i = i', j = j' \\
p_{ij}p_{i'j'} & \text{if } i = i', j \neq j' \\
p_{ij}p_{i'j'} & \text{if } i \neq i', 
\end{cases} \quad (4.14)$$

for every $\xi_N$. This probability structure will be used in the evaluation of the variance of the estimated mean.

### 4.2 Infinite Sequences of Populations and Samples

In this section, we give the framework for our discussion of the large sample properties of estimators. In this section, we will treat the elements of the finite population as fixed quantities. Let $\xi$ be a fixed sequence,

$$\xi = \{(m_i, y_i, p_i) : i = 1, 2, \ldots\}, \quad (4.15)$$
where \( m_i \) is the number of elements in cluster \( i \),

\[
y_i = (y_{i1}, y_{i2}, \ldots, y_{im_i})'
\]  

(4.16)

is the vector of target variables for cluster \( i \).

\[
p_i = (p_{i1}, p_{i2}, \ldots, p_{im_i})'
\]  

(4.17)

is the vector of response probabilities for the elements in cluster \( i \), as defined in (4.13), and

\[
0 < p_{ij} \leq 1 \quad \text{for} \quad j = 1, 2, \ldots, m_i.
\]

Let a sequence of finite populations,

\[
\{\xi_{N_{c(t)}} : t = 1, 2, \ldots\}
\]  

(4.18)

be created from \( \xi \). For the \( t \)-th finite population \( \xi_{N_{c(t)}} \), there are a total of \( N_{c(t)} \) clusters and

\[
N(t) = \sum_{i=1}^{N_{c(t)}} m_i
\]  

(4.19)

is the total number of elements, where

\[
0 < N_{c(1)} < N_{c(2)} < \ldots < N_{c(t)} < \ldots
\]  

(4.20)

We also assume that the finite populations in the sequence are nested,

\[
\xi_{N_{c(1)}} \subset \xi_{N_{c(2)}} \subset \ldots \subset \xi_{N_{c(t)}} \subset \ldots
\]  

(4.21)

Let a probability sample be drawn from each \( \xi_{N_{c(t)}} \) in (4.18),

\[
\{s_{N_{c(t)}} \subset \xi_{N_{c(t)}} : t = 1, 2, \ldots\}
\]  

(4.22)
where sample \( s_{n_{c(t)}} \) has \( n_{c(t)} \) clusters,

\[
n_{c(t)} < N_{c(t)} \quad (4.23)
\]

for \( t = 1, 2, \ldots \) and

\[
n_{c(1)} < n_{c(2)} < \ldots < n_{c(t)} < \ldots \quad (4.24)
\]

The total number of elements which are eligible to be in the sample \( s_{n_{c(t)}} \) is denoted by

\[
n_{t} = \sum_{i \in s_{n_{c(t)}}} m_i. \quad (4.25)
\]

The \( i \)-th cluster in \( \xi_{N_{c(t)}} \) is selected with the inclusion probability,

\[
\pi_{i(t)} = n_{c(t)} \left( \sum_{k=1}^{n_{c(t)}} w_k \right)^{-1} w_i \quad (4.26)
\]

for \( i = 1, 2, \ldots, N_{c(t)} \), where

\[
0 < \pi_{i(t)} < 1, \quad (4.27)
\]

and \( w_i > 0 \) are characteristics associated with each cluster. Note that the sequence of samples \( \{ s_{n_{c(t)}} , t = 1, 2, \ldots \} \) is not necessarily nested although \( \{ \xi_{N_{c(t)}} , t = 1, 2, \ldots \} \) are nested. For convenience, we will often drop the subscript \( t \) which indicates the finite population and the associated sample, and we use the first \( n_{c(t)} \) subscripts to identify the \( n_{c(t)} \) clusters in sample \( s_{n_{c(t)}} \). For example, (4.25) will be written as

\[
n = \sum_{i=1}^{n_c} m_i. \quad (4.28)
\]

For the sequence \( \{ \xi_{N_{c(t)}} , s_{n_{c(t)}} \} \), we will use \( n \to \infty \) or \( N \to \infty \) to indicate the process of \( t \to \infty \) when the subscript \( t \) is omitted.
Within the \( i \)-th cluster in sample \( s_{n_{c(t)}} \), member \( j \) responds and provides the information \( y_{ij} \) with probability \( p_{ij} \). The total number of respondents in a sample \( s_{n_{c(t)}} \) is denoted by

\[
n_{r(t)} = \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} r_{ij},
\]

where \( r_{ij} \) is defined in (4.12).

### 4.3 The Horvitz-Thompson Estimator with Nonresponse Adjustment

We assume the finite population and associated cluster sample sequence described in Section 4.2. For cluster \( i \), we denote the cluster total by

\[
y_i = \sum_{j=1}^{m_i} y_{ij},
\]

for \( i = 1, 2, \ldots \). If we know the response probabilities \( p_{ij} \), then an estimator of \( y_i \) is

\[
\hat{y}_i = \sum_{j=1}^{m_i} y_{ij} p_{ij}^{-1} r_{ij},
\]

and we denote the estimation error by

\[
b_i = \hat{y}_i - y_i.
\]

Therefore, a Horvitz-Thompson estimator for \( \bar{Y} \) is

\[
\hat{\mu}_{\text{adj,HT}} = \sum_{i=1}^{N_c} \frac{\hat{y}_i}{\pi_i^{-1}t_i} = \sum_{i=1}^{N_c} \sum_{j=1}^{m_i} \frac{1}{\pi_i^{-1}y_{ij}r_{ij}},
\]

where \( \pi_i \) is the inclusion probability for cluster \( i \).
where $N$ is the total number of elements in $\xi_{N_c}$ defined in (4.19) and is assumed known.

We write

$$\hat{\mu}_{\text{adj,HT}} = \hat{d}_1 + \hat{d}_2,$$  \hspace{1cm} (4.34)

where

$$\hat{d}_1 = N^{-1} \sum_{i=1}^{N_c} \pi_i^{-1} (\tilde{y}_i - y_i) t_i = N^{-1} \sum_{i=1}^{N_c} \pi_i^{-1} b_i t_i,$$ \hspace{1cm} (4.35)

$$\hat{d}_2 = N^{-1} \sum_{i=1}^{N_c} \pi_i^{-1} y_i t_i.$$ \hspace{1cm} (4.36)

Note that $\hat{d}_2$ is not observed, but is an useful expression for our discussion. We give some basic results associated with the Horvitz-Thompson estimator $\hat{\mu}_{\text{adj,HT}}$ in Lemma 4.2.

**Lemma 4.2** Let a cluster sample $s_{n_c}$ with $n_c$ cluster be drawn from a finite population $\xi_{N_c}$ of size $N_c$. Assume the sequence described in Section 4.2, and assume that (4.14) holds. Let $\hat{d}_1$ and $\hat{d}_2$ be defined by (4.35) and (4.36), respectively. Then

(i) $E (\hat{d}_1 | \xi_N) = 0$,

(ii) $E (\hat{d}_2 | \xi_N) = \bar{y}$,

(iii) $V (\hat{d}_1 | \xi_N) = E (\hat{d}_1^2 | \xi_N) = N^{-2} \sum_{i=1}^{N_c} \sum_{j=1}^{m_i} \pi_i^{-1} (\pi_j^{-1} - 1) y_{ij}^2$,

(iv) $V (\hat{d}_2 | \xi_N) = \frac{1}{2} N^{-2} \sum_{i,i'} (\pi_i \pi_{i'} - \pi_{i,j} \pi_{i',j}) (\pi_i^{-1} y_i - \pi_{i'}^{-1} y_{i'})^2$,

(v) $\text{cov}(\hat{d}_1, \hat{d}_2 | \xi_N) = E \{ (\hat{d}_1 - E (\hat{d}_1 | \xi_N)) (\hat{d}_2 - E (\hat{d}_2 | \xi_N)) | \xi_N \} = 0$,

where $V (\hat{\mu} | \xi_N) = E (\hat{\mu}^2 | \xi_N) - (E (\hat{\mu} | \xi_N))^2$. 


Proof. We prove each result.

Proof of (i). For each cluster,

\[
E(\hat{y}_i | t_i = 1) = E \left\{ \sum_{j=1}^{m_i} p_{ij}^{-1} r_{ij} y_{ij} | t_i = 1 \right\} = y_i. \tag{4.37}
\]

Therefore,

\[
E(\hat{d}_1 | \xi_N) = N^{-1} \sum_{i=1}^{N_c} \pi_i^{-1} \left\{ \left( \hat{y}_i - y_i \right) t_i | \xi_N \right\}
= N^{-1} \sum_{i=1}^{N_c} E \left\{ \left( \hat{y}_i - y_i \right) | t_i = 1 \right\}
= N^{-1} \sum_{i=1}^{N_c} \{ y_i - y_i \} = 0. \tag{4.38}
\]

Proof of (ii). Using \( E(t_i | \xi_N) = \pi_i \), we have

\[
E(\hat{d}_2 | \xi_N) = N^{-1} \sum_{i=1}^{N_c} \pi_i^{-1} y_i E(t_i | \xi_N) = N^{-1} \sum_{i=1}^{N_c} y_i = \bar{Y}. \tag{4.39}
\]

Proof of (iii). Let \( b_i \) be defined by (4.32). Then, by (4.37),

\[
E(\hat{b}_i | t_i = 1) = 0. \tag{4.40}
\]

If \( i \neq i' \),

\[
E(b_i b_{i'} | t_i t_{i'} = 1) = \left[ \sum_{j,j'} E \left\{ p_{ij}^{-1} p_{ij'}^{-1} r_{ij} y_{ij} r_{ij'} y_{ij'} | t_i t_{i'} = 1 \right\} \right] - y_i y_{i'}
= \left( \sum_{j,j'} y_{ij} y_{ij'} \right) - y_i y_{i'}
= 0, \tag{4.41}
\]
If \( i = i' \), using (4.37),

\[
E \left( b_i^2 | t_i = 1 \right) = \left\{ \sum_{j \neq i} E \left( p_{ij}^{-1} p_{ij'}^{-1} y_{ij} y_{ij'} r_{ij} r_{ij'} | t_i = 1 \right) \right\} - y_i^2
\]

\[
= y_i^2 + \sum_{j=1}^{m_i} p_{ij}^{-1} y_{ij}^2 - y_i^2 - \sum_{j=1}^{m_i} y_{ij}^2
\]

\[
= \sum_{j=1}^{m_i} \left( p_{ij}^{-1} - 1 \right) y_{ij}^2.
\]

(4.42)

Therefore, by (4.41) and (4.42),

\[
E \left( d_1^2 | \xi_N \right) = N^{-2} \sum_{i,i'} \pi_i^{-1} \pi_{i'}^{-1} E \left\{ b_i b_{i'} | t_i t_{i'} = 1 \right\} \Pr (t_i t_{i'} = 1 | \xi_N)
\]

\[
= N^{-2} \sum_{i=1}^{N_i} \pi_i^{-1} E \left\{ b_i^2 | (t_i = 1, \xi_N) \right\}
\]

\[
= N^{-2} \sum_{i=1}^{N_i} \sum_{j=1}^{m_i} \pi_i^{-1} \left( p_{ij}^{-1} - 1 \right) y_{ij}^2,
\]

(4.43)

and from (i), \( V \left( d_1 | \xi_N \right) = E \left( d_1^2 | \xi_N \right) \).

Proof of (iv). Since \( \hat{d}_2 \) is a Horvitz-Thompson estimator for cluster totals by Theorem 9A.5 of Cochran (1977, p. 260), the variance of \( \hat{d}_2 \) is

\[
V \left( \hat{d}_2 | \xi_N \right) = \frac{1}{2N^2} \sum_{i,i'} (\pi_i \pi_{i'} - \pi_{i'} \pi_i) \left[ \pi_i^{-1} y_i - \pi_{i'}^{-1} y_{i'} \right]^2.
\]

Proof of (v). By (i), we have

\[
\text{cov} \left( \hat{d}_1, \hat{d}_2 | \xi_N \right) = E \left( \hat{d}_1 \hat{d}_2 | \xi_N \right).
\]

(4.44)

For any \( i \) and \( i' \),

\[
E (b_i y_{i'} | t_i t_{i'} = 1) = y_{i'}, E (b_i | t_i t_{i'} = 1) = 0.
\]

(4.45)
by (4.44). It follows that

\[
\text{cov}(\hat{d}_1, \hat{d}_2|\xi_N) = N^{-2} \sum_{i,i'} E(b_i|t_i,t_i' = 1) \Pr(t_i,t_i' = 1|\xi_N) \pi_i^{-1} \pi_i'^{-1} = 0.
\]

The design-unbiasedness of the Horvitz-Thompson estimator \(\hat{\mu}_{\text{adj.HT}}\) in (4.33),

\[
E(\hat{\mu}_{\text{adj.HT}}|\xi_N) = \bar{Y},
\]

follows from Lemma 4.2 immediately and its design-variance is

\[
V(\hat{\mu}_{\text{adj.HT}}|\xi_N) = N^{-2} \sum_{i=1}^{N_c} \sum_{j=1}^{m_i} \left[ gpp_i^{-1} \left( \mu_i - 1 \right) y_{ij}^2 \right]
+ \frac{1}{2} N^{-2} \sum_{i,i'=1}^{N_c} \left[ (\pi_i g p_i - \pi_i') \left( \pi_i^{-1} y_{ij} - \pi_i'^{-1} y_{ij}' \right)^2 \right].
\]

Sometimes we need to consider the weighted mean of \(Y\). Therefore, it is convenient to give expressions for weighted means used later. We assume that \(\alpha_{ij}\) is a weight associated with element \(j\) in cluster \(i\), and define the weighted population mean of \(Y\) by

\[
\bar{Y}_\alpha = N^{-1} \sum_{i=1}^{N_c} \sum_{j=1}^{m_i} y_{ij} \alpha_{ij}.
\]

The HT estimator of \(\bar{Y}_\alpha\) is

\[
\hat{\mu}_{\alpha\text{-adj.HT}} = N^{-1} \sum_{i=1}^{N_c} \sum_{j=1}^{m_i} \pi_i^{-1} y_{ij} \alpha_{ij}.
\]
Then, by Lemma 4.2, we have the results:

\[ E(\hat{\mu}_{\text{adj-HT}}|\xi_N) = \bar{Y}_a. \] (4.50)

and

\[
V(\hat{\mu}_{\text{adj-HT}}|\xi_N) = \sum_{i=1}^{N_c} \sum_{j=1}^{m_i} \left[ \pi_i^{-1} (p_{ij} - 1) y_{ij}^2 \right] \\
+ \frac{1}{2} N^{-2} \sum_{i \neq i'}^{N_c} \left[ (\pi_i \pi_i' - \pi_{ii'}) \left( \pi_i^{-1} y_{ai} - \pi_i'^{-1} y_{ai'} \right) \right].
\] (4.51)

where

\[ y_{ai} = \sum_{j=1}^{m_i} y_{ij} \alpha_{ij} \] (4.52)

for \( i = 1, 2, ..., N_c. \)

For the full response case, all \( p_{ij} = 1 \), then \( \hat{\mu}_{\text{adj-HT}} \) becomes the usual Horvitz-Thompson estimator with fixed sample size, and the first term of the (ii) in (4.47) will vanish. Lemma 4.2 gives a decomposition of the nonresponse adjusted Horvitz-Thompson estimator for a cluster sample in a manner such that the two parts of the decomposition are uncorrelated. The variances of the two parts represent the cluster-to-cluster variance and the element-within-cluster variance.

Lemma 4.3 gives sufficient conditions for \( \hat{\mu}_{\text{adj-HT}} \) in (4.33) to be consistent for \( \bar{Y} \). The lemmas are extensions of the results in Fuller and Isaki (1981) to cluster sampling in the presence of nonresponse.
Lemma 4.3 Let the sequence of finite populations \( \{ \xi_{N(t)}, t = 1, 2, \ldots \} \) and the sequence of the associated cluster samples \( \{ s_{n(t)}, t = 1, 2, \ldots \} \) be as described in Section 4.2. Let

\[
g_{ii'}(t) = \left( \pi_i(t) \pi_i'(t) - \pi_{i'}(t) \right) I_{\{ \pi_i(t) \pi_i'(t) - \pi_{i'}(t) > 0 \}},
\]

(4.53)

for \( t = 1, 2, \ldots \), and \( i, i' = 1, \ldots, N(c(t)), \) where \( I_A \) is the indicator function for event \( A \).

Assume that

(i) The relation in (4.14) holds, and there exist constants \( \lambda_1 > 0, \lambda_2 > 0 \), such that for \( \lambda_1 < \pi_i(t) < \lambda_2 < 1, \lambda_1 < p_{ij} < 1 \).

(ii) For some \( \delta_1 > 0, \delta_3 > 0, \delta_4 > 0 \), satisfying \( \delta_3^{-1} + \delta_4^{-1} = 1 \),

\[
N_{c(t)}^{-2} \sum_{i \neq i' = 1}^{N(c(t))} g_{ii'}^{(t)} = O \left( n_{c(t)}^{-2} \delta_1 \right),
\]

and

\[
N_{c(t)}^{-1} \sum_{i = 1}^{N(c(t))} |y_i|^{2\delta_4} = O \left( 1 \right).
\]

(4.56)

(4.57)

If \( \delta_3 = 1 \), then we replace the condition (4.57) by

\[
|y_i| < M_1 > \infty
\]

(4.58)
for some $M_1 > 0$, where $t = 1, 2, \ldots$ and $i = 1, \ldots, N_c(t)$. If $\delta_4 = 1$, then we replace the condition (4.57) by

$$g_{\nu(t)} n_{c(t)}^{2\delta_1} < M_2 < \infty$$

(4.59)

for some $M_2 > 0$, where $t = 1, 2, \ldots$ and $i, i' = 1, \ldots, N_c(t)$.

(iii) For some $\delta_2 > 0$,

$$N^{-2}_{c(t)} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} y_{ij}^2 = O \left( n_{c(t)}^{-2\delta_2} \right).$$

(4.60)

Then

$$V \left( \hat{\mu}_{\text{adj,HT}} | \xi_{N_c(t)} \right) = O \left( n_{c(t)}^{-\min(\delta_1, \delta_2)} \right).$$

(4.61)

and

$$\hat{\mu}_{\text{adj,HT}} - \bar{Y}(t) = O_p \left( n_{c(t)}^{-\min(\delta_1, \delta_2)} \right).$$

(4.62)

where $\hat{\mu}_{\text{adj,HT}}$ is defined in (4.33) and $\bar{Y}(t)$ is the finite population mean of $\xi_{N_c(t)}$.

Proof. By Lemma 4.2, the variance of $\hat{\mu}_{\text{adj,HT}}$ is

$$V \left( \hat{\mu}_{\text{adj,HT}} | \xi_{N_c(t)} \right) = \sigma_1^2(t) + \sigma_2^2(t),$$

(4.63)

where

$$\sigma_1^2(t) \equiv \frac{1}{2} N_{c(t)}^{-2} \sum_{i,i'=1}^{N_c(t)} \left[ (\pi_{i(t)} \pi_{i'(t)} - \pi_{i'(t)} \pi_{i(t)}) \left( \pi_{i(t)}^{-1} y_i - \pi_{i'(t)}^{-1} y_{i'} \right) \right],$$

(4.64)

and

$$\sigma_2^2(t) \equiv N_{c(t)}^{-2} \sum_{i=1}^{N_c(t)} m_i \sum_{j=1}^{m_i} \left[ \pi_{i(t)}^{-1} (P_{ij}^{-1} - 1) y_{ij}^2 \right].$$

(4.65)
When $\delta_3 > 1$ and $\delta_4 > 1$, using the Hölder inequality and the Jensen inequality,

\[
\sigma^2_{1(t)} \leq \frac{1}{2} N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'}(t) \left[ \pi_{ii'}^{-1} y_i - \pi_{ii'}^{-1} y_{i'} \right]^2 \\
\leq \frac{1}{2} \left( N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'}^{\delta_3} \right)^{\delta_3^{-1}} \left( N_{c(t)}^{-2} \sum_{i \neq i'} \left| \pi_{ii'}^{-1} y_i - \pi_{ii'}^{-1} y_{i'} \right|^{2 \delta_4} \right)^{\delta_4^{-1}} \\
\leq \frac{1}{2} \left( N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'}^{\delta_3} \right)^{\delta_3^{-1}} \left( N_{c(t)}^{-2} \sum_{i \neq i'} \left( \pi_{ii'}^{-2 \delta_4} |y_i|^{2 \delta_4} + \pi_{ii'}^{-2 \delta_4} |y_{i'}|^{2 \delta_4} \right)^{2 \delta_4-1} \right)^{\delta_4^{-1}} \\
\leq \frac{1}{2} \left( N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'}^{\delta_3} \right)^{\delta_3^{-1}} \left( N_{c(t)}^{-2} \gamma^{2 \delta_4-1} \left( N_{c(t)} + 1 \right) \lambda^{-2 \delta_4} \sum_{i = 1}^{N_{c(t)}} |y_i|^{2 \delta_4} \right)^{\delta_4^{-1}} \\
\leq 2^{(1-\delta_4^{-1})} \lambda^{-2} \left( N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'}^{\delta_3} \right)^{\delta_3^{-1}} \left( N_{c(t)}^{-1} \sum_{i = 1}^{N_{c(t)}} |y_i|^{2 \delta_4} \right)^{\delta_4^{-1}}. \quad (4.66)
\]

Thus, by condition (ii),

\[
\sigma^2_{1(t)} = O \left( n_{c(t)}^{-2 \delta_4} \right) O \left( 1 \right) = O \left( n_{c(t)}^{-2 \delta_4} \right). \quad (4.67)
\]

If $\delta_3 = 1$, then by the Jensen inequality and condition (4.58),

\[
\sigma^2_{1(t)} \leq \frac{1}{2} N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'}(t) \left[ \pi_{ii'}^{-1} y_i - \pi_{ii'}^{-1} y_{i'} \right]^2 \\
\leq \frac{1}{2} N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'}(t) \left[ \pi_{ii'}^{-2} |y_i|^2 + \pi_{ii'}^{-2} |y_{i'}|^2 \right] \\
\leq N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'}(t) 2 \lambda^{-2} M^2. \quad (4.68)
\]

Thus, by condition (4.56), we have

\[
\sigma^2_{1(t)} = O \left( n_{c(t)}^{-2 \delta_4} \right). \quad (4.69)
\]
If $\delta_1 = 1$, then by condition (4.59) and the Jensen inequality,

$$
\sigma^2_{1(t)} \leq \frac{1}{2} N_{c(t)}^{-2} \sum_{t \neq t'} g_{w(t)} \left[ \pi_{i(t)}^{-1} y_i - \pi_{i'}^{-1} y_{i'} \right]^2 \\
\leq \frac{1}{2} N_{c(t)}^{-2} M_2 n_{c(t)}^{-2\delta_1} \sum_{t \neq t'} 2 \left( \pi_{i(t)}^{-2} y_i^2 + \pi_{i'}^{-2} y_{i'}^2 \right) \\
\leq N_{c(t)}^{-2} n_{c(t)}^{-2\delta_1} M_2 \lambda_1^{-2} \left( N_{c(t)} - 1 \right) \sum_{i=1}^{N_{c(t)}} y_i^2, \quad (4.70)
$$

and using condition (4.57), we have

$$
\sigma^2_{1(t)} = O \left( n_{c(t)}^{-2\delta_1} \right). \quad (4.71)
$$

Using conditions (i) and (ii),

$$
\sigma^2_{2(t)} \leq \lambda_1^{-1} \left( \lambda^{-1} - 1 \right) N_{c(t)}^{-2} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} y_{ij}^2 = O \left( n_{c(t)}^{-2\delta_2} \right). \quad (4.72)
$$

Therefore by (4.72),

$$
V \left( \hat{\mu}_{adj, HT} \mid \xi_{c(t)} \right) = O \left( n_{c(t)}^{-2\min(\delta_1, \delta_2)} \right). \quad (4.73)
$$

Thus, by (4.46) and the Chebyshev inequality,

$$
\hat{\mu}_{adj, HT} - \bar{Y}_{(t)} = O_p \left( n_{c(t)}^{-\min(\delta_1, \delta_2)} \right). \quad (4.74)
$$

For a special case of Lemma 4.3, if the clusters are selected by simple random sampling, the inclusion probability is

$$
\pi_{i(t)} = n_{c(t)} N_{c(t)}^{-1}, \quad (4.75)
$$
the joint inclusion probability is
\[ \pi_{i(t)'} = n_c(t) \left( n_c(t) - 1 \right) \frac{1}{N_c(t)} \left( N_c(t) - 1 \right)^{-1}. \] (4.76)

and
\[ \pi_{i(t)} \pi_{i'(t)} - \pi_{i'(t)} = n_c(t) \frac{n_c(t)}{N_c(t)} \left( N_c(t) - 1 \right)^{-1} \left( N_c(t) - n_c(t) \right) > 0, \] (4.77)

for \( i \neq i' \). Thus, for all \( i \neq i' \),
\[ n_c(t) g_{i(t)}' = n_c(t) \frac{n_c(t)}{N_c(t)} \left( N_c(t) - 1 \right)^{-1} \left( N_c(t) - n_c(t) \right) < 1. \] (4.78)

Therefore, for simple random sampling, condition (4.59) will be satisfied by choosing \( \delta_1 = 1/2 \), and we will have (4.62) provided that (4.60) holds.

Condition (4.57) is for cluster totals, and condition (4.60) is for the sum of squares within each cluster. Often we assume that the conditions in Lemma 4.3 hold for \( \delta_1 = \delta_2 = 1/2 \), and the results in (4.62) become
\[ \tilde{\mu}_{adjHT} - \bar{Y}_{(t)} = O_p \left( n_c(t)^{-1/2} \right). \] (4.79)

Under mild conditions, it is possible to construct a sequence of designs for cluster samples, such that the nonresponse adjusted Horvitz-Thompson estimator \( \tilde{\mu}_{adjHT} \) has design variance of order \( O \left( n_c(t)^{-1} \right) \). Lemma 4.4 extends Lemma 2 of Isaki and Fuller (1982) to cluster sampling with nonresponse in elements.

**Lemma 4.4** Let the sequence of finite populations \( \{ \xi_{N_c(t)} : t = 1, 2, \ldots \} \) and the sequence of associated cluster samples \( \{ s_{n_c(t)} : t = 1, 2, \ldots \} \) be as described in Section 4.2. Assume that
(i) Conditions (4.14), (4.54), and (4.55) hold.

(ii) \[ N^{-1} \sum_{i=1}^{N_c(t)} \left( \sum_{j=1}^{m_i} y_{ij}^2 \right) = O(1). \] (4.80)

Then there exists a sequence of designs with cluster inclusion probabilities \( \pi_{i(t)} \) such that

\[ E \left\{ \left( \hat{\mu}_{\text{adj-HT}} - \bar{Y}_{(t)} \right)^2 | \xi_N \right\} = O \left( n^{-1} \right), \] (4.81)

where \( \hat{\mu}_{\text{adj-HT}} \) is in (4.33).

**Proof.** The proof uses the construction of a *one-per-stratum sampling design* described in Isaki and Fuller (1982). For each \( \xi_{N_{c(t)}} \), arrange the clusters of \( \xi_{N_{c(t)}} \) in natural order and form the cumulative sum of probabilities \( T_{i(t)} = \sum_{k=1}^{i} \pi_{k(t)} \). Divide the population into \( n_{c(t)} \) strata by placing a boundary at the points \( T_{i(t)} = 1, 2, ..., n_{c(t)} \). If a boundary falls in the interval \( \left( \sum_{k=1}^{i_0} \pi_{k(t)}, \sum_{k=1}^{i_0+1} \pi_{k(t)} \right) \), the cluster \( i_0 \) will fall in two adjacent strata. Using the one-per-stratum method given in Fuller (1970, p. 217), select one cluster in each stratum with inclusion probability \( \pi_{i(t)} \). This can be done so that the joint probability of including a cluster \( i \) that belongs to one and the only one stratum \( q \), and a cluster \( i' \) that belongs to one and only one stratum \( q' \), with

\[ q - q' \geq \lambda_2 \left( 1 - \lambda_2 \right)^{-1}, \] (4.82)

is

\[ \pi_{ii'(t)} = \pi_i(t) \pi_{i'(t)}, \] (4.83)
where \( \lambda_2 < 1 \) is the upper bound for all \( \pi_{i(t)} \) as in (4.54). Since all \( \pi_{i(t)} > \lambda_1 > 0 \), there are at most \( \lambda_1^{-1} \) clusters in each stratum. Therefore, by (4.82), for each cluster \( i = 1, 2, \ldots, N_{c(t)} \), there are at most

\[
k_1 \equiv \lambda_1^{-1} \left[ 2\lambda_2 (1 - \lambda_2)^{-1} + 1 \right] \tag{4.84}
\]

clusters such that (4.83) may not hold. Because \( \pi_{i(t)} < \lambda_2 < 1 \), for any \( i, i' \), we have

\[
\left| g_{i'v(t)} \right| = \left| \pi_{i(t)} \pi_{i'(t)} - \pi_{i'v(t)} \right| \leq 2 \left| \pi_{i(t)} \right| \leq 2\lambda_2, \tag{4.85}
\]

where \( g_{iv(t)} \) is defined as in (4.53). By the Hölder inequality, for each \( i = 1, 2, \ldots, N_{c(t)} \),

\[
y_{i}^2 = \left( \sum_{j=1}^{m_i} y_{ij} \right)^2 \leq \left( \sum_{j=1}^{m_i} y_{ij} \right) m_i \leq m \sum_{j=1}^{m_i} y_{ij}^2. \tag{4.86}
\]

Thus, using the Hölder inequality, (4.86), (4.84), and (4.85), we have

\[
\sigma_{1(t)}^2 \equiv \frac{1}{2} N_{c(t)} \sum_{i,i'} \left[ \left( \pi_{i(t)} \pi_{i'(t)} - \pi_{i'v(t)} \right) \left( \pi_{i(t)}^{-1} y_i - \pi_{i'(t)}^{-1} y_{i'} \right)^2 \right]
\leq \frac{1}{2} N_{c(t)} \sum_{i,i'} g_{i'v(t)} \pi_{i'(t)} \lambda_1^{-2} m \left[ \sum_{j=1}^{m_i} y_{ij}^2 + \sum_{j=1}^{m_{i'}} y_{ij}^2 \right]
\leq 2 N_{c(t)} \sum_{i,i'} g_{i'v(t)} \lambda_1^{-2} m \lambda_2 \lambda_1^{-2} m_k \sum_{j=1}^{m_i} y_{ij}^2
\leq (4\lambda_2 \lambda_1^{-2} m_k m) N_{c(t)} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} y_{ij}^2. \tag{4.87}
\]

And by condition (4.80),

\[
\sigma_{1(t)}^2 = O \left( n_{c(t)}^{-1} \right). \tag{4.88}
\]
Also, by (4.54) and (4.80),

\[ \sigma^2_{2(t)} = N_{e(t)}^{-2} \sum_{i=1}^{m_n} \sum_{j=1}^{N_{c_i(t)}} \left[ \pi_{i(t)}^{-1} \left( p_{ij}^{-1} - 1 \right) y_{ij}^2 \right] \]

\[ \leq \left( \lambda_t^{-1} \left( \lambda_t^{-1} - 1 \right) \right) \sum_{i=1}^{N_{e(t)}} \sum_{k=1}^{m_n} \sum_{j=1}^{N_{c(t)}} \frac{y_{ij}^2}{n_{c_i(t)}} \]

\[ = O \left( n_{c_i(t)}^{-1} \right). \tag{4.89} \]

Therefore, from Corollary 4.2.1, (4.88) and (4.89),

\[ E \left\{ \left( \hat{\mu}_{adj-HT} - \bar{y}_{(t)} \right)^2 \right\} | \xi_{N_{c_i(t)}} \]

\[ = V \left( \hat{\mu}_{adj-HT} - \bar{y}_{(t)} \right) | \xi_{N_{c_i(t)}} \]

\[ = \sigma^2_{1(t)} + \sigma^2_{2(t)} = O \left( n_{c_i(t)}^{-1} \right). \tag{4.90} \]

Lemma 4.4 states that if the population means of cluster sum of squares, \( \sum_{j=1}^{m_n} y_{ij}^2 \), are bounded, then under the design described in the proof of Lemma 4.4, we will have (4.81), provided the inclusion probabilities for clusters, and the response probabilities for elements are bounded. For convenience, we will call the design described in Lemma 4.4 a one-per-stratum design.

Now we consider the regression coefficients constructed with the inverse of the nonresponse probability as weights in the weighted regression. Associated with cluster in \( \xi_{N_{c_i(t)}} \) described in Section 4.2, let

\[ x_i = \left( x_{i1}', x_{i2}', ..., x_{im_i}' \right) \tag{4.91} \]
be the matrix of auxiliary variables, with
\[
x_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijq}),
\]  
(4.92)
the vector of auxiliary information for element \( j \) in cluster \( i \). Let the estimated regression coefficients be
\[
\hat{\beta}(t) = \left( \sum_{i=1}^{N_{e(t)}} \pi_{i(t)}^{-1} \left( \sum_{j=1}^{m_i} x'_{ij} x_{ij} r_{ij} p_{ij}^{-1} \right) \right)^{-1} \left( \sum_{i=1}^{N_{e(t)}} \pi_{i(t)}^{-1} t_i \left( \sum_{j=1}^{m_i} x'_{ij} r_{ij} p_{ij}^{-1} y_{ij} \right) \right),
\]  
(4.93)
and we assume \( \sum_{i,j} x'_{ij} x_{ij} r_{ij} p_{ij}^{-1} \pi_{i(t)}^{-1} \) is nonsingular. Define the population regression coefficients by
\[
\beta(t) = \left( \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} x'_{ij} x_{ij} \right)^{-1} \left( \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} x'_{ij} y_{ij} \right).
\]  
(4.94)
We will also consider more general regression coefficients,
\[
\hat{\eta}(t) = \left( \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} x'_{ij} x_{ij} w_{ij} \right)^{-1} \left( \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} x'_{ij} y_{ij} r_{ij} w_{ij} \right),
\]  
(4.95)
where the weights \( w_{ij} \) are bounded. That is, there exist \( \lambda_3 > 0 \) and \( \lambda_4 > 0 \) such that
\[
0 < \lambda_3 < w_{ij} < \lambda_4.
\]  
(4.96)
We denote the population regression coefficients associated with \( \hat{\eta}(t) \) by
\[
\eta(t) = \left( \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} x'_{ij} x_{ij} w_{ij} \pi_{i(t)} p_{ij} \right)^{-1} \left( \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} x'_{ij} y_{ij} w_{ij} \pi_{i(t)} p_{ij} \right).
\]  
(4.97)
If we choose
\[
w_{ij} = \pi_{i(t)}^{-1} p_{ij}^{-1},
\]  
(4.98)
then $\hat{\eta}(t)$ is the expression for $\hat{\beta}(t)$ and $\eta(t)$ is the expression for $\beta(t)$. Theorem 4.1 gives sufficient conditions for $\hat{\eta}(t)$ to be consistent for $\eta(t)$. If $A = (a_{ij})$, we define the power operation by elementwise operation

$$A^{*\delta} = \left( |a_{ij}|^\delta \right).$$  \hfill (4.99)

**Definition 4.2** Let $\{A_t = (a_{ij(t)})_{m \times n}, t = 1, 2, \ldots \}$ be a sequence of $m$ by $n$ nonsingular matrices whose elements $a_{ij(t)}$ are random variables. Let $A = (a_{ij})_{m \times n}$ be an $m$ by $n$ matrix and let $\{g_t : t = 1, 2, \ldots \}$ be a sequence of positive numbers. Then we say

$$A_t - A = O_p(g_t)$$  \hfill (4.100)

if and only if

$$a_{ij(t)} - a_{ij} = O_p(g_t)$$  \hfill (4.101)

for $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, n$. We say

$$A_t - A = o_p(g_t)$$  \hfill (4.102)

if and only if

$$a_{ij(t)} - a_{ij} = o_p(g_t)$$  \hfill (4.103)

for $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, n$. 

Lemma 4.5 and Lemma 4.6 give some results which will be used in the proof of Theorem 4.1.
Lemma 4.5 (Continuous mapping theorem). Let $Y_t$ be a sequence of real valued $q$-dimensional random vectors, such that

$$\lim_{t \to \infty} Y_t = Y.$$  

If $f : \mathbb{R}^q \to \mathbb{R}^p$ is a continuous function, then

$$\lim_{t \to \infty} [f(Y_t) - f(Y)] = 0. \quad (4.104)$$

Proof. (See Theorem 5.1.4 of Fuller (1976), p. 188)

Lemma 4.6 Let $\{A_t : t = 1, 2, \ldots\}$ be a sequence of nonsingular $n \times n$ matrices whose elements $a_{ij}(t)$ are random variables. If there exist a nonsingular matrix $A$, such that for some $\delta > 0$,

$$A_t - A = O_p \left(t^{-\delta}\right), \quad (4.105)$$

then

$$A_t^{-1} - A^{-1} = O_p \left(t^{-\delta}\right). \quad (4.106)$$

If the $O_p$ in (4.104) is replaced by $o_p$, then (4.107) holds by replacing $O_p$ with $o_p$.

Proof. By Definition 4.2, and (4.105), for each element of $A_t - A$,

$$a_{ij}(t) - a_{ij} = O_p \left(t^{-\delta}\right). \quad (4.107)$$

Denote the $(i, j)$-th element of $A_t^{-1}$ and $A^{-1}$ by $b_{ij}(t)(A_t)$ and $b_{ij}(A)$, respectively. The elements of the inverses are continuous functions of the elements of $A_t$ and $A$, respectively (Theorem 1.48, Rickart, 1960). Therefore, by Lemma 4.7, for $i, j = 1, 2, \ldots, n$,

$$\lim_{t \to \infty} \left(b_{ij}(t) - b_{ij}\right) = 0, \quad (4.108)$$
and $b_{ij(t)} = O_p(1)$. From the relation,

$$A_t^{-1} - A^{-1} = -A^{-1} (A_t - A) A_t^{-1}, \quad (4.109)$$

and (4.107), we have for $i, j = 1, 2, ..., n$,

$$b_{ij(t)} - b_{ij} = O_p \left( t^{-6} \right). \quad (4.110)$$

Thus, (4.106) holds. Similarly, if we replace the $O_p$ in (4.106), then (4.107) will hold using the same argument as for $O_p$, replacing $O_p$ in (4.107) by $o_p$. ■

Now we give conditions for estimated regression coefficients for cluster sampling to be consistent in the presence of nonresponse.

**Theorem 4.1** Let the sequence of finite populations $\{\xi_{N(t)} : t = 1, 2, \ldots\}$ and the associated cluster samples $\{s_{n(t)} : t = 1, 2, \ldots\}$ be as described in Section 4.2. Assume that

(i) The nonresponse behavior can be modeled by Poisson sampling and (.14) holds.

(ii) There exist constants $0 < \lambda_1 < \lambda_2 < 1$ such that for all $t = 1, 2, \ldots, i = 1, 2, \ldots, N_{c(t)}$, and $j = 1, \ldots, m_i$,

$$0 < \lambda_1 < \pi_{i(t)} < \lambda_2 < 1, \quad \lambda_1 < p_{ij} \leq 1, \quad (4.111)$$

$$0 < \lambda_3 < w_{ij} < \lambda_4.$$

We also assume that there exists $m > 0$, such that the total number of elements in any cluster is bounded,

$$m_i \leq m. \quad (4.112)$$
(iii) There exist \( \delta_3 > 0 \) and \( \delta_4 > 0 \) with \( \delta_3^{-1} + \delta_4^{-1} = 1 \) such that

\[
N_{c(t)}^{-1} \sum_{i \neq i'} g_{ii'}^{\delta_3} = O\left(n_{c(t)}^{-\delta_3}\right),
\]

\[
N_{c(t)}^{-1} \sum_{i=1}^{m_i} \sum_{j=1}^{n_i} \left\{ [x_{ij}^* (x_{ij}, y_{ij})] \ast\ast (2\delta_4) \right\} = O(1),
\]

where \( g_{ii'} \) is defined in (4.53), and the exponential operation \( \ast\ast \) is defined in (4.99). It is understood that if \( \delta_3 = 1 \), then the absolute value of every element in matrices \( x_{ij}^* (x_{ij}, y_{ij}) \) is bounded by some constant \( M > 0 \), if \( \delta_4 = 1 \), then condition (4.113) is replaced by

\[
n_{c(t)} g_{ii'} < M
\]

for some \( M > 0 \), where \( t = 1, 2, ..., i = 1, 2, ..., N_{c(t)} \), and \( j = 1, ..., m_i \).

Then

\[
\hat{\eta}(t) - \eta(t) = O_p \left(n_{c(t)}^{-1/2}\right),
\]

where \( \hat{\eta}(t) \) and \( \eta(t) \) are in (4.95) and (4.97), respectively.

**Proof.** Let

\[
\left( \hat{Q}(t), \hat{H}(t) \right) = N_{c(t)}^{-1} \sum_{i=1}^{m_i} \pi_i(t) t_i \left\{ \sum_{j=1}^{m_i} x_{ij}^* x_{ij} P_{ij}^{-1} (x_{ij}^*, y_{ij}^*) \right\},
\]

and

\[
\left( Q(t), H(t) \right) = N_{c(t)}^{-1} \sum_{i=1}^{m_i} \sum_{j=1}^{n_i} x_{ij}^* \left( x_{ij}^*, y_{ij}^* \right),
\]

where

\[
(x_{ij}^*, y_{ij}^*) = P_{ij}^{1/2} \pi_i(t)^{1/2} \omega_i^{1/2} (x_{ij}, y_{ij}).
\]
Then

\[ \eta(t) = Q(t)^{-1} H(t), \]

\[ \eta(t) = Q(t)^{-1} H(t). \]

(4.120)

and each element of \( Q(t) \) and \( H(t) \) is an adjusted Horvitz-Thompson estimator of the form of \( \mu_{adj,HT} \) in (4.33). Denote the \((\ell, k)\)-th element of \( Q(t) \), \((\ell, k = 1, 2, ..., q)\) by

\[ \hat{q}_{\ell k(t)} = \sum_{i=1}^{N_c(t)} \pi_i(t)^{-1} \left\{ \sum_{j=1}^{m_i} \sum_{r_{ij}} P_{ij}^{-1} x_{ij}^* \right\}. \]

(4.121)

If \( \delta_3 > 1 \) in condition (ii), and \( w_{ij} < 4 \),

\[ N_c(t)^{-1} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} \left| x_{ij}^* x_{ij}^* \right|^{2\delta_4} \leq N_c(t)^{-1} \sum_{i=1}^{N_c(t)} \left\{ \sum_{j=1}^{m_i} \left| x_{ij}^* x_{ij}^* P_{ij} \pi(t) W_{ij} \right|^{2\delta_4} \right\} \]

\[ \leq N_c(t)^{-1} \sum_{i=1}^{N_c(t)} \left\{ \sum_{j=1}^{m_i} \left| x_{ij}^* x_{ij}^* \right|^{2\delta_4} 4^{2\delta_4} \right\}. \]

(4.122)

Therefore, from condition (4.114), we have

\[ N_c(t)^{-1} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} \left[ \left( x_{ij}^* x_{ij}^* \right)^{*} \left( 2\delta_4 \right) \right] = O(1). \]

(4.123)

If \( \delta_3 = 1 \), then for \( \ell, k = 1, 2, ..., q \), using condition (iii),

\[ \left| x_{ij}^* x_{ij}^* \right| \leq w_{ij} \left| x_{ij}^* x_{ij}^* \right| \leq 4 M < \infty \]

(4.124)

for some \( M > 0 \). If \( \delta_4 = 1 \),

\[ N_c(t)^{-1} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} \left( x_{ij}^* x_{ij}^* \right)^2 \leq 4 N_c(t)^{-1} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} x_{ij}^2 x_{ij}^2 = O(1). \]

(4.125)
In all cases, conditions of Lemma 4.3 are met with (4.113), and we have

\[ \hat{Q}(t) - Q(t) = O_p \left( n_{e(t)}^{-1/2} \right). \]  

(4.126)

Thus, by Lemma 4.6,

\[ \hat{Q}_1^{-1}(t) - Q_1^{-1}(t) = O_p \left( n_{e(t)}^{-1/2} \right). \]  

(4.127)

Using similar arguments, we have

\[ \hat{H}(t) - H(t) = O_p \left( n_{e(t)}^{-1/2} \right). \]  

(4.128)

Therefore,

\[ \hat{\eta}(t) - \eta(t) = \hat{Q}_1^{-1}(t)\hat{H}(t) - Q_1^{-1}(t)H(t) \]
\[ = \left( \hat{Q}_1^{-1}(t) + O_p \left( n_{e(t)}^{-1/2} \right) \right) \left( H(t) + O_p \left( n_{e(t)}^{-1/2} \right) \right) - Q_1^{-1}(t)H(t) \]
\[ = O_p \left( n_{e(t)}^{-1/2} \right). \]  

(4.129)

We apply the results of Theorem 4.1 to regression coefficient estimators such as \( \hat{\beta}_1(t) \) in (4.93), or the estimator

\[ \hat{\gamma}(t) = \left( \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} X'_{ij} \pi_i^{-1(t)T_{ij}}X_{ij} \right)^{-1} \left( \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} X'_{ij} \pi_i^{-1(t)T_{ij}y_{ij}} \right), \]  

(4.130)
associated with the population regression coefficient matrix,

\[ \gamma(t) = \left( \sum_{i=1}^{N_e(t)} \sum_{j=1}^{m_i} x'_{ij} p_{ij} x_{ij} \right)^{-1} \left( \sum_{i=1}^{N_e(t)} \sum_{j=1}^{m_i} x'_{ij} p_{ij} y_{ij} \right). \]  

(4.131)

The results are given by Corollary 4.1.1.

**Corollary 4.1.1** Under the conditions of Theorem 4.1, we have

\[ \hat{\beta}(t) - \beta(t) = \mathcal{O}_p \left( \frac{1}{n_{e(t)}} \right), \]

(4.132)

\[ \hat{\gamma}(t) - \gamma(t) = \mathcal{O}_p \left( \frac{1}{n_{e(t)}} \right), \]

(4.133)

where \( \hat{\beta}(t), \beta(t), \hat{\gamma}(t), \) and \( \gamma(t) \) are defined in (4.93), (4.94), (4.130), and (4.131), respectively.

**Proof.** Let

\[ w_{ij} = p_{ij}^{-1} \pi_{i(t)}^{-1}, \]

(4.134)

in \( \hat{\eta}(t) \) of (4.95), where \( w_{ij} \) are bounded,

\[ 0 < \lambda_2^{-1} < w_{ij} < \lambda_1^{-2}. \]

(4.135)

Then by Theorem 4.1, (4.132) follows. Let

\[ w_{ij} = \pi_{i(t)}^{-1}, \]

(4.136)

in \( \hat{\eta}(t) \) of (4.95), then

\[ 0 < \lambda_2^{-1} < w_{ij} < \lambda_1^{-1}, \]

and using Theorem 4.1, (4.133) follows. \( \blacksquare \)
Under the conditions of Theorem 4.1, we can obtain a consistent regression estimator of the population mean when the vector of ones, denoted by $J$, is in the space $C(X)$ generated by the column vector of $X$, where

$$X = \left( x_{11}', \ldots, x_{1m_1}', \ldots, x_{N_1} ', \ldots, x_{N_2} m_{N_2} ' \right)'.$$

(4.137)

**Corollary 4.1.2** For the sequence of finite populations and associated samples of Theorem 4.1, assume that conditions (i), (ii), and (iii) of Theorem 4.1 hold. Assume that

$$J \in C(X),$$

(4.138)

where $J$ is a $N(t)$-dimensional vector whose elements are ones and, $C(X)$ is the space generated by the column vectors of $X$. Then

$$\hat{\mu} - \bar{Y}(t) = O_p \left( n_{e(t)}^{-1/2} \right),$$

(4.139)

where

$$\hat{\mu} = \bar{X}_n \hat{\beta}(t),$$

(4.140)

$\hat{\beta}(t)$ is defined in (4.93), and $\bar{Y}(t)$ is the finite population mean.

**Proof** By Corollary 4.1.1 of Theorem 4.1, we have

$$\hat{\beta}(t) - \beta(t) = O_p \left( n_{e(t)}^{-1/2} \right),$$

(4.141)

where $\beta(t)$ is defined in (4.94).
Therefore,

\[ \hat{\mu} - \bar{Y}(t) = \bar{X} (\hat{\beta}(t) - \beta(t)) + \bar{X} \beta(t) - \bar{Y}(t) \]
\[ = O_p \left( n^{-1/2} c(t) \right) + N_{(t)}^{-1} Y' \left( X'X \right)^{-1} X' I y. \tag{4.142} \]

where \( y = (y_{11}, ..., y_{1m_1}, ..., y_{Nc1}, ..., y_{NcNc})' \). Since \( J \in C(X) \), there exists a \( q \)-dimensional vector, \( c \), such that

\[ J = Xc. \tag{4.143} \]

Thus,

\[ J' \left( X'X + X' I \right) = c'X' \left( X'X + X' I \right) = 0, \tag{4.144} \]

and (4.142) becomes

\[ \hat{\mu} - \bar{Y}(t) = O_p \left( n^{-1/2} c(t) \right). \tag{4.145} \]

Condition (4.138) will be trivially satisfied if the first element for each \( x_{ij} \) is one. If we choose a single \( X \) variable that is identically equal to one, then the regression estimator defined in (4.140) becomes the scaled Horvitz-Thompson estimator (SHT),

\[ \hat{\mu}_{SHT} = \left( \sum_{i=1}^{N} \sum_{j=1}^{m_i} \pi_{ij(t)}^{-1} r_{ij} \right)^{-1} \left( \sum_{i=1}^{N} \sum_{j=1}^{m_i} \pi_{ij(t)}^{-1} r_{ij} y_{ij} \right), \tag{4.146} \]

where \( \pi_{ij(t)} \) is the observation probabilities defined as in Definition 4.1.
4.4 Regression Estimator for Nonresponse

In this section, we investigate regression estimators for cluster sampling in the presence of nonresponse. If one estimates the population mean without adjusting for nonresponse, and the respondents are different from the nonrespondents for the measured characteristics, the nonresponse will introduce bias. Using auxiliary information can reduce the bias. The scaled Horvitz-Thompson estimator in (4.146) is an example of a consistent estimator obtained by including the inverse of the response probabilities $p_{ij}^{-1}$ in the weighting for respondents. However, in most cases, $p_{ij}$ is unknown, and we need to estimate $p_{ij}$ in order to correct for nonresponse. For example, if we assume the elements within each cluster have the same response probability, then $p_{ij}$ may be estimated by the response rate within cluster $i$ as $\hat{m}_i m_i^{-1}$, where $\hat{m}_i = \sum_{j=1}^{m_i} r_{ij}$ is the number of respondents.

Another way to reduce the bias and improve the efficiency of the estimator is to use the regression estimator. A regression estimator of $\hat{Y}$ is,

$$\hat{\mu}_{\text{Reg}} = \bar{X}\hat{\gamma}_{(t)}$$  \hspace{1cm} (4.147)

where $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijq})$ is the vector of auxiliary variables, $\hat{\gamma}_{(t)}$ is defined in (4.130). The regression estimator $\hat{\mu}_{\text{Reg}}$ in (4.147) was investigated by Mickey (1959), Isaki and Fuller (1982), Wright (1983) and Fuller, Loughin and Baker (1994). We write the data
in the matrix form:

\[ X = (x'_{11}, \ldots, x'_{1m_1}, \ldots, x'_{N_{e1}}, \ldots, x'_{N_{e m_{N_e}}}) \]  

(4.148)

\[ y = (y_{11}, \ldots, y_{1m_1}, \ldots, y_{N_{e1}}, \ldots, y_{N_{e m_{N_e}}})' \]  

(4.149)

\[ r = (r_{11}, \ldots, r_{1m_1}, \ldots, r_{N_{e1}}, \ldots, r_{N_{e m_{N_e}}})' \]  

(4.150)

and

\[ R = \text{diag}(r), \]  

(4.151)

where \( r_{ij} \) is our response indicator. The observation probabilities and response probabilities are written in the same manner:

\[ \pi^*_{(t)} = \left( \pi_{1(t)} p_{11}, \ldots, \pi_{1(t)} p_{1m_1}, \ldots, \pi_{N_{e(t)}} p_{N_{e(t)}1}, \ldots, \pi_{N_{e(t)}} p_{N_{e(t)} m_{N_e(t)}} \right)', \]  

(4.152)

\[ p = \left( p_{11}, \ldots, p_{1m_1}, \ldots, p_{N_{e1}}, \ldots, p_{N_{e m_{N_e}}} \right)', \]  

(4.153)

\[ \Pi_{(t)}^* = \text{diag} \left( \pi^*_{(t)} \right), \]  

(4.154)

\[ P = \text{diag}(p), \]  

(4.155)

\[ \Pi_{(t)} = \Pi_{(t)}^* P^{-1}, \]  

(4.156)

where \( \pi_{(t)} = \left( \pi_{1(t)}, \ldots, \pi_{N_{e(t)}} \right)' \) is the vector of inclusion probabilities for clusters. Thus, \( \hat{\gamma} \) in (4.130) can be expressed as

\[ \hat{\gamma}_{(t)} = \left( X' \Pi_{(t)}^{-1} RX \right)^{-1} \left( X' \Pi_{(t)}^{-1} Ry \right) \]  

(4.157)
and the $\gamma(t)$ in (4.131) can be written as

$$\gamma(t) = (X'PX)^{-1}(X'Py). \quad (4.158)$$

We denote the residuals of the population regression by

$$a(t) = y - X\gamma(t). \quad (4.159)$$

Our goal is to examine the consistency of the regression estimator in (4.147). When the conditions of Theorem 4.1 are met,

$$\text{plim} (\hat{\mu}_{\text{Reg}} - \bar{Y}) = \text{plim} \left( \bar{X} \left( \hat{\gamma}(t) - \gamma(t) \right) - N^{-1} \sum_{i,j} a_{ij} \right)$$

$$= \text{plim} (-\bar{a}). \quad (4.160)$$

where

$$\bar{a} = N^{-1} \sum_{i=1}^{N(t)} \sum_{j=1}^{m_i} a_{ij}. \quad (4.161)$$

Therefore, $\hat{\mu}_{\text{Reg}}$ is design consistent if

$$\text{plim} \bar{a} = 0. \quad (4.162)$$

Theorem 4.3 gives a necessary and sufficient condition for $\bar{a} = 0$. The sufficiency of this theorem was given in Särndal (1980), Isaki and Fuller (1982), Wright (1983), and Fuller, Loughin, and Baker (1994). Theorem 4.3 has a similarity to Zyskind's theorem (Zyskind (1967)).
Theorem 4.2 Let \( a \) be defined by (4.159), and \( \bar{a} \) by (4.161). A necessary and sufficient condition for

\[
\bar{a} = N^{-1} \sum_{i=1}^{N} a_i = 0,
\]

(4.163)

for any \( y \) is that there exist a \( q \)-dimension vector \( c \) such that

\[
J = PXc
\]

(4.164)

where \( J \) is an \( N \)-dimension column vector whose elements are one, \( P \) is defined in (4.155), and \( X \) is defined in (4.148).

Proof. By (4.158) and (4.159),

\[
a = \left[I - X(X'PX)^{-1}(X'P)\right]y.
\]

(4.165)

Therefore, \( \bar{a} = N^{-1}J'a = 0 \) for any \( y \) if and only if

\[
M = J' \left[I - X(X'PX)^{-1}(X'P)\right] = 0.
\]

(4.166)

Letting \( X^* = P^{1/2}X \),

\[
M = J'P^{-1/2} \left[I - X^*(X^*X^*)^{-1}X^*\right]P^{1/2},
\]

(4.167)

where \( P^{-1/2}P^{-1/2} = P^{-1} \). Note that \( X^*(X^*X^*)^{-1}X^* \) is the projection matrix for the space generated by the column vectors of \( X^* \), where the space is denoted by \( C(X^*) \). That is, for every vector \( x \in C(X^*) \),

\[
X^* \left(X^*X^*\right)^{-1}X^* x = x.
\]

(4.168)
Thus, $M = 0$ is equivalent to

$$P^{-1/2}J \in C(X^*) .$$  \hspace{1cm} (4.169)

Therefore, a necessary and sufficient condition for $M = 0$ is that the vector $P^{-1/2}J$ belongs to $C(X^*)$. Equivalently, there exists a $q$-vector $c$ such that

$$J = PXc .$$  \hspace{1cm} (4.170)

Therefore, a necessary and sufficient condition for $\bar{a} = 0$ for any $y$ is that (4.170) holds for some vector $c$, or equivalently, there exists a $c$, such that

$$\bar{p}_{ij}^{-1} = x_{ij}c ,$$  \hspace{1cm} (4.171)

for all $i = 1, ..., N_c, j = 1, ..., m_i$.

By Theorem 4.2, if $\gamma_{(t)}$ is consistent, and if the auxiliary variables $x$ satisfy the condition (4.171), then the regression estimator $\hat{\mu}_{\text{Reg}}$ in (4.147) will be a consistent estimator for $\bar{Y}_n$ even if the weights in the regression are not proportional to the observation probabilities.

**Corollary 4.2.1** For the sequence of finite populations and associated samples of Theorem 4.1, assume that conditions (i), (ii), and (iii) of Theorem 4.1 hold. If (4.164) holds, then

$$\hat{\mu}_{\text{Reg}} - \bar{Y}_{(t)} = O_p \left( n_{c(t)}^{-1/2} \right) ,$$  \hspace{1cm} (4.172)

where $\hat{\mu}_{\text{Reg}}$ is defined in (4.147) and $\bar{Y}_{(t)}$ is the finite population mean.

**Proof** By Corollary 4.1.1 of Theorem 4.1, we have

$$\hat{\gamma}_{(t)} - \gamma_{(t)} = O_p \left( n_{c(t)}^{-1/2} \right) ,$$  \hspace{1cm} (4.173)
where \( \hat{\gamma}(t) \) and \( \gamma(t) \) are defined in (4.130) and (4.131), respectively. Therefore,

\[
\hat{\mu}_{\text{Reg}} - \hat{Y}(t) = \bar{X} \left( \hat{\gamma}(t) - \gamma(t) \right) + \bar{a} \\
= O_p \left( n_{e(t)}^{-1/2} \right)
\]

(4.174)

Because \( \bar{a} = 0 \) for every \( y \).

Under the conditions in Theorem 4.1, \( \hat{\mu}_{\text{Reg}} \) will be a consistent estimator of \( \hat{Y} \) if the auxiliary variable \( x \) is such that they explain the nonresponse behavior in an "inverse linear" manner as described in (4.171). We will discuss nonresponse adjustment in the next section.

4.5 Regression Estimator with Nonresponse Adjustment

In Section 4.4, we discussed the consistency of the regression estimator \( \hat{\mu}_{\text{Reg}} \) of (4.147). In this section, we will present results for regression estimators with a nonresponse adjustment in the weights. Since we do not know the response probabilities, \( p_{ij} \), we use estimated \( p_{ij} \) to construct the regression estimator.

We assume that the response probabilities \( p_{ij} \) depend on some auxiliary variables. Post-stratification is an example. In post-stratification, based on the response pattern, we divide the population into different categories such that within each category, the response probabilities for elements are assumed to be the same. That is, the response probabilities satisfy
a classification model with \( L \) classification levels.

\[
p_{ij} = c_{ij} \theta \tag{4.175}
\]

for \( i = 1, 2, \ldots, N_c \) and \( j = 1, \ldots, m_i \), where \( c_{ij} \) is the \( L \)-dimensional vector whose elements are zeros and ones indicating the category of element \((i,j)\), and \( \theta = (\theta_1, \ldots, \theta_L)' \) is the response probability vector for \( L \) categories. It is possible to estimate the response probabilities within each category and reduce the bias by using these estimated probabilities in the regression estimator. See, for example, Little (1986).

Other estimation procedures can be used to estimate the response probabilities. Folsom and Witt (1994) used a logistic regression model for response probabilities. In the logistic model,

\[
p_{ij} = \{1 + \exp(-x_{ij} \theta)\}^{-1}. \tag{4.176}
\]

Now we examine the effect of using estimated response probabilities in a regression estimator. We assume that response probabilities are estimated by estimating the unknown model parameters of a true model. Model parameters are estimated from auxiliary information of both respondents and nonrespondents. That is, if cluster \( i \) is selected, we assume that \( x_{ij} \) is available for the estimation of \( p_{ij} \) for all \( j = 1, 2, \ldots, m_i \). For example, for a human survey, the auxiliary variables might be age, gender, and education. If we are interested in income, but some individuals refuse to report their income, we assume we are still able to obtain the auxiliary information for these individuals.
We assume the response probability \( p_{ij} \) is a function of auxiliary variables:

\[
 p_{ij} = f(x_{ij}, \theta) \tag{4.177}
\]

for \( i = 1, \ldots, N_c \), and \( j = 1, \ldots, m_i \), where \( \theta \) is a \( q_0 \)-dimensional vector of unknown parameters and \( f \) is a known function. Denote the estimator of \( \theta \) in (4.177) by \( \hat{\theta} \), then \( p_{ij} \) is estimated by

\[
 \hat{p}_{ij} = f(x_{ij}, \hat{\theta}). \tag{4.178}
\]

If we include the inverse of the estimated response probability \( \hat{p}_{ij}^{-1} \) as an additional regressor, we denote the new auxiliary variables by

\[
 (x_{ij}, \hat{p}_{ij}^{-1}) \equiv x_{ij}^{**}. \tag{4.179}
\]

If we use the new auxiliary variables \( x_{ij}^{**} \) to perform the regression, then we obtain the regression estimator,

\[
 \hat{\mu}^{**} = X^{**}\hat{\beta}^{**}, \tag{4.180}
\]

where

\[
 \hat{\beta}^{**} = (X^{**}\Pi_{(i)}^{-1}RX^{**})^{-1}(X^{**}\Pi_{(i)}^{-1}Ry), \tag{4.181}
\]

\[
 X^{**} = (X, \hat{p}^{-1}), \tag{4.182}
\]

\[
 \hat{p}^{-1} = (\hat{p}_{11}^{-1}, \ldots, \hat{p}_{1,m_1}^{-1}, \ldots, \hat{p}_{N_c,1}^{-1}, \ldots, \hat{p}_{N_c,m_{N_c}}^{-1}), \tag{4.183}
\]

\[
 \bar{X}^{**} = (\bar{X}, \bar{p}^{-1}), \tag{4.184}
\]

\[
 \bar{p}^{-1} = N_{(i)}^{-1} \sum_{i,j} \hat{p}_{ij}^{-2}r_{ij} \pi_i^{-1}, \tag{4.185}
\]
X, y, R, and \( \Pi_{(t)} \) are defined in (4.148), (4.149), (4.151), and (4.156), respectively. Note that we write matrices for the entire finite population. In practice, the actual matrices used in the regression are for respondents only.

Theorem 4.4 states conditions for the adjusted regression estimator \( \hat{\mu}^{*\ast} \) in (4.180) to be consistent.

**Theorem 4.4** Let the sequence of finite populations \( \{\xi_{N(t)}, \ t = 1, 2, \ldots\} \) and the associated sequence of cluster samples \( \{s_{n(t)}, \ t = 1, 2, \ldots\} \) be as described in Section 4.2. Assume that

(i) Respondents in each selected cluster consist of a Poisson sample such that (4.14) holds.

(ii) There exist constants \( \lambda_1 > 0, \lambda_2 > 0, \) and \( m > 0, \) such that

\[
0 < \lambda_1 < \pi_i(t) < \lambda_2 < 1, \quad 0 < \lambda_1 < p_{ij} \leq 1.
\]

(iii) There exist \( \delta_3 > 0 \) and \( \delta_4 > 0 \) such that \( \delta_3^{-1} + \delta_4^{-1} = 1 \) and

\[
N_{c(t)}^{-1} \sum_{i \neq i'} N_{c(t)} \delta_{ii'} = O \left( n_{c(t)}^{-\delta_3} \right) \quad \text{(4.187)}
\]

\[
N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \left\{ [x_{ij} (x_{ij}, y_{ij})] * (2\delta_4) \right\} = O(1),
\]

where \( g_{ii'}(t) \) is defined in (4.53). It is understood that if \( \delta_3 = 1, \) then the absolute value of every element in the matrix \( x_{ij} (x_{ij}, y_{ij}) \) is bounded by some constant \( M > 0, \) if \( \delta_4 = 1, \)
then condition (4.187) is replaced by

\[ n_{c(t)} g_{it}(t) < M \]  

(4.189)

for some \( M > 0 \), \( t = 1, 2, \ldots \), \( i = 1, 2, \ldots \), \( N_{c(t)} \), and \( j = 1, \ldots, m_i \).

(iv) The response probabilities satisfy

\[ p = f(x, \theta), \]  

(4.190)

where \( \theta \) is an unknown \( q_0 \)-dimensional parameter vector and \( f : R^q \times R^\infty \to R \). For every \( \epsilon > 0 \), there exists a \( \delta > 0 \), such that if \( |\theta - \theta^{(0)}| < \delta \), then \( |f(x, \theta) - f(x, \theta^{(0)})| < \epsilon \) for every \( x \). We also assume

\[ 0 < \lambda_1 < f(x, \theta) \leq 1. \]  

(4.191)

(v) Suppose there is a true parameter \( \theta^{(0)} \) and an estimator \( \hat{\theta} \) for \( \theta^{(0)} \) such that,

\[ \hat{\theta}_{(t)} - \theta^{(0)} = o_p(1), \]  

(4.192)

Then

\[ \text{plim}_{t \to \infty} \left( \hat{\mu}^{**} - \hat{Y}_{(t)} | \xi_{N_{c(t)}} \right) = 0, \]  

(4.193)

where \( \hat{\mu}^{**} \) is in (4.180), \( \hat{p}_{ij} = f(x_{ij}, \hat{\theta}_{(t)}) \), and \( \hat{Y}_{(t)} \) is the finite population mean of \( \xi_{N_{c(t)}} \).

Proof. Let

\[ \left( \hat{Q}^{**}, \hat{H}^{**} \right) = N_{(t)}^{-1} X^{**} \Pi^{-1} R(X^{**}, y) \]  

(4.194)
\[
= \left( N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} \pi_{i(t)}^{-1} \sum_{j=1}^{m_i} \left( \frac{\mathbf{x}_{ij}}{\hat{p}_{ij}^{-1}} \right) r_{ij} \left\{ \left( (\mathbf{x}_{ij}, \hat{p}_{ij}^{-1}), y_{ij} \right) \right\} \right)
\]

and

\[
\left( \hat{\mathbf{Q}}, \hat{\mathbf{H}} \right) = \left( N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} \pi_{i(t)}^{-1} \sum_{j=1}^{m_i} \left( \frac{\mathbf{x}_{ij}}{\hat{p}_{ij}^{-1}} \right) r_{ij} \left\{ \left( (\mathbf{x}_{ij}, \hat{p}_{ij}^{-1}), y_{ij} \right) \right\} \right)
\]

(4.195)

First, we show that

\[
\left( \hat{\mathbf{Q}}^{**}, \hat{\mathbf{H}}^{**} \right) - \left( \hat{\mathbf{Q}}, \hat{\mathbf{H}} \right) = \alpha_p (1).
\]

(4.196)

Let

\[
\mathbf{p}^{-1} = \left( p_{11}^{-1}, ..., p_{m_1}^{-1}, ..., p_{N_{(t)} m_{N_{(t)}}}^{-1}, ..., p_{N_{(t)} m_{N_{(t)}}}^{-1} \right),
\]

(4.197)

and let \( \Delta \mathbf{p}^{-1} = \hat{\mathbf{p}}^{-1} - \mathbf{p}^{-1} \), where \( \hat{\mathbf{p}}^{-1} \) is defined in (4.183), then

\[
\hat{\mathbf{Q}}^{**} = \left( N_{(t)}^{-1} \right) \left( \begin{array}{c} \mathbf{X}' \\ \mathbf{p}^{-1/2} + \Delta \mathbf{p}^{-1/2} \end{array} \right) \Pi^{-1} \mathbf{R} \left( \mathbf{X}, \mathbf{p}^{-1} + \Delta \mathbf{p}^{-1} \right)
\]

\[
= \hat{\mathbf{Q}} + \left( \begin{array}{c} \mathbf{0} \\ (0, b) \end{array} \right) + \left( \begin{array}{c} (0, 0) \end{array} \right).
\]

(4.198)

where

\[
\mathbf{b} = \left( N_{(t)}^{-1} \right) \left( \begin{array}{c} \mathbf{X}' \\ \mathbf{p}^{-1/2} \end{array} \right) \Pi^{-1} \mathbf{R} \Delta \mathbf{p}^{-1}
\]

\[
= N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} \sum_{j=1}^{m_i} \left( \frac{\mathbf{x}_{ij}'}{\hat{p}_{ij}^{-1}} \right) \pi_{i(t)}^{-1} r_{ij} \left( \hat{p}_{ij}^{-1} - p_{ij}^{-1} \right),
\]

(4.199)

\[
d = N_{(t)}^{-1} \Delta \mathbf{p}^{-1/2} \Pi \mathbf{R} \Delta \mathbf{p}^{-1}
\]
\[ \hat{Q} - \hat{Q} = o_p(1). \]  

Therefore, it is enough to show that \(b = o_p(1)\) and \(d = o_p(1)\) in order to conclude that \(\hat{Q} - \hat{Q} = o_p(1)\).

**Proof of \(d = o_p(1)\)** By condition (iv), since \(f\) is bounded, given \(\delta > 0\), let \(\delta_1 = \sqrt{\lambda_1 \delta}\). then there exists \(\epsilon_1 > 0\), such that for \(\hat{\theta}_{(t)}\), \(\theta^{(0)}\) satisfying
\[
|\hat{\theta}_{(t)} - \theta^{(0)}| < \epsilon_1, 
\]
then
\[
|\hat{p}_{ij}^{-1} - p_{ij}^{-1}| = \left| \left[ f \left( x_{ij}, \hat{\theta}_{(t)} \right) \right]^{-1} - \left[ f \left( x_{ij}, \theta^{(0)} \right) \right]^{-1} \right| < \delta_1 
\]
for \(i = 1, 2, ..., N_{\epsilon(t)}\), and \(j = 1, ..., m_i\).

By (4.192), for a given \(\epsilon > 0\), there exist \(t_0\), such that for all \(t > t_0\),
\[
\Pr \left( |\hat{\theta}_{(t)} - \theta^{(0)}| > \frac{1}{2} \epsilon_1, \xi_{N_{\epsilon(t)}} \right) < \epsilon. 
\]
From (4.202), for all \(t > t_0\),
\[
\Pr \left( |\hat{p}_{ij}^{-1} - p_{ij}^{-1}| < \delta \text{ for } i = 1 < 2, ..., N_{\epsilon(t)}, \right.
\]
and \(j = 1, 2, ..., m_i | \xi_{N_{\epsilon(t)}} \)
\[
\geq \Pr \left( |\hat{\theta}_{(t)} - \theta^{(0)}| < \epsilon_1 | \xi_{N_{\epsilon(t)}} \right) 
\]
\[
> 1 - \epsilon. 
\]

From condition (ii), if for \(i = 1, 2, ..., N_{\epsilon(t)}\) and \(j = 1, 2, ..., m,\)
\[
|\hat{p}_{ij}^{-1} - p_{ij}^{-1}| < \delta_1, 
\]
then

$$|d| \leq N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \lambda_i^{-1} \delta_i^2 = \lambda_1^{-1} \delta_1^2. \quad (4.206)$$

Thus,

$$Pr \left( |d| > \delta |\xi_{N_{c(t)}} \right) = \epsilon. \quad (4.207)$$

Therefore, we have

$$d = o_p(1). \quad (4.208)$$

Proof of $b = o_p(1)$ The proof of $b = o_p(1)$ is similar to the proof of $d = o_p(1)$. For the $\ell$-th ($\ell = 1, 2, ..., q$) element of $b$, using the Hölder inequality and the Jensen inequality,

$$|b_\ell| = \left| \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \pi_{i(t)}^{-1} r_{ij} x_{ij\ell} \left( \hat{p}_{ij}^{-1} - p_{ij}^{-1} \right) \right|$$

$$\leq \left( \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x_{ij\ell}^2 \right)^{1/2} \left( \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \pi_{i(t)}^{-1} r_{ij} \left( \hat{p}_{ij}^{-1} - p_{ij}^{-1} \right)^2 \right)^{1/2}$$

$$\leq \left( \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x_{ij\ell}^2 \right)^{1/4} \lambda_1^{-1/2} \left( \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \pi_{i(t)}^{-1} r_{ij} \left( \hat{p}_{ij}^{-1} - p_{ij}^{-1} \right)^2 \right)^{1/2}$$

$$= \left( \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x_{ij\ell}^2 \right)^{1/4} \lambda_1^{-1/2} d^{1/2}. \quad (4.209)$$

Hence, by condition (iii), since $\delta_4 \geq 1$, there exist $M > 0$, such that

$$\sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x_{ijk}^4 < M,$$

and when (4.202) holds, using (4.206),

$$|b_\ell| \leq M^{1/4} \lambda_1^{-1} \delta_1. \quad (4.210)$$
For the \((q + 1)\)-th element of \(b\),
\[
|b_{q+1}| = \left| \sum_{i=1}^{N_{(t)}} \sum_{j=1}^{m_{i}} \pi_{ij}^{-1} r_{ij} \left( \hat{p}_{ij}^{-1} - p_{ij}^{-1} \right) \right|
\leq \lambda_{i}^{-3/2} \left( \sum_{i=1}^{N_{(t)}} \sum_{j=1}^{m_{i}} r_{ij} \pi_{ij}^{-1} \left( \hat{p}_{ij}^{-1} - p_{ij}^{-1} \right)^{2} \right)^{1/2}
\leq \lambda_{i}^{-3/2} d_{1/2}.
\]

Therefore, when (4.202) holds, using (4.206),
\[
|b_{q+1}| \leq \lambda_{i}^{-2} \delta_{i}.
\]  \hspace{1cm} (4.212)

Thus, by arguments used to prove \(d = o_{p}(1)\), we have
\[
b = o_{p}(1),
\]  \hspace{1cm} (4.213)

and, by (4.208), we have \(\hat{Q}^{**} - \hat{Q} = o_{p}(1)\).

We now show that \(\hat{H}^{**} - \hat{H} = o_{p}(1)\). Using
\[
\hat{H}^{**} - \hat{H} = N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} \sum_{j=1}^{m_{i}} \left( \begin{array}{c} 0 \\ \hat{p}_{ij}^{-1} - p_{ij}^{-1} \end{array} \right) \pi_{ij}^{-1} r_{ij} y_{ij} \equiv \left( \begin{array}{l} 0 \\ h \end{array} \right),
\]  \hspace{1cm} (4.214)

and condition (iii),
\[
h = N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} \sum_{j=1}^{m_{i}} \pi_{ij}^{-1} r_{ij} y_{ij} \left( \hat{p}_{ij}^{-1} - p_{ij}^{-1} \right) = o_{p}(1).
\]  \hspace{1cm} (4.215)

Hence,
\[
\hat{H}^{**} - \hat{H} = o_{p}(1),
\]  \hspace{1cm} (4.216)

and (4.196) is true.
Using Lemma 4.3,

$$\left( \hat{Q} - Q, \hat{H} - H \right) = O_p \left( n_{e(t)}^{-1/2} \right),$$

where

$$\left( Q, H \right) = N_{e(t)}^{-1} \sum_{i=1}^{N_{e(t)}} \sum_{j=1}^{m_i} p_{ij} \left( x'_{ij} \right) \left( \left( x_{ij}, p_{ij}^{-1} \right), y_{ij} \right),$$

and

$$\left( \hat{Q}, \hat{H} \right) = O_p (1).$$

Also, by the absolute continuity of $f$, using arguments similar to those in the proof of $d = o_p (1)$, we have

$$\overline{p^{-1}} - N_{e(t)}^{-1} \sum_{i,j} p_{ij}^{-2} r_{ij} \pi_{i(t)}^{-1} = o_p (1).$$

Using Lemma 4.3,

$$N_{e(t)}^{-1} \sum_{i,j} p_{ij}^{-2} r_{ij} \pi_{i(t)}^{-1} - \overline{p^{-1}} = o_p \left( n_{e(t)}^{-1/2} \right),$$

where $\overline{p^{-1}} = N_{e(t)}^{-1} \sum_{i,j} p_{ij}^{-1}$ is the finite population mean of $p_{ij}^{-1}$. Therefore, $\overline{p^{-1}}$ is consistent for $\overline{p^{-1}}$,

$$\overline{p^{-1}} - \overline{p^{-1}} = o_p (1).$$

Writing $\hat{\mu}^{**}$ in (4.180) as:

$$\hat{\mu}^{**} = (\overline{X}, \overline{p^{-1}}) \hat{Q}^{***} \hat{H}^{**}$$

$$= (\overline{X}, \overline{p^{-1}}) \hat{Q}^{***} \hat{H}^{**} + (0, \overline{p^{-1}} - \overline{p^{-1}}) \hat{Q}^{***} \hat{H}^{**}.$$
By Corollary 4.3.1, we have
\[
(\tilde{X}, \tilde{p}^{-1}) \tilde{Q}^{-1}\tilde{H} - \tilde{Y} = o_p\left(n_{e(t)}^{-1/2}\right). \tag{4.224}
\]

Thus, using (4.196), (4.219), (4.222), and (4.223),
\[
\hat{\mu}^{**} - \tilde{Y}_{(t)} = (\tilde{X}, \tilde{p}^{-1}) \tilde{Q}^{-1}\tilde{H} - \tilde{Y}_{(t)} + o_p(1)
\]
\[
= o_p(1). \tag{4.225}
\]

Therefore, \(\hat{\mu}^{**}\) is a design consistent estimator for \(\tilde{Y}_{(t)}\).

Another procedure that incorporates estimated response probabilities into the estimation is to use the inverses of the estimated response probabilities as weights in the regression. Let
\[
\hat{\m}^{-1} = \text{diag} \left(\hat{\m}^{-1}\right), \tag{4.226}
\]
where \(\hat{\m}^{-1}\) is defined in (4.183), then a weighted regression estimator, using the inverse of \(\hat{\m}_{ij}\), is
\[
\hat{\mu}_{\text{weight}} = \tilde{X}'\hat{\beta}_{\text{weight}}, \tag{4.227}
\]
where
\[
\hat{\beta}_{\text{weight}} = \left(X'\Pi_{(t)}^{-1}R\hat{\m}^{-1}X\right)^{-1}\left(X'\Pi_{(t)}^{-1}R\hat{\m}^{-1}y\right) \tag{4.228}
\]
and \(X, y, R, \Pi_{(t)}\) are defined in (4.148), (4.149), (4.151), and (4.156), respectively. Theorem 4.5 gives conditions for estimator \(\hat{\mu}_{\text{weight}}\) to be design consistent.
Theorem 4.5  For the sequence of finite populations and associated cluster samples of
Theorem 4.4, assume that conditions (i) - (v) of Theorem 4.4 hold. Assume

\[ \mathbf{J} \in \mathbb{C}(\mathbf{X}). \]

where \( \mathbf{J} \) is the vector whose elements are all ones, and \( \mathbb{C}(\mathbf{X}) \) is the space generated by
the column vectors of \( \mathbf{X} \). Then

\[ \lim_{t \to \infty} \left( \hat{\mu}_{p\text{-weight}} - \tilde{Y}_{(t)} | \xi_{N_{(t)}} \right) = 0. \]

Proof. Let

\[ \left( \hat{\mathbf{A}}, \hat{\mathbf{B}} \right) = \mathbf{X}^\prime \mathbf{B}_t^{-1} \mathbf{R} \hat{\mathbf{F}}^{-1} (\mathbf{X}, \mathbf{y}) \]

and

\[ \left( \hat{\mathbf{A}}, \hat{\mathbf{B}} \right) = N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} \sum_{j=1}^{m_i} \pi_{i(t)}^{-1} \mathbf{X}_i \hat{\mathbf{p}}_{ij}^{-1} \mathbf{r}_{ij} (\mathbf{x}_{ij}, \mathbf{y}_{ij}). \]

then

\[ \left( \hat{\mathbf{A}}, \hat{\mathbf{B}} \right) - \left( \mathbf{A}, \mathbf{B} \right) = N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} \sum_{j=1}^{m_i} \pi_{i(t)}^{-1} \mathbf{X}_i \Delta \hat{\mathbf{p}}_{ij}^{-1} \mathbf{r}_{ij} (\mathbf{x}_{ij}, \mathbf{y}_{ij}) \]

\[ \equiv (\Delta \mathbf{A}, \Delta \mathbf{B}), \]

where \( \Delta \hat{p}_{ij}^{-1} = \hat{p}_{ij}^{-1} - p_{ij}^{-1} \). For the \((\ell, k)\)-th element of \( \Delta \mathbf{A} \),

\[ \Delta a_{\ell k} = N_{(t)}^{-1} \sum_{i=1}^{N_{(t)}} \sum_{j=1}^{m_i} \pi_{i(t)}^{-1} x_{ij} \ell x_{ij} \Delta \hat{p}_{ij}^{-1}. \]
Using the Hölder inequality, and \( \pi_{\ell(t)} < \lambda_1^{-1} \),

\[
|\Delta a_{\ell k}| \leq \left( N_{c(t)}^{-1} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} x_{ij}^2 \right)^{1/2} \left( \lambda_1^{-1} N_{c(t)}^{-1} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} \pi_{\ell(t)}^{-1} r_{ij} \left( \Delta p_{ij}^{1/2} \right)^2 \right)^{1/2} \\
= \left( N_{c(t)}^{-1} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} x_{ij}^2 \right)^{1/2} \lambda_1^{-1/2} d^{1/2},
\]

(4.235)

where \( d \) is defined in (4.200). Therefore, by condition (iii) of Theorem 4.4 and \( \delta_4 \geq 1 \), there exists an \( M > 0 \), such that for all \( (\ell, k), \)

\[
N_{c(t)}^{-1} \sum_{i=1}^{N_c(t)} \sum_{j=1}^{m_i} x_{ij}^2 \ll M.
\]

(4.236)

Thus,

\[
|\Delta a_{\ell k}| \leq M^{1/2} \lambda_1^{-1/2} d^{1/2}.
\]

(4.237)

Using similar arguments to those used in the proof of \( d = o_p(1) \) in Theorem 4.4, we have for \( \ell, k = 1, ..., q, \)

\[
\Delta a_{\ell k} = o_p(1)
\]

(4.238)

Similarly, using the condition in (iii) of Theorem 4.3, for each element of \( \Delta B \),

\[
|\Delta b_{\ell t}| \leq M^{1/2} \lambda_1^{-1/2} d^{1/2},
\]

(4.239)

and

\[
\Delta b_{\ell t} = o_p(1)
\]

(4.240)

for \( \ell = 1, ..., q \). Therefore, from (4.238) and (4.240), we have

\[
(\tilde{A}, \tilde{B}) - (\hat{A}, \hat{B}) = o_p(1),
\]

(4.241)
and by Lemma 3.2,

\[ \widehat{\mathbf{A}}^{-1} - \mathbf{A}^{-1} = o_p(1). \]  

(4.242)

Using (4.229) and Theorem 4.2, we have

\[ \mathbf{X} \widehat{\mathbf{A}}^{-1} \mathbf{B} - \bar{Y}(t) = O_p \left( n_{\epsilon(t)}^{-1/2} \right). \]  

(4.243)

By Corollary 4.3.1 of Lemma 4.3,

\[ \widehat{\mathbf{A}} - N_{\epsilon(t)}^{-1} \sum_{i=1}^{N_{\epsilon(t)}} \sum_{j=1}^{m_i} \mathbf{x}_{ij} \mathbf{x}_{ij} = O_p \left( n_{\epsilon(t)}^{-1/2} \right), \]  

(4.244)

\[ \mathbf{B} - N_{\epsilon(t)}^{-1} \sum_{i=1}^{N_{\epsilon(t)}} \sum_{j=1}^{m_i} \mathbf{x}_{ij} \mathbf{y}_{ij} = O_p \left( n_{\epsilon(t)}^{-1/2} \right), \]  

(4.245)

thus, using conditions (iii) of Theorem 4.4, and the Hölder inequality,

\[ (\mathbf{A}, \mathbf{B}) = O_p(1). \]  

(4.246)

Thus,

\[ \mu_{p\text{-weight}} - \bar{Y}(t) = \mathbf{X} \widehat{\mathbf{A}}^{-1} \mathbf{B} - \bar{Y}(t) \]

\[ = \mathbf{X} \left( \mathbf{A}^{-1} + \widehat{\mathbf{A}}^{-1} - \mathbf{A}^{-1} \right) \left( \mathbf{B} + \mathbf{B} - \mathbf{B} \right) - \bar{Y}(t) \]

\[ = \mathbf{X} \left( \mathbf{A}^{-1} + o_p(1) \right) \left( \mathbf{B} + o_p(1) \right) - \bar{Y}(t) \]

\[ = \mathbf{X} \mathbf{A}^{-1} \mathbf{B} - \bar{Y}(t) + o_p(1) \]

\[ = o_p(1). \]  

(4.247)

That is, \( \mu_{p\text{-weight}} \) is design consistent for \( \bar{Y}(t) \).
Estimators $\hat{\mu}^{**}$ in (4.180) and $\hat{\mu}_{p{\text{--weight}}}$ in (4.227) are both regression estimators which incorporate the estimated response probabilities into the estimation of $\bar{Y}_{(t)}$ in order to obtain the consistency. Estimator $\hat{\mu}^{**}$ uses the inverse of the estimated response probabilities as a new control variable while $\hat{\mu}_{p{\text{--weight}}}$ uses the inverse of estimated response probabilities as regression weights. If the control variables include intercept, both estimators will yield consistent estimators given the conditions in Theorem 4.5. If the control variables do not include intercept, then under the conditions of Theorem 4.5, $\hat{\mu}^{**}$ will be consistent and $\hat{\mu}_{p{\text{--weight}}}$ may not be consistent. Therefore, when condition (4.229) holds, $\hat{\mu}_{p{\text{--weight}}}$ may be chosen due to the simplicity of computation comparing to $\hat{\mu}^{**}$, otherwise, a safe choice for an consistency estimator is by using $\hat{\mu}^{**}$ which does not require the condition (4.229).

4.6 Variance of Regression Estimator

In previous sections, we discussed several regression estimators associated with the non-response behavior. In this section, we will evaluate the error of the regression estimator and give the variance expression for an approximation to the error. We give the result in Theorem 4.5.

**Theorem 4.5** Let the sequence of finite populations $\{\xi_{N(t)}^{(t)}, \ t = 1, 2, \ldots\}$ and the associated sequence of cluster samples $\{s_{n(t)}, \ t = 1, 2, \ldots\}$ be as described in Section 4.2. Let

$$\hat{p}_{ij}^{-1} = f (x_{ij}, \hat{\theta}_{(t)})$$

(4.248)
be an estimator of the inverse of the response probabilities. Let the regression coefficient estimator be

$$\hat{\beta} = \left( X' \Pi_{(t)}^{-1} \hat{P}^{-1} R X \right)^{-1} \left( X' \Pi_{(t)}^{-1} \hat{P}^{-1} R y \right),$$

(4.249)

where $X$, $y$, $\Pi_{(t)}$, $\hat{P}^{-1}$, and $R$ are defined by (4.148), (4.149), (4.156), (4.226), and (4.151), respectively.

Let the finite population regression coefficients be

$$\beta = (X'X)^{-1} X'y,$$

(4.250)

and the population residuals be

$$a = y - X\beta.$$  

(4.251)

Assume that:

(i) Respondents in each selected cluster consist of a Poisson sample such that (4.14) holds.

(ii) There exist constants $\lambda_1 > 0$, $\lambda_2 > 0$, and $m > 0$ such that

$$0 < \lambda_1 < \pi_{i(t)} < \lambda_2 < 1, \quad 0 < \lambda_1 < p_{ij} \leq 1, \quad 0 < \lambda_1 < p_{ij} \leq 1,$$

(4.252)

and $m_i \leq m$.

(iii) There exist $\delta_3 > 0$ and $\delta_4 > 0$, such that $\delta_3^{-1} + \delta_4^{-1} = 1$, and

$$N_{c(t)}^{-2} \sum_{i \neq i'} g_{ii'(t)} = O \left( n_{c(t)}^{-\delta_3} \right),$$

(4.253)

$$N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \left( x_{ij} \left| y_{ij}, \alpha_{ij} \right. \right) \ast \ast (2\delta_4) = O (1),$$

(4.254)
where the exponential operation ** is defined in (4.100).

\[ g_{ii'}(t) = \left( \pi_{i}(t)\pi_{i'}(t) - \pi_{ii'}(t) \right) I_{\{\pi_{i}(t)\pi_{i'}(t) - \pi_{ii'}(t) > 0\}} \]  

(4.255)

and \( a_{ij} \) are elements of the residual vector \( a \) in (4.251). It is understood that if \( \delta_3 = 1 \), then the absolute value of every element in the matrix

\[ \left( x'_{ij} (x_{ij}, y_{ij}) \right)^* \tag{4.256} \]

is bounded by some constant \( M > 0 \). If \( \delta_4 = 1 \), then condition (4.253) is replaced by

\[ n_{c(t)} g_{ii'}(t) < M \]  

(4.257)

for some \( M > 0, t = 1, 2, ..., i = 1, 2, ..., N_{c(t)}, \) and \( j = 1, ..., m_i \). We also assume that

\[ \lim_{t \to \infty} N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x'_{ij} (x_{ij}, y_{ij}) = (Q, H), \]  

(4.258)

and \( Q \) is positive definite.

(iv) The inverses of the response probabilities satisfy

\[ p_{ij}^{-1} = f \left( x_{ij}, \theta^{(0)} \right), \]  

(4.259)

where \( \theta^{(0)} \) is the unknown \( q_0 \)-dimensional parameter vector. The function \( f (x, \theta), f : \nabla^{q \times R^{q_0} \to [1, \lambda^{-1}]} \) has continuous first and second derivatives.

(v) The estimator \( \hat{\theta}_{(t)} \) is such that

\[ \hat{\theta}_{(t)} - \theta^{(0)} = O_p \left( n_{c(t)}^{-1/2} \right) \]  

(4.260)
(vi) For any \( \theta \) in an open set containing \( \theta^{(0)} \) and for \( \ell = 1, 2, \ldots, q_0 \),

\[
N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x'_{ij} a_{ij} = 0, \quad (4.261)
\]

\[
N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x'_{ij} a_{ij} p_{ij} \frac{\partial f (x_{ij}, \theta)}{\partial \theta} = o(1). \quad (4.262)
\]

Then,

\[
n^{1/2} (\hat{\beta} - \beta) = n_{c(t)}^{1/2} Q_{N_{c(t)}}^{-1} \hat{H} + o_p (1) \quad (4.263)
\]

with

\[
V \left( n_{c(t)}^{1/2} Q_{N_{c(t)}}^{-1} \hat{H} \mid \xi_{N_{c(t)}} \right) = n_{c(t)} Q_{N_{c(t)}}^{-1} V_H Q_{N_{c(t)}}^{-1} \quad (4.264)
\]

where

\[
\hat{H} = N_{c(t)}^{-1} \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} x'_{ij} a_{ij} \pi^{-1}_{i(t)} p^{-1}_{i(t)} r_{ij}, \quad (4.265)
\]

\[
Q_{N_{c(t)}} = N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x'_{ij} x_{ij}, \quad (4.266)
\]

\[
V_H = N_{c(t)}^{-2} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \pi^{-1}_{i(t)} \left( p^{-1}_{i(t)} - 1 \right) h_{ij} h'_{ij}
+ 0.5 N_{c(t)}^{-2} \sum_{i \neq i'} \left[ (\pi_{i(t)} \pi_{i'(t)} - \pi_{i'(t)}) (h_{i.} - h_{i'}) (h_{i'} - h_{i'}) \right], \quad (4.267)
\]

\[
h_{ij} = a_{ij} x'_{ij}, \quad (4.268)
\]

\[
h_{i.} = \pi^{-1}_{i(t)} \sum_{j=1}^{m_i} h_{ij} = \sum_{j=1}^{m_i} a_{ij} x'_{ij} \pi^{-1}_{i(t)}, \quad (4.269)
\]

and \( \pi_{i(t)} \) is the joint inclusion probability for cluster \( i \) and \( i' \).

Proof. We take the Taylor expansion of \( \hat{p}_{ij}^{-1} \):

\[
\hat{p}_{ij}^{-1} - p_{ij}^{-1} = \left( \frac{\partial f (x_{ij}, \theta^{(1)})}{\partial \theta} \right)' (\hat{\theta}_{(t)} - \theta^{(0)}) \quad (4.270)
\]
\[ \hat{p}_{ij}^{-1} - p_{ij}^{-1} = \left( \frac{\partial f(x_{ij}, \theta^{(0)})}{\partial \theta} \right)' (\hat{\theta}(t) - \theta^{(0)}) + (\hat{\theta}(t) - \theta^{(0)})' \left( \frac{\partial f(x_{ij}, \theta^{(2)})}{\partial \theta \partial \theta'} \right) (\hat{\theta}(t) - \theta^{(0)}) , \] (4.271)

where \( \theta^{(1)} \) and \( \theta^{(2)} \) are two vectors in an open set containing \( \theta^{(0)} \).

Let

\[
\hat{Q} = N_{c(t)}^{-1} \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} x_{ij} x_{ij} \pi_{c(t)}^{-1} \hat{p}_{ij}^{-1} \tau_{ij},
\]
(4.272)

\[
\tilde{Q} = N_{c(t)}^{-1} \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} x_{ij} x_{ij} \pi_{i(t)}^{-1} p_{ij}^{-1} \tau_{ij},
\]
(4.273)

\[
Q_{N_{c(t)}} = N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x_{ij} x_{ij},
\]
(4.274)

where \( \tau_{ij} \) are the indicator variables for response, defined by (4.151). Then, by Theorem 4.1,

\[
\tilde{Q} - Q_{N_{c(t)}} = O_P \left( n_{c(t)}^{-1/2} \right)
\]
(4.275)

and using (4.270),

\[
\hat{Q} - Q_{N_{c(t)}} = \tilde{Q} - Q_{N_{c(t)}} + \hat{G},
\]
(4.276)

where

\[
\hat{G} = \sum_{\ell=1}^{q_0} \hat{G}_{c(\ell)} (\hat{\theta}(c(\ell)) - \theta^{(0)}),
\]
(4.277)

\[
\hat{G}_{c(\ell)} = N_{c(t)}^{-1} \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} x_{ij} x_{ij} \pi_{c(t)}^{-1} \tau_{ij} x_{ij},
\]
(4.278)

\[
\tau_{ij(\ell)} = \frac{\partial f(x_{ij}, \theta^{(1)})}{\partial \theta^{(1)}} ,
\]
and \( \hat{\theta}_{(t)\ell} \) and \( \theta_{(0)}^{(0)} \) are the \( \ell \)-th element of \( \hat{\theta}_{(t)} \) and \( \theta^{(0)} \). By condition (iv), \( \tau_{ij\ell} \) is bounded by some constant, say \( M_0 \), in an open set containing \( \theta^{(0)} \).

Thus, by conditions (4.252) and (4.256),

\[
N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \left( \sum_{j=1}^{m_i} x_{ij} x_{ij} \tau_{ij\ell}^{-1} \right) \ast \ast (2\delta_4) = O(1). \tag{4.279}
\]

Also, for \( h, h' = 1, 2, \ldots, q \),

\[
\sum_{j=1}^{m_i} \left| x_{ijh} x_{ijh'} \tau_{ij\ell}^{-1} \right|^2 \leq \lambda_1^{-2} M_0^2 \sum_{j=1}^{m_i} \left| x_{ijh} x_{ijh'} \right|^2. \tag{4.280}
\]

Thus,

\[
N_{c(t)}^{-2} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \left| x_{ijh} x_{ijh'} \tau_{ij\ell}^{-1} \right|^2 \leq N_{c(t)}^{-2} \lambda_1^{-2} M_0^2 \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \left| x_{ijh} x_{ijh'} \right|^2 \leq N_{c(t)}^{-1} \lambda_1^{-2} M_0^2 m^{1-\epsilon_4^{-1}} \left( N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \left| x_{ijh} x_{ijh'} \right|^2 \right)^{\epsilon_4^{-1}}. \tag{4.281}
\]

Using condition (4.256),

\[
N_{c(t)}^{-2} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \left( x_{ij} x_{ij} \tau_{ij\ell}^{-1} \right) \ast \ast 2 = O_p \left( n_{c(t)}^{-1/2} \right). \tag{4.282}
\]

Now the conditions of Lemma 4.3 are met, and by the conclusion of Lemma 4.3,

\[
\hat{G}_\ell = G_{N_{c(t)}\ell} + O_p \left( n_{c(t)}^{-1/2} \right), \tag{4.283}
\]

where

\[
G_{N_{c(t)}\ell} = N_{c(t)}^{-1} \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} x_{ij} x_{ij} \tau_{ij\ell}. \tag{4.284}
\]
By condition (4.256),
\[ \mathbf{G}_{N(t)} = O(1), \]  
and it follows that
\[ \mathbf{G}_{t} \left( \hat{\theta}(t) \right) = \left( \mathbf{G}_{N(t)} + O_p \left( n_{e(t)}^{-1/2} \right) \right) O_p \left( n_{e(t)}^{-1/2} \right) = O_p \left( n_{e(t)}^{-1/2} \right). \]  
Thus, from (4.280),
\[ \mathbf{G} = O_p \left( n_{e(t)}^{-1/2} \right), \]  
and using (4.276),
\[ \mathbf{Q} - \mathbf{Q}_{N(t)} = \mathbf{Q} - \mathbf{Q}_{N(t)} + O_p \left( n_{e(t)}^{-1/2} \right) = O_p \left( n_{e(t)}^{-1/2} \right). \]  
Because \( \mathbf{Q} \), the limit of \( \mathbf{Q}_{N(t)} \), is positive definite, by Lemma 4.6,
\[ \mathbf{Q}^{-1} - \mathbf{Q}_{N(t)}^{-1} = O_p \left( n_{e(t)}^{-1/2} \right). \]  
Now, we evaluate \( \hat{\beta} - \beta \). Using (4.251) and (4.273),
\[ \hat{\beta} - \beta = \mathbf{Q}^{-1} \left\{ \sum_{i=1}^{n_{e(t)}} \sum_{j=1}^{m_{t}} x_{ij} (a_{ij} + x_{ij} \beta) p_{ij}^{-1} r_{ij} \right\} - \beta = \left( \mathbf{Q}_{N(t)}^{-1} + O_p \left( n_{e(t)}^{-1/2} \right) \right) \left\{ \hat{\mathbf{H}} + \hat{\mathbf{E}} + \hat{\mathbf{F}} \right\}, \]  
where
\[ \hat{\mathbf{H}} = \sum_{i=1}^{n_{e(t)}} \sum_{j=1}^{m_{t}} x_{ij} a_{ij} p_{ij}^{-1} r_{ij}, \]  
(4.291)
\[ \hat{E} = \sum_{\ell=1}^{\infty} \hat{E}_\ell \left( \hat{\theta}(t) - \theta_\ell^{(0)} \right), \quad (4.292) \]
\[ \hat{E}_\ell = N^{-1}_{c(t)} \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} x'_i a_{ij} \pi_{c(t)}^{-1} r_{ij} \frac{\partial f(x_{ij}, \theta^{(0)})}{\partial \theta_\ell}, \quad (4.293) \]
\[ \hat{F} = \sum_{\ell,s=1}^{\infty} \hat{F}_{\ell s} \left( \hat{\theta}(t) - \theta_\ell^{(0)} \right) \left( \hat{\theta}(t) - \theta_s^{(0)} \right), \quad (4.294) \]
\[ \hat{F}_{\ell s} = 0.5 N^{-1}_{c(t)} \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} x'_i a_{ij} \pi_{c(t)}^{-1} r_{ij} \frac{\partial^2 f(x_{ij}, \theta^{(0)})}{\partial \theta_\ell \partial \theta_s}. \quad (4.295) \]

By arguments similar to those used in the proof of (4.283), and using condition (4.256), we have
\[ \hat{E}_\ell = \mathbb{E}_{N_{c(t)}} + \mathcal{O}_p\left(n^{-1/2}_{c(t)}\right), \quad (4.296) \]
where
\[ \mathbb{E}_{N_{c(t)}} = N^{-1}_{c(t)} \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} x'_i a_{ij} \pi_{c(t)}^{-1} r_{ij} \frac{\partial f(x_{ij}, \theta^{(0)})}{\partial \theta_\ell}. \quad (4.297) \]
By condition (4.265),
\[ \mathbb{E}_{N_{c(t)}} = o(1). \quad (4.298) \]
Then, from (4.294),
\[ \hat{E} = \sum_{\ell=1}^{\infty} \left( \mathbb{E}_{N_{c(t)}} + \mathcal{O}_p\left(n^{-1/2}_{c(t)}\right) \right) \left( \hat{\theta}(t) - \theta_\ell^{(0)} \right) \]
\[ = \mathcal{O}_p\left(n^{-1/2}_{c(t)}\right). \quad (4.299) \]
By the same arguments, using condition (4.256), we have
\[ \hat{F}_{\ell s} = \mathbb{F}_{N_{c(t)}} + \mathcal{O}_p\left(n^{-1/2}_{c(t)}\right), \quad (4.300) \]
where

\[
F_{N_{\ell(t)}} = \frac{1}{2} N_{\ell(t)}^{-1} \sum_{i=1}^{N_{\ell(t)}} \sum_{j=1}^{m_i} x_{ij} a_{ij} p_{ij} \frac{\partial^2 f(x_{ij}, \theta^{(2)})}{\partial \theta_i \partial \theta_j}
\]

\[= O(1). \quad (4.301)\]

Therefore,

\[\hat{F} = O_p \left( n_{\ell(t)}^{-1} \right). \quad (4.302)\]

Now we examine \(\hat{H}\) of (4.293). By Lemma 4.3,

\[\hat{H} = H_{N_{\ell(t)}} = O_p \left( n_{\ell(t)}^{-1/2} \right), \quad (4.303)\]

where

\[H_{N_{\ell(t)}} = N_{\ell(t)}^{-1} \sum_{i=1}^{N_{\ell(t)}} \sum_{j=1}^{m_i} x_{ij}' a_{ij} = 0. \quad (4.304)\]

by condition (4.263). Thus, by (4.303),

\[\hat{H} = O_p \left( n_{\ell(t)}^{-1/2} \right). \quad (4.305)\]

We apply (4.301), (4.302), and (4.305) to (4.292),

\[\hat{\beta} - \beta = \left( Q_{N_{\ell(t)}}^{-1} + O_p \left( n_{\ell(t)}^{-1/2} \right) \right) \left( \hat{H} + o_p \left( n_{\ell(t)}^{-1/2} \right) + O_p \left( n_{\ell(t)}^{-1} \right) \right) \]

\[= Q_{N_{\ell(t)}}^{-1} \hat{H} + o_p \left( n_{\ell(t)}^{-1/2} \right), \quad (4.306)\]

and

\[n_{\ell(t)}^{1/2} \left( \hat{\beta} - \beta \right) = n_{\ell(t)}^{1/2} Q_{N_{\ell(t)}}^{-1} \hat{H} + o_p (1). \quad (4.307)\]
By Lemma 4.2, the design variance of \( \hat{H} \) is

\[
V_{\hat{H}} = V \left( \hat{H} | \xi_{N_{c(t)}} \right) = \sum_{i=1}^{N_{c(t)}} \sum_{j=1}^{m_i} \left[ \pi_{i(t)}^{-1} \left( \pi_{ij}^{-1} - 1 \right) \right] h_{ij} h'_{ij} + 0.5 N_{c(t)}^{-2} \left[ \left( \pi_{i(t)} \pi_{ij(t)} - \pi_{i'j(t)} \right) (h_{i} - h_{i'}) (h_{i} - h_{i'})' \right].
\] (4.308)

and (4.264) follows.

The variance expression in (4.266) can be estimated easily for some special sampling designs. For example, for a simple random sample, the inclusion probabilities and joint inclusion probabilities are

\[
\pi_{i(t)} = n_{c(t)} N_{c(t)}^{-1},
\]
(4.309)

\[
\pi_{i'j(t)} = n_{c(t)} (n_{c(t)} - 1) N_{c(t)}^{-1} \left( N_{c(t)} - 1 \right)^{-1},
\]
(4.310)

and

\[
\pi_{i(t)} \pi_{i'(t)} - \pi_{i'j(t)} = n_{c(t)} N_{c(t)}^{-2} \left( N_{c(t)} - 1 \right) \left( N_{c(t)} - n_{c(t)} \right).
\]
(4.311)

Therefore, we estimate the \( p_{ij}^{-1} \) using the estimator (4.250), and then compute the regression coefficients \( \hat{\beta} \) in (4.251). Let the residual from the regression be

\[
\hat{\alpha} = y - X\hat{\beta}.
\]
(4.312)

Then the estimated cluster totals for

\[
\hat{h}_{ij} = \hat{a}_{ij} x_{ij},
\]
(4.313)
are

\[ \hat{h}_i = \sum_{j=1}^{m_i} \hat{a}_{ij} x_{ij}^t \pi_{it(t)}^{-1} \hat{p}_{ij}^{-1} r_{ij}, \]  

(4.314)

where \( r_{ij} \) is the indicator variable for response for the \( j \)-th element in selected cluster \( i \). It follows that an estimator of \( V_H \) is

\[
\tilde{V}_H = N_{c(t)}^{-2} \sum_{i=1}^{n_{c(t)}} \sum_{j=1}^{m_i} \pi_{it(t)}^{-2} \left( \hat{p}_{ij}^{-1} - 1 \right) \hat{p}_{ij}^{-1} \hat{h}_{ij} \hat{h}_{ij}' r_{ij} \\
+ 0.5 N_{c(t)}^{-2} \sum_{i \neq i' = 1}^{n_{c(t)}} \left[ \left( \pi_{it(t)} \pi_{i't(t)} - \pi_{ii'(t)} \right) \pi_{ii'(t)}^{-1} \left( \hat{h}_{ii} - \hat{h}_{i'i} \right) \left( \hat{h}_{ii} - \hat{h}_{i'i} \right)' \right] 
\]

(4.315)

Thus, an estimator for (4.266) is

\[
\tilde{V} \left( n_{c(t)}^{1/2} Q_{N_{c(t)}}^{-1} \hat{H} \right) = n_{c(t)} Q^{-1} \tilde{V}_H \hat{Q}^{-1}. 
\]

(4.316)
5 REGRESSION WEIGHTING METHODS FOR SIPP DATA

In this chapter, we discuss regression weighting procedures for the Survey of Income and Program Participation (SIPP). The construction of regression weights for such data is investigated.

5.1 The SIPP Data and the Problem

The Census Bureau designed the Survey of Income and Program Participation (SIPP) to provide improved information on income and participation in government programs. Characteristics associated with persons and households which may have impact on income and program participation are collected in the SIPP surveys.

The SIPP is a multistage stratified systematic sample of the noninstitutionalized resident population of the United States. The sample selection for SIPP has three stages: the selection of primary sampling units (PSUs), the selection of address units in sample PSUs, and the determination of persons and households to be included in the sample for the initial and subsequent interviews, the additional requirement for longitudinal surveys. The frame for the selection of sample PSUs consists of a listing of U.S. counties and independent cities, along with population counts and other data for those units from the most recent census of population. For the survey later than 1985, there is a total of 230 sample PSUs. For more details about the sample selection for SIPP, see Jabine, King and Petroni (1990). The SIPP
sample is the sum of four equal sized rotation groups. Each month one rotation group was interviewed. One cycle of four interviews for the four groups is called a wave. Several waves which cover a period of time are called a panel. For example, Panel 1987, composed of seven waves, contains the SIPP-interviewed people from February 1987 through May 1989. A description of SIPP and discussion of SIPP research can be found in Jabine, King, and Petroni (1990), Cavanaugh (1987), Petroni, Huggins, and Cannody (1989). The survey produces two kinds of estimates: cross-sectional and longitudinal. We consider estimation for the panel 1987 longitudinal sample. In order to be a part of the longitudinal sample, the respondent must provide data at each of seven interview periods. About 80% of those that responded at the first interview (Wave One) also responded at the remaining six interviews. A total of 30,766 people interviewed in Wave One were eligible for the 1987 panel longitudinal sample. A total of 24,429 individuals completed all seven interviews. Estimation for the longitudinal sample uses information from all Wave One respondents and also uses control information from the Current Population Survey. We compare alternative estimators that use the information in different ways.

Longitudinal estimators are derived from the weights assigned to the people in the longitudinal sample. Many weighting procedures have been investigated for the longitudinal sample. See Judkins, Hubble, Dorsch, McMillen, and Ernst (1984), Ernst, Hubble, and Judkins (1984), Kobilarcik (1986), Kaspryzk (1989), Petroni (1992), and Folsom and Witt (1994). The current weighting scheme at the U.S. Census Bureau is described by Waite
The procedure makes two adjustments to the base weights, where the base weights are the reciprocals of the probabilities of selection. The adjustments attempt to compensate for nonresponse and undercoverage, using variables thought to be highly correlated with SIPP variables of interest. The first stage adjustment is of the post stratification type. The cells are defined by characteristics of people who were eligible in the Wave One sample. The second stage adjustment is a raking procedure, performed after the first adjustment, using data from the Current Population Survey as controls. Huggins (1987), Ernst and Gillman (1988), Huggins and Fay (1988), Lepkowski (1989), and Folsom and Witt (1994) discussed the adjustment for incomplete SIPP longitudinal data. Some of the research was conducted for SIPP panels earlier than 1987.

We treat the Panel 1987 SIPP data as a three-phase sample. We consider the phase I sample to be the Current Population survey. In the analysis, we assume zero error in these estimates. The phase II sample is the 1987 Wave One data. Phase II included all the people who were eligible and participated in the survey during Wave One. The phase III sample is defined as the subsample from the phase II sample which includes all people who participated in the survey from Wave One through Wave Seven unless they died or moved to an ineligible address. The phase III sample is also called the longitudinal sample of panel 1987.

Because of the confidentiality restrictions, the PSUs and strata information are not available to the public. The SIPP data set released to the public is called the SIPP Research
File, from which our study was conducted. This file contains pseudo clusters and pseudo strata created by the Census Bureau by scrambling the original identification. Therefore, in the rest of this chapter, we use clusters to refer to the pseudo clusters in the SIPP Research File, and when we use strata we are referring to pseudo strata which are provided in the SIPP Research File. There are 72 pseudo strata and 1904 pseudo clusters in the SIPP data.

### 5.2 Notation and Simple Estimators

Let \( x_{ijk} \) be the vector of observations on the \( x \)-variables for the \( k \)-th individual in the \( j \)-th cluster of stratum \( i \), where

\[
x_{ijk} = (x_{ijk1}, x_{ijk2}, \ldots, x_{ijkp}),
\]

\( i = 1, 2, \ldots, L \) is the stratum identification, \( j = 1, \ldots, n_i \) is the cluster within stratum identification, \( k = 1, 2, \ldots, m_{ij} \) is the individual within cluster identification, and \( x_{ijkl} \) is the observation on the \( l \)-th variable for individual \( ijk \), where \( l = 1, 2, \ldots, p \). Characteristics in different samples are identified by I, II, or III according to the sampling phase. In sample \( \tau = I, II, III \), we define the data matrices

\[
\begin{align*}
X^{(\tau)} &= (x_{ijk}) \quad \text{which is an } n^{(\tau)} \times p \text{ matrix}, \\
Y^{(\tau)} &= (y_{ijk}) \quad \text{which is an } n^{(\tau)} \times q \text{ matrix}, \\
Z^{(\tau)} &= (z_{ijk}) \quad \text{which is an } n^{(\tau)} \times r \text{ matrix}, \\
G^{(\tau)} &= [1, X^{(\tau)}] \quad \text{which is an } n^{(\tau)} \times (p + 1) \text{ matrix},
\end{align*}
\]
\[ T^{(r)} = [1, Y^{(r)}] \] which is an \( n^{(r)} \times (q + 1) \) matrix.

where

\begin{equation}
    n^{(r)} = \sum_{i=1}^{L} \sum_{j=1}^{n_i^{(r)}} m_{ij}^{(r)}
\end{equation}

is the total number of individuals in sample \( \tau \), \( n_i^{(r)} \) is the number of clusters in stratum \( i \).
\( m_{ij}^{(r)} \) is the number of individuals in cluster \( j \) of stratum \( i \). If no confusion will result, the sample identification will be omitted. For example, we may write \( X^{(III)} \) simply as \( X \).

The \( X \)-variables are control variables for the phase I sample, the \( Y \)-variables are control variables for the phase II sample, and the \( Z \)-variables are the variables of interest. We assume that in the phase I sample, only \( X \)-variables are observed and that the vector of sample totals of the \( X \)-variables, denoted by \( X_I \), is available. In the phase II sample, we observe \( Y \) and \( X \), and in the phase III sample, we observe \( X, Y, \) and \( Z \). The matrix of initial weights in the phase II sample is denoted by

\[ W^{(II)} = \text{diag} \left( w_{ijk}^{(0,II)} \right) \quad n^{(II)} \times n^{(II)}. \quad (5.4) \]

In the phase II sample, the initial weights \( w_{ijk}^{(0,II)} \) are the inverse of inclusion probabilities adjusted for control variables: age, gender and race, such that the weighted sum, using \( w_{ijk}^{(0,II)} \) as weights, will yield the population values for these variables. Since the phase III sample is a subsample from the phase II sample which includes only respondents, the initial weights in the phase III sample are obtained by adjusting the initial weights \( w_{ijk}^{(0,II)} \) using control variables \( Y \). In the SIPP data, \( Y \) variables are indicator variables for
the noninterview adjustment cells. These noninterview adjustment cells are formed using auxiliary variables that were believed to be correlated with response. The initial weights in the phase III sample are adjusted within each cell. That is, for each element \((i, j, k)\) in the phase III sample, let \(\ell\) be the cell to which \((i, j, k)\) belongs, then the initial weight for \((i, j, k)\) in the phase III sample is

\[
\omega_{ijk}^{(0, III)} = \omega_{ijk}^{(0, II)} \frac{\sum_{(i', j', k') \in \ell, (i', j', k') \in \text{phase II}} \omega_{i'j'k'}^{(0, II)}}{\sum_{(i', j', k') \in \ell, (i', j', k') \in \text{phase II}} \omega_{i'j'k'}^{(0, II)}}.
\]  

(5.5)

The sum of the weights \(\omega_{ijk}^{(0, III)}\) equals the sum of \(\omega_{ijk}^{(0, II)}\) within each noninterview adjustment cell \(\ell\),

\[
\sum_{(i,j,k) \in \ell} \omega_{ijk}^{(0, III)} = \sum_{(i,j,k) \in \ell} \omega_{ijk}^{(0, II)}.
\]  

(5.6)

A second set of initial weights is also used in our analysis. The second weight is the product of the initial weight \(\omega_{ijk}^{(0, II)}\) and the inverse of the estimated response probability \(\hat{p}_{ijk}\) as the weight in the computation. That is, we define the weight for the phase III sample as,

\[
\omega_{ijk}^{(0, III)} = \omega_{ijk}^{(0, II)} \hat{p}_{ijk}^{-1},
\]  

(5.7)

where the \(\hat{p}_{ijk}\) are estimated from the phase II sample. We give the details of the method of estimating \(\hat{p}_{ijk}\) in Section 5.3. For convenience, we call the weights in (5.7) initial weights.
in the rest of this chapter, and we let

\[ W^{(III)} = \text{diag} \left( w_{ijk}^{(0,III)} \right), \quad (5.8) \]

which is an \( n^{(III)} \) by \( n^{(III)} \) diagonal matrix.

The regression coefficient matrices for regression estimation are identified by the sample phase where the regression is applied. For example, \( \hat{\beta}_{Y,X}^{(II)} \) is the least squares estimate of the \( p \times q \) regression coefficient matrix \( \beta_{Y,X} \) obtained by the weighted regression of \( Y \) on \( X \) in the phase II sample. Therefore, we have

\[
\begin{pmatrix}
\hat{\beta}_{Y,X}^{(r)} \\
\hat{\beta}_{0}^{(r)}
\end{pmatrix} = \left( G^{(r)'} W^{(r)} G^{(r)} \right)^{-1} G^{(r)'} W^{(r)} Y^{(r)},
\quad (r = II, III), \quad (5.9)
\]

where \( W^{(II)} \) are defined in (5.4) and \( W^{(III)} \) is defined in (5.8). The total number of individuals in the population is denoted by \( N \) and the population means of the variables are denoted by \( \mu \).

A subscript indicating the phase of the sample is applied to estimated totals. For example,

\[
\hat{X}_{II} = \sum_{i=1}^{L} \sum_{j=1}^{m_{i}^{(II)}} \sum_{k=1}^{n_{i}^{(II)} m_{ij}^{(II)}} w_{ijk}^{(0,II)} x_{ijk}
\quad (5.10)
\]

is the estimated total for \( X \) computed from sample II using the initial weights. Let \( f_{i} \) be the sampling rate for the \( i \)-th stratum, where \( f_{i} = N_{i}^{-1} n_{i} \) and let \( m_{ij} \) be the number of elements in the \( ij \)-th cluster. Then the estimated covariance matrix for \( \hat{X}_{II} \) is

\[
\hat{V} \left( \hat{X}_{II} \right) = \sum_{i=1}^{L} \left( n_{i} - 1 \right)^{-1} n_{i} \left( 1 - f_{i} \right) \sum_{j=1}^{n_{i}} \left( x_{ij} - \bar{x}_{i..} \right) \left( x_{ij} - \bar{x}_{i..} \right) \quad (5.11)
\]
where
\[ x_{ij} = \sum_{k=1}^{m_{ij}} w_{ijk} x_{ijk}, \quad \bar{x}_{..} = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}. \] (5.12)

Similarly, if we have a variable \( Y \), with \( ij\)-th observation \( y_{ijk} \), then the estimated covariance matrix between \( \hat{X}_{II} \) and \( \hat{Y}_{II} \) is
\[ \text{Cov} (\hat{X}_{II}, \hat{Y}_{II}) = \sum_{i=1}^{L} (n_i - 1)^{-1} n_i (1 - f_i) \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})' (y_{ij} - \bar{y}_{..}). \] (5.13)

These are the basic estimators for totals based on weights associated with the sampling design.

### 5.3 Estimation of Response Probabilities

In Chapter 4, we discussed the consistency of regression estimators using the inverse of estimated probabilities. In this section, we describe the model we used to estimate the response probabilities for SIPP. We use the notation of Section 5.2. Assume that sample II is the full sample for the SIPP data, and the phase II sample consists of the respondents who responded on all seven interviews. We denote the response probability associated with the individual \((i, j, k)\) by \( p_{ijk} \) and let the indicator variable for response be \( r_{ijk} \). The estimation procedure for \( p_{ijk} \) is described in the following steps.

**Step 1.** In the phase II sample, regress the indicator variable for response \( r \) on \( Y \), and on both \( X \) and \( Y \), respectively, and calculate the predicted value from the regressions,
\[ r_{\text{Reg on } Y} = Y^{(II)} \left( Y^{(II)'Y^{(II)}} \right)^{-1} Y^{(II)}' r \equiv \hat{r}_{ijk}, \] (5.14)
and
\[
\mathbf{r}_{\text{Reg. on } \mathbf{X} \text{ and } \mathbf{Y}} = (\mathbf{X}^{(II)}, \mathbf{Y}^{(II)}) \left( (\mathbf{X}^{(II)}, \mathbf{Y}^{(II)})' (\mathbf{X}^{(II)}, \mathbf{Y}^{(II)}) \right)^{-1} (\mathbf{X}^{(II)}, \mathbf{Y}^{(II)})' \mathbf{r}. 
\]

(5.15)

where \( \mathbf{r} = (r_{ijk}) \) is the \( n^{(II)} \)-dimensional column vector. We denote the difference of the predicted values from two regressions by
\[
diff = \mathbf{f}_{\text{Reg. on } \mathbf{Y}} - \mathbf{f}_{\text{Reg. on } \mathbf{X} \text{ and } \mathbf{Y}} = (\text{diff}_{ijk})
\]

(5.16)

and denote the sample mean of \( \mathbf{f}_{\text{Reg. on } \mathbf{Y}} \) by
\[
\bar{\mathbf{f}}_{\text{Reg. on } \mathbf{Y}} = n^{(II)} \mathbf{J} \mathbf{f}_{\text{Reg. on } \mathbf{Y}} = n^{(II)} \sum_{i,j,k} \hat{r}_{ijk},
\]

(5.17)

where \( \mathbf{J} \) is the vector whose elements are all ones.

Step 2. Estimate the parameter vector, \( \mathbf{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)' \), of a logistic model,
\[
\frac{1}{1 - p_{ijk}} = 1 + \exp \left\{ \theta_0 + \theta_1 \log \left[ (1 - r_{ijk})^{-1} \right] + \theta_2 \text{diff}_{ijk} + \theta_3 \text{diff}^2_{ijk} + \theta_4 \left( \hat{r}_{ijk} - \bar{f}_{\text{Reg. on } \mathbf{Y}} \right) \right\}. 
\]

(5.18)

Denote the estimates of \( \mathbf{\theta} \) from (5.18) by \( \mathbf{\hat{\theta}} \), and calculate the estimates for \( p_{ijk} \) by
\[
\hat{p}_{ijk} = 1 - \left( 1 + \exp \left\{ \hat{\theta}_0 + \hat{\theta}_1 \log \left[ \hat{r}_{ijk} (1 - \hat{r}_{ijk})^{-1} \right] + \hat{\theta}_2 \text{diff}_{ijk} + \hat{\theta}_3 \text{diff}^2_{ijk} + \hat{\theta}_4 \text{diff}_{ijk} \right\} \right)^{-1}. 
\]

(5.19)

The estimated \( \mathbf{\hat{\theta}} \) is
\[
\hat{\mathbf{\theta}} = \begin{pmatrix} 0.035 & 1.033 & 6.158 & -5.280 & 6.577 \\ 0.052 & 0.036 & 0.317 & 1.794 & 2.506 \end{pmatrix}'.
\]

(5.20)
where the standard errors are shown in the parentheses. Since the regressors in the logistic model (5.19) are functions of predicted values from regressions, the degrees of freedom for $F$-statistics to test the significance of the logistic regression coefficients are the numbers of auxiliary variables in regressions of (5.14) and (5.15). Thus, the $F$-statistics for $\hat{\theta}_1$ in (5.19) is

$$\left( \frac{-1.033}{0.036} \right)^2 / 97 = 8.49,$$

with 97 and 24,248 degrees of freedom. Similarly, the $F$-statistics for $\hat{\theta}_2$ is

$$\left( \frac{6.158}{0.317} \right)^2 / 79 = 4.78,$$

with 79 and 24,248 degrees of freedom.

The $F$-statistic for the intercept $\hat{\theta}_0$ is

$$\left( \frac{0.035}{0.052} \right)^2 = 0.45,$$

with 1 and 24,248 degrees of freedom, and $F$-statistics for $\hat{\theta}_3$ and $\hat{\theta}_4$ can be obtained by

$$\left( \frac{-5.280}{1.794} \right)^2 = 8.66 \quad \text{and} \quad \left( \frac{6.577}{2.506} \right)^2 = 6.89,$$

respectively, both with 1 and 24,248 degrees of freedom. The $F$-statistics indicate that all of the logistic regression coefficients except the intercept are significant at the 1% level. The estimated $\tilde{p}_{ijk}$ in (5.19) will be used as the estimated response probability for individual $(i, j, k)$ later in the mean estimators.

Note that if we assume that the response probability for individual $(i, j, k)$ is $p_{ijk}$, and the respondents form a Poisson sample, then the expected value of the total number of
respondents in the phase II sample is

\[ \sum_{i,j,k} p_{ijk}. \]  

(5.21)

The estimated response probabilities in (5.19) are such that

\[ \sum_{i,j,k} \hat{p}_{ijk} = n^{(III)}, \]  

(5.22)

which is the sample size of the phase III sample. The estimated response probabilities \( \hat{p}_{ijk} \) in (5.19) are used in constructing the initial weights for the phase III sample described in (5.7).

To investigate the goodness of fit of the function, we compare the estimates with the realization. We divide the phase II sample into eight categories corresponding to each row in Table 5.1. Each individual belongs to one category according to the estimated response probability. For example, for individual \((i, j, k)\), if

\[ 0.45 \leq \hat{p}_{ijk} < 0.55, \]  

(5.23)

then this individual is classified into the category which corresponds to "0.45 \(\leq p_{ijk} < 0.55" in the column "Estimated Response Probability." In Table 5.1, the column "Total Observations" contains the total number of individuals in the phase II sample who fell into the corresponding category. The column "Mean of \( \hat{p}_{ijk} \)" shows the mean value of \( \hat{p}_{ijk} \) within each category. The column "Response Rate" is the percentage of the respondents
Table 5.1 Estimated response probabilities

<table>
<thead>
<tr>
<th>Estimated Response Probability</th>
<th>Total Observation</th>
<th>Difference (%)</th>
<th>Mean of $\hat{p}_{ijk}$ (%)</th>
<th>Response Rate (%)</th>
<th>Min $\hat{p}_{ijk}$ (%)</th>
<th>Max $\hat{p}_{ijk}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 \leq \hat{p}_{ijk} &lt; .25$</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$.25 \leq \hat{p}_{ijk} &lt; .35$</td>
<td>9</td>
<td>0.44</td>
<td>33.78</td>
<td>33.33</td>
<td>31.91</td>
<td>34.95</td>
</tr>
<tr>
<td>$.35 \leq \hat{p}_{ijk} &lt; .45$</td>
<td>246</td>
<td>-1.86</td>
<td>41.64</td>
<td>43.50</td>
<td>35.30</td>
<td>44.91</td>
</tr>
<tr>
<td>$.45 \leq \hat{p}_{ijk} &lt; .55$</td>
<td>654</td>
<td>1.37</td>
<td>50.76</td>
<td>49.39</td>
<td>45.10</td>
<td>55.00</td>
</tr>
<tr>
<td>$.55 \leq \hat{p}_{ijk} &lt; .65$</td>
<td>1647</td>
<td>0.09</td>
<td>60.81</td>
<td>60.72</td>
<td>55.04</td>
<td>64.99</td>
</tr>
<tr>
<td>$.65 \leq \hat{p}_{ijk} &lt; .75$</td>
<td>4645</td>
<td>-1.36</td>
<td>70.70</td>
<td>72.06</td>
<td>64.00</td>
<td>74.99</td>
</tr>
<tr>
<td>$.75 \leq \hat{p}_{ijk} &lt; .85$</td>
<td>14081</td>
<td>0.49</td>
<td>81.03</td>
<td>80.54</td>
<td>75.00</td>
<td>85.00</td>
</tr>
<tr>
<td>$.85 \leq \hat{p}_{ijk} \leq .100$</td>
<td>9484</td>
<td>-0.13</td>
<td>87.47</td>
<td>87.60</td>
<td>85.00</td>
<td>95.39</td>
</tr>
</tbody>
</table>

within the category in the phase II sample, that is, in category $\ell$,

$$\text{Response Rate } \ell = \frac{\text{Total Observations}}{\sum_{(i,j,k) \in \ell} r_{ijk}}. \quad (5.24)$$

The column "Difference" is the difference between the mean of the $\hat{p}_{ijk}$ and the response rate within each category,

$$\text{difference} = \text{(mean of } \hat{p}_{ijk}\text{)} - \text{(response rate)}, \quad (5.25)$$

These differences are small in absolute value, but the deviation for the category $0.65 \leq \hat{p}_{ijk} < 0.75$ is about two binomial standard errors. All estimated response probabilities exceed 25%, and the category $0.75 \leq \hat{p}_{ijk} < 0.85$ contains 46% of the individuals in the phase II sample. The column "Min $\hat{p}_{ijk}$" indicates the minimum $\hat{p}_{ijk}$ within each category and "Max $\hat{p}_{ijk}$" indicates the maximum $\hat{p}_{ijk}$. 
5.4 Regression Weighting Schemes

The procedures we consider use auxiliary information in different ways. We outline several regression weighting schemes for the phase III sample.

5.4.1 Regression Weighting by Individuals

Estimators of this section can be given a regression representation in which the regression coefficients are computed using sums of squares and products among individuals.

5.4.1.1 Three-Phase Estimator (Estimation Scheme One)

We give the steps for constructing the three-phase regression estimator.

Step 1. In sample II, construct weights by regressing $Y$ on $X$. Let the regression weights be

$$w_{ijk}^{(1,II)} = w_{ijk}^{(0,II)} \left\{ 1 + N \left[ 0, X_I - \bar{X}_{II} \right] \left( G^{(II)} W^{(II)} G^{(II)} \right)^{-1} [1, x_{ijk}]' \right\}$$

(5.26)

where $i = 1, ..., L$, $j = 1, ..., n_i^{(II)}$, $k = 1, 2, ..., m_{ij}^{(II)}$, $G^{(II)}$ is defined in (5.2). $w_{ijk}^{(0,II)}$ and $W^{(II)}$ is defined by (5.4). The weights are such that

$$\sum_{ijk} w_{ijk}^{(1,II)} [1, x_{ijk}] = [N, X_I].$$

(5.27)

Step 2. In sample II, estimate the mean of $Y$, $\mu_Y$, using the weights in (5.26),

$$\mu_Y^{(1)} = N^{-1} \sum_{ijk} w_{ijk}^{(1,II)} y_{ijk}$$
\[ \hat{Y}_{II} + \left( \bar{X}_I - \bar{X}_{II} \right) \hat{\beta}_{Y.X}^{(II)}. \]  

(5.28)

where \( \hat{\beta}_{Y.X}^{(II)} = \left( G^{(II)} W^{(II)} G^{(II)} \right)^{-1} G^{(II)} W^{(II)} Y^{(II)}. \)

**Step 3.** In sample III, using (5.28) as the controls, regress \( Z \) on \( X \) and \( Y \) to construct the regression weights

\[
\begin{align*}
  w_{ijk}^{(1,III)} &= w_{ijk}^{(0,III)} \left\{ 1 + N \left[ 0, \bar{X}_I - \bar{X}_{III}, \mu_Y^{(1)} - \bar{Y}_{III} \right] \\
  &\quad \left( F^{(III)} W^{(III)} F^{(III)} \right)^{-1} [1, x_{ijk}, y_{ijk}] \\
  \end{align*}
\]

(5.29)

where

\[
\left( \bar{X}_{III}, \bar{Y}_{III} \right) = \left[ \sum_{ijk} w_{ijk}^{(0,III)} \right]^{-1} \sum_{ijk} w_{ijk}^{(0,III)} (x_{ijk}, y_{ijk}),
\]

(5.30)

\[
F^{(III)} = \left[ 1, X^{(III)}, Y^{(III)} \right].
\]

(5.31)

and \( w_{ijk}^{(0,III)} \) are defined by (5.7).

**Step 4.** In sample III, estimate \( \mu_Z \) based on the weights in (5.29):

\[
\begin{align*}
  \hat{\mu}_{\text{three-phase}} &= \left[ \sum_{ijk} w_{ijk}^{(1,III)} \right]^{-1} \sum_{ijk} w_{ijk}^{(1,III)} z_{ijk} \\
  &= \hat{Z}_{III} + \left[ \bar{X}_I - \bar{X}_{III}, \mu_Y^{(1)} - \bar{Y}_{III} \right] \hat{\beta}_{Z.XY} \\
  \hat{\beta}_{0}^{(III)} \left[ \begin{array}{c}
  Z_{III} \\
  \beta \end{array} \right] &= \left( F^{(III)} W^{(III)} F^{(III)} \right)^{-1} \left( F^{(III)} W^{(III)} z^{(III)} \right).
\end{align*}
\]

(5.32)

This completes the description of the three-phase procedure. To estimate the covariance matrix of \( \hat{\mu}_{\text{three-phase}} \), we use the Taylor expansion,

\[
\hat{\mu}_{\text{three-phase}} = \hat{Z}_{III} + \hat{H} \delta + O_{\mu} \left( n^{(III)} \right)^{-1}
\]

(5.33)
where $\hat{H} = [\hat{X}_I - \hat{X}_{III}, \hat{X}_I - \hat{X}_{II}, \hat{Y}_{II} - \hat{Y}_{III}]$, $\beta = [\psi', (\beta_{YX}\psi)'', \psi']'$, $\beta_{Z, XY} = [\psi', \psi']'$, and the partition of $\beta_{Z, XY}$ conforms to the partition $[X, Y]$. Thus,

$$V(\hat{\mu}_{\text{three-phase}}) \approx V(\hat{Z}_{III}) + \text{cov}(\hat{Z}_{III}, \hat{H}) \delta + \delta' \text{cov}(\hat{H}, \hat{Z}_{III}) + \delta' V(\hat{H}) \delta. \quad (5.34)$$

Covariance matrices between two mean estimators from different samples are estimated using the larger sample, assigning the observations not in the small sample zero weights. For example, the weights for sample II that can be used to construct sample III estimates are

$$w_{ijk}^{(0,III)} = \begin{cases} w_{ijk}^{(0,III)} & \text{if } (i, j, k) \in III \\ 0 & \text{otherwise} \end{cases} \quad (5.35)$$

for $(i, j, k) \in II$. Then

$$\hat{X}_{III} = \sum_{III} w_{ijk}^{(0,III)} x_{ijk} = \sum_{II} \hat{w}_{ijk}^{(0,II)} x_{ijk}. \quad (5.36)$$

If we assume the finite population correction is negligible, some estimated covariance matrices are

$$\hat{\text{Cov}}(\hat{Z}_{III}, \bar{Y}_{II}) = \sum_{i=1}^{L} (n_i^{(II)} - 1)^{-1} n_i^{(II)} \sum_{j=1}^{n_i^{(II)}} (a_{ij} - \bar{a}_{..})'(b_{ij} - \bar{b}_{..}), \quad (5.37)$$

$$\hat{V}(\hat{Z}_{III}) = \sum_{i=1}^{L} (n_i^{(II)} - 1)^{-1} n_i^{(II)} \sum_{j=1}^{n_i^{(II)}} (a_{ij} - \bar{a}_{..})'(a_{ij} - \bar{a}_{..}), \quad (5.38)$$

$$\hat{V}(\bar{Y}_{III}) = \sum_{i=1}^{L} (n_i^{(II)} - 1)^{-1} n_i^{(II)} \sum_{j=1}^{n_i^{(II)}} (c_{ij} - \bar{c}_{..})'(c_{ij} - \bar{c}_{..}), \quad (5.39)$$

and

$$\hat{V}(\bar{Y}_{II}) = \sum_{i=1}^{L} (n_i^{(II)} - 1)^{-1} n_i^{(II)} \sum_{j=1}^{n_i^{(II)}} (b_{ij} - \bar{b}_{..})'(b_{ij} - \bar{b}_{..}), \quad (5.40)$$
where

\[ a_{ij} = \hat{N}^{-1} \sum_{k=1}^{m_{ij}} \hat{w}_{ijk}^{(0,II)} (z_{ijk} - \hat{\nu}_{III}), \quad (5.41) \]

\[ b_{ij} = \hat{N}^{-1} \sum_{k=1}^{m_{ij}} \hat{w}_{ijk}^{(0,II)} (y_{ijk} - \hat{\mu}_{II}), \quad (5.42) \]

\[ c_{ij} = \hat{N}^{-1} \sum_{k=1}^{m_{ij}} \hat{w}_{ijk}^{(0,III)} (y_{ijk} - \hat{\nu}_{III}), \quad (5.43) \]

\[ (\hat{\nu}_{III}, \hat{\mu}_{III}) = \hat{N}^{-1} \sum_{(i,j,k) \in III} w_{ijk}^{(0,III)} (y_{ijk}, z_{ijk}), \quad (5.44) \]

\[ \hat{\nu}_{II} = \hat{N}^{-1} \sum_{(i,j,k) \in II} w_{ijk}^{(0,II)} y_{ijk}, \quad (5.45) \]

\[ (\tilde{a}_i, \tilde{b}_i, \tilde{c}_i) = n_i^{-1} \sum_{j=1}^{n_i} (a_{ij}, b_{ij}, c_{ij}), \quad i = 1, 2, ..., L \quad (5.46) \]

and the weights are such that

\[ \hat{N} = \sum_{(i,j,k) \in II} w_{ijk}^{(0,II)} = \sum_{(i,j,k) \in III} w_{ijk}^{(0,III)} = \sum_{(i,j,k) \in II} \hat{w}_{ijk}^{(0,II)}. \quad (5.47) \]

To be totally correct, the multiplier in \( \hat{V} \{ \hat{\nu}_{III} \} \) should be \( (n_i^{(III)} - 1)^{-1} n_i^{(III)} \), where \( n_i^{(III)} \) is the number of primary sampling units in the phase III sample. We use the multiplier \( (n_i^{(II)} - 1)^{-1} n_i^{(II)} \) for simplicity for both sample sizes, because, with about 100 primary sampling units per stratum, the multiplier has little effect. We estimate \( \delta \) in (5.34) by the least squares procedure of Step 3. The estimated covariance matrices and regression coefficients are used to estimate the covariance matrix, \( V \left( \tilde{\mu}_{\text{three-phase}} \right) \) of (5.34).

### 5.4.1.2 Estimation Scheme Two

Estimation scheme two is an approximation to the procedure currently used by the Census Bureau to construct weights for the SIPP data. In this procedure, the information from the
respondents of Wave One is used to construct weights for the panel adjusted for nonresponse. Then the population information from the Current Population Survey is used to create final weights. We give the steps required to construct the estimator.

**Step 1.** In the phase II sample, estimate the controls for \( Y \) using initial sampling weights:

\[
\hat{Y}_{II} = \sum_{(i,j,k)} w_{ijk}^{(0,II)} Y_{ijk},
\]

\[
\hat{\mu}_Y^{(2)} = \hat{Y}_{II} = \left[ \sum_{(i,j,k)} w_{ijk}^{(0,II)} \right]^{-1} \hat{Y}_{II},
\]

(5.48)

where \( w_{ijk}^{(0,II)} \) is defined by (5.4).

**Step 2.** In the phase III sample, construct weights using \( \hat{\mu}_Y^{(2)} \) as the population control:

\[
u_{ijk}^{(2,III)} = w_{ijk}^{(0,III)} \left\{ 1 + N \left[ 0, \hat{\mu}_Y^{(2)} - \hat{Y}_{III} \right] \left( T^{(III)^T} W^{(III)} T^{(III)} \right)^{-1} \left[ 1, y_{ijk} \right] \right\}, \]

(5.49)

where \( T^{(III)} \) is as in (5.2), and \( w_{ijk}^{(0,III)} \) is defined by (5.7). These weights satisfy

\[
\sum_{ijk} \nu_{ijk}^{(2,III)} [1, y_{ijk}] = N \left[ 1, \hat{\mu}_Y^{(2)} \right].
\]

**Step 3.** Estimate \( \mu_Z \) and \( \mu_X \) using weights (5.49):

\[
\left( \hat{\mu}_Z^{(2)}, \hat{\mu}_X^{(2)} \right) = \sum_{ijk} \nu_{ijk}^{(2,III)} (z_{ijk}, x_{ijk})
\]

\[
= \left( \hat{Z}_{III}, \hat{X}_{III} \right) + \left( \hat{Y}_{II} - \hat{Y}_{III} \right) \left( \beta_{Z,Y}^{(III)}, \beta_{X,Y}^{(III)} \right).
\]

(5.50)
Step 4. Construct weights using the regression of $Z$ on $X$, and using $\bar{X}_I$ as the control:

$$u^{(2,III)}_{ijk} = u^{(2,III)}_{ijk} \left\{ 1 + N \left[ 0, \bar{X}_I - \mu^{(2)}_X \right] \left( G^{(III)} U^{(III)}_3 G^{(III)} \right)^{-1} \left[ 1, \mathbf{x}_{ijk} \right]' \right\}. \quad (5.51)$$

where $U^{(III)}_3 = \text{diag}(u^{(2,III)}_{ijk})$.

Step 5. Estimate $\mu_Z$ using weights (5.51):

$$\hat{\mu}_{\text{scheme } #2} = \sum_{(i,j,k)} u^{(2,III)}_{ijk} x_{ijk}$$

$$= \hat{\mu}^{(2)}_Z + (\bar{X}_I - \hat{\mu}^{(2)}_X) \hat{\beta}^{(III)}_{Z,X}. \quad (5.52)$$

The estimate of the covariance matrix of $\hat{\mu}_{\text{scheme } #2}$ is based on the Taylor expansion

$$\hat{\mu}_{\text{scheme } #2} = \hat{Z}_{III} + \left( \bar{X}_I - \bar{X}_{III} \right) \hat{\beta}^{(III)}_{Z,X} + \left( \bar{Y}_I - \bar{Y}_{III} \right) \left\{ \hat{\beta}^{(III)}_{Z,Y} - \beta^{(III)}_{Z,Y} \beta^{(III)}_{X,Y} \right\}$$

$$= \hat{Z}_{III} + \hat{K} \gamma + O_p \left( n^{(III)^{-1}} \right), \quad (5.53)$$

where $\hat{K} = \left[ \bar{X}_I - \bar{X}_{III}, \bar{Y}_I - \bar{Y}_{III} \right]$ and $\gamma = [\beta_{Z,X} \left( \beta_{Z,Y} - \beta_{X,Y} \beta_{Z,X} \right)]'$. Using the same procedure as used for three-phase estimation, we can estimate the covariance matrix of $\hat{\mu}_{\text{scheme } #2}$.

5.4.1.3 Estimation Scheme Three

Estimation scheme three differs from scheme two only in that the totals for the first nonresponse adjustment are regression estimated totals using the Current Population Survey data as control variables. We outline the steps in the estimation.
Step 1. As in Steps 1 - 2 of three-phase estimation, define regression weights for the phase II sample and estimate the mean of $Y$:

$$w_{ijk}^{(3,II)} = w_{ijk}^{(0,II)} \left\{ 1 + \left[ 0, X_I - \bar{X}_{II} \right] \left( G^{(III)} Y W^{(III)} G^{(III)} \right)^{-1} [1, x_{ijk}]^T \right\}. \quad (5.54)$$

$$\hat{\mu}_Y^{(3)} = \sum_{ijk} w_{ijk}^{(3,II)} y_{ijk} = \bar{Y}_{II} + \left( \bar{X}_I - \bar{X}_{II} \right) \hat{\beta}_{Y,X}^{(II)}. \quad (5.55)$$

where $w_{ijk}^{(0,II)}$ is defined by (5.4).

Step 2. In the phase III sample, regress $Z$ on $Y$, using the $\hat{\mu}_Y^{(3)}$ in (5.55) as the control for $Y$, to create weights

$$u_{ijk}^{(3,III)} = w_{ijk}^{(0,III)} \left\{ 1 + N \left[ 0, \hat{\mu}_Y^{(3)} - \bar{Y}_{III} \right] \left( T^{(III)} Y W^{(III)} T^{(III)} \right)^{-1} [1, y_{ijk}]^T \right\}. \quad (5.56)$$

where $w_{ijk}^{(0,III)}$ is defined by (5.7). These weights satisfy

$$\sum_{ijk} u_{ijk}^{(3,III)} y_{ijk} = N \left[ 1, \hat{\mu}_Y^{(3)} \right].$$

Step 3. In the phase III sample, use the weights in (5.56) to estimate the mean of $X$ and $Z$:

$$\hat{\mu}_Z^{(3)} = N^{-1} \sum_{ijk} u_{ijk}^{(3,III)} z_{ijk} = \bar{Z}_{III} + \left( \hat{\mu}_Y^{(3)} - \bar{Y}_{III} \right) \hat{\beta}_{Z,Y}^{(III)}.$$

$$\hat{\mu}_X^{(3)} = N^{-1} \sum_{ijk} u_{ijk}^{(3,III)} x_{ijk} = \bar{X}_{III} + \left( \hat{\mu}_Y^{(3)} - \bar{Y}_{III} \right) \hat{\beta}_{Z,Y}^{(III)}. \quad (5.57)$$

Step 4. In the phase III sample, construct the regression weights, using the regression of $Z$ on $X$ and $\hat{\mu}^{(3)}_Z$ as controls, to create

$$w_{ijk}^{(3,III)} = u_{ijk}^{(3,III)} \left\{ 1 + N \left[ 0, \bar{X}_I - \hat{\mu}^{(3)}_X \right] \left( G^{(III)} Y W^{(III)} G^{(III)} \right)^{-1} [1, x_{ijk}]^T \right\}. \quad (5.58)$$
where $U_2^{(III)} = \text{diag}(u_{ijk}^{(3,III)})$. These weights satisfy
\[
\sum_{ijk} u_{ijk}^{(3,III)} [1, x_{ijk}] = N [1, \hat{\mu}_x^{(3)}].
\]

**Step 5.** Estimate $\mu_z$ using weights (5.58).

\[
\hat{\mu}_{\text{scheme } 3} = \sum_{ijk} u_{ijk}^{(3,III)} z_{ijk}
\]
\[
= \hat{\mu}_z^{(3)} + (\bar{X}_i - \hat{\mu}_x^{(3)}) \beta_{Z,Z}^{(III)}.
\]

The estimate of the covariance matrix of $\hat{\mu}_{\text{scheme } 3}$ is based on the Taylor expansion
\[
\hat{\mu}_{\text{scheme } 3} = \hat{Z}_{III} + \hat{H} \alpha + O_p \left(n^{(III)^{-1}}\right),
\]
where $\hat{H}$ is in (5.33) and
\[
\alpha = (\beta_{Z,X}, (\beta_{Z,Y} - \beta_{X,Y} \beta_{X,Z})', (\beta_{Z,Y} - \beta_{X,Y} \beta_{Z,X})').
\]

Using (5.61), the covariance matrix, $V(\hat{\mu}_{\text{scheme } 3})$, can be estimated as described for three-phase estimation.

**5.4.1.4 Other Estimators**

Some simple estimation procedures are compared to the three-phase estimator and schemes #2 and #3. These estimators are

**Sample mean:**
\[
\hat{Z}_{III} = \left(\sum_{ijk} u_{ijk}^{(0,III)}\right)^{-1} \sum_{ijk} u_{ijk}^{(0,III)} z_{ijk}
\]

**Regression estimator using only $X$-variables:**
\[
\hat{\mu}_{\text{Reg on } X} = \hat{Z}_{III} + (\bar{X}_i - \hat{Z}_{III}) \hat{\beta}_{Z,X}.
\]
Two-phase estimator (on $Y$ only):

$$\hat{\mu}_{\text{Reg on } Y} = \hat{Z}_{III} + \left(\bar{Y}_{II} - \bar{Y}_{III}\right)\hat{\beta}_{Z,Y}$$ (5.64)

The covariance matrices of these estimators can be estimated similarly as described for three-phase estimation.

### 5.4.2 Three-phase Weighting Using Cluster Totals

In SIPP, the primary sampling units are clusters of individuals residing in the same household. An alternative method of defining the regression weights to be used in estimation is to use cluster totals to define the regression coefficients. This is theoretically superior to regressions based on individuals provided individuals are given initial weights that are proportional to the inverses of the inclusion probabilities. In this section, we give an alternative form of the three-phase estimator for cluster sampling. We will use the cluster totals to create the regression weights. These cluster totals will be calculated using two sets of initial weights: the weights $w_{ij}^{(0,III)}$ defined in (5.5), and the weights $w_{ijk}^{(0,III)}$ defined in (5.7), which are adjusted by the estimated response probabilities. We will use the weights $w_{ijk}^{(0,III)}$ to describe the estimation procedures. We outline the construction of the estimator in three steps.

**Step 1.** In the phase II sample, calculate the cluster totals for variables $X$ and $Y$

$$\{(x_{ij}, y_{ij}) = \sum_{k=1}^{m_{ij}^{(II)}} w_{ijk}^{(0,II)} (x_{ijk}, y_{ijk})\}$$ (5.65)
Regression weights are constructed for each cluster by regressing $Y$ on $X$.

$$\begin{align*}
u_{ij}^{(i,II)} &= 1 + \left[ 0, \mathbf{X}_i - \bar{X}_{II} \right] \left( \mathbf{G}^{(II)'} \mathbf{G}^{(II)} \right)^{-1} \begin{bmatrix} 1 \\ x'_{ij} \end{bmatrix}, \tag{5.66}
\end{align*}$$

where $i = 1, 2, \ldots, L; \quad j = 1, \ldots, n_i^{(II)}$, and $\mathbf{G}^{(II)} = ([1, x_{ij}])$ is a $\sum_{i=1}^{L} n_i^{(II)} \times (p + 1)$ matrix. The weights are such that

$$\begin{align*}
\sum_{i,j,k} u_{ij}^{(i,II)} &= N, \tag{5.67} \\
\sum_{ij} w_{ij}^{(i,II)} x_{ij} &= X_i. \tag{5.68}
\end{align*}$$

**Step 2.** In the phase II sample, estimate the mean of $Y$, $\mu_Y$, using the weights in (5.66).

$$\begin{align*}
\tilde{\mu}_Y^{(I)} &= N^{-1} \sum_{ij} u_{ij}^{(i,II)} y_{ij} \\
&= \tilde{Y}_{II} + (\bar{X}_I - \bar{X}_{II}) \tilde{\beta}_{Y,X}^{(II)}, \tag{5.69}
\end{align*}$$

where $\left( \tilde{\beta}_{0}, \tilde{\beta}_{Y,X}^{(II)} \right)' = \left( \mathbf{G}^{(II)'} \mathbf{G}^{(II)} \right)^{-1} \mathbf{G}^{(II)'} \tilde{Y}^{(II)}$ are the regression coefficients based on cluster totals, $\tilde{Y}^{(II)} = (y_{ij})$.

**Step 3.** In the phase III sample, calculate the cluster totals:

$$\begin{align*}
(x_{ij}, y_{ij}, z_{ij}) &= \sum_{k=1}^{m_{ij}^{(III)}} w_{ijk}^{(0,III)} (x_{ijk}, y_{ijk}, z_{ijk}), \tag{5.70}
\end{align*}$$

where $w_{ijk}^{(0,III)}$ are defined in (5.7), and let

$$\begin{align*}
\tilde{X}^{(III)} &= (x_{ij}), \quad \tilde{Y}^{(III)} = (y_{ij}), \quad \tilde{Z}^{(III)} = (z_{ij}), \quad \tilde{X}^{(III)} = (1, \tilde{X}^{(III)}, \tilde{Y}^{(III)}).
\end{align*}$$
Regress $Z$ on $X$ and $Y$ based on these cluster totals to construct the regression weights for each cluster

$$w_{ij}^{(1,III)} = 1 + N \left[ 0, \bar{X}_l - \bar{X}_{III}, \mu_Y^{(1)} - \bar{Y}_{III} \right] \left( \bar{F}^{(III)} \bar{F}^{(III)} \right)^{-1} \begin{bmatrix} 1 \\ x_{ij} \\ y_{ij} \end{bmatrix}, \quad (5.71)$$

where

$$\left( \bar{X}_{III}, \bar{Y}_{III} \right) = N^{-1} \sum_{i,j} (x_{ij}, y_{ij}).$$

These weights are such that

$$\sum_{ij} w_{ij}^{(1,III)} = N,$$

and

$$N^{-1} \sum_{ij} w_{ij}^{(1,III)} (x_{ij}, y_{ij}) = \left( \bar{X}_l, \mu_Y^{(1)} \right), \quad (5.72)$$

where $N$ is the number of clusters.

**Step 4.** In the phase III sample, estimate the mean of $Z$, denoted by $\mu_Z$, using the weight in $(5.71)$:

$$\bar{\mu}_{\text{three-phase cluster}} = N^{-1} \sum_{ij} w_{ij}^{(1,III)} z_{ij},$$

$$= \bar{z}_{III} + \left[ \bar{X}_l - \bar{X}_{III}, \mu_Y^{(1)} - \bar{Y}_{III} \right] \bar{\beta}_{Z,XY}, \quad (5.73)$$

where

$$\begin{bmatrix} \bar{\beta}_0 \\ \bar{\beta}_{Z,XY} \end{bmatrix} = \left( \bar{F}^{(III)} \bar{F}^{(III)} \right)^{-1} \bar{F}^{(III)} \bar{z}^{(III)}.$$
Thus, we obtain the estimator of the mean for \( Z \) based on the regression coefficients calculated from cluster totals. To estimate the covariance matrix of \( \hat{\mu}_{\text{three-phase cluster}} \) we use the Taylor expansion,

\[
\hat{\mu}_{\text{three-phase cluster}} = \hat{Z}_{III} + \hat{H}\delta + O_p\left(n^{-1}\right),
\]

(5.74)

where \( \hat{H} \) and \( \delta \) are as in (5.33). The variance of this three-phase estimator is calculated as in (5.34) using the initial weights \( w_{ijk}^{(0,III)} \) in (5.7).

### 5.4.3 Minimum Variance Estimators

We now give an alternative derivation of regression estimators of the mean of \( Z \). The estimator is a linear function of the vector of sample means

\[
\hat{\mu}_Z = \hat{Q}'\lambda,
\]

(5.75)

where \( \hat{Q}' = \begin{bmatrix} \hat{Z}_{III}, \bar{X}_I - \bar{X}_{III}, \bar{X}_I - \bar{X}_{II}, \bar{Y}_I - \bar{Y}_{III} \end{bmatrix} \), and the coefficient matrix \( \lambda \) is to be determined in order to minimize the variance of \( \hat{\mu}_Z \). For example, the simple estimator (5.62) is of the form of (5.75) with \( \lambda = [I, 0, 0, 0]' \). Similarly, the estimator in (5.52) of Scheme 2 can be written as

\[
\hat{\mu}_{\text{scheme #2}} = \hat{Z}_{III} + \hat{K}\hat{\gamma} = \hat{Q}' \begin{bmatrix} I, \hat{\beta}'_{Z,X}, 0, (\hat{\beta}_{Z,Y} - \hat{\beta}_{X,Y}\hat{\beta}_{Z,X})' \end{bmatrix}'
\]

(5.76)

in the form of (5.75) with \( \lambda = \begin{bmatrix} I, \hat{\beta}'_{Z,X}, 0, (\hat{\beta}_{Z,Y} - \hat{\beta}_{X,Y}\hat{\beta}_{Z,X})' \end{bmatrix}' \). The regression estimator in (5.63) can be written as

\[
\hat{\mu}_{\text{Reg on } X} = \hat{Q}' \begin{bmatrix} I, \hat{\beta}'_{Z,X}, 0, 0 \end{bmatrix}'
\]

(5.77)
in the form of (5.75) with \( \mathbf{A} = [I, \hat{\beta}_{Z,X}', 0, 0]' \). The two-phase estimator in (5.64) can be written

\[
\hat{\mu}_{\text{Reg on } Y} = \hat{Q}' [I, 0, 0, \hat{\beta}_{Z,Y}]',
\]

in the form of (5.75) with \( \mathbf{A} = [I, 0, 0, \hat{\beta}_{Z,Y}]' \).

To determine the optimal \( \mathbf{A} \), we consider the linear model representation

\[
\hat{Q} = [I, 0, 0, 0]' \mu_Z + \mathbf{e}.
\]  

(5.79)

We write the covariance matrix of \( \mathbf{e} \) as

\[
\mathbf{V}(\mathbf{e}) = \mathbf{\Sigma} = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{bmatrix},
\]

(5.80)

where

\[
\Sigma_{11} = \mathbf{V}(\hat{\beta}_{III}),
\]

\[
\Sigma_{12} = \begin{bmatrix}
\text{Cov}(\hat{\beta}_{III}, -\bar{X}_{III}) & \text{Cov}(\hat{\beta}_{III}, -\bar{Y}_{II}) & \text{Cov}(\hat{\beta}_{III}, \bar{Y}_{II} - \bar{Y}_{III})
\end{bmatrix},
\]

\[
\Sigma_{22} = \mathbf{V}[\bar{X}_I - \bar{X}_{III}, \bar{X}_I - \bar{X}_{II}, \bar{Y}_{II} - \bar{Y}_{III}]'.
\]

Note that representation (5.79) is based on the assumption that

\[
E \{ \bar{X}_I - \bar{X}_{III} \} = 0,
\]

\[
E \{ \bar{X}_I - \bar{X}_{II} \} = 0,
\]

\[
E \{ \bar{Y}_{II} - \bar{Y}_{III} \} = 0.
\]
For these assumptions to be satisfied, the weights used to construct the estimators $(\hat{Z}_{III}, \hat{X}_{III}, \hat{Y}_{III})$ must be the inverses of the selection probabilities. Using the representation (5.79), the best linear unbiased estimator of (BLUE) $\mu_Z$ is the generalized least square estimator:

$$
\hat{\mu}_{\text{BLUE}} = \left\{ (I, 0)' \Sigma^{-1} (I, 0)' \right\}^{-1} \left\{ (I, 0) \Sigma^{-1} \hat{Q} \right\}
$$

$$
= \Sigma_{11:2} \left( \Sigma_{11:2}^{-1} - \Sigma_{11:2}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \right) \hat{Q}
$$

$$
= \hat{Z}_{III} - \Sigma_{12} \Sigma_{22}^{-1} \left( \hat{X}_I - \hat{X}_{III}, \hat{X}_I - \hat{X}_{III}, \hat{Y}_I - \hat{Y}_{III} \right)'
$$

$$
= \lambda_{\text{BLUE}}' \hat{Q},
$$

(5.81)

where

$$
\Sigma_{11:2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21},
$$

(5.82)

$$
\lambda_{\text{BLUE}} = \begin{bmatrix}
I \\
-\Sigma_{22}^{-1} \Sigma_{21}
\end{bmatrix}.
$$

(5.83)

That is,

$$
\hat{\mu}_{\text{BLUE}} = \hat{Q}' \lambda_{\text{BLUE}}.
$$

(5.84)

It follows from the linear model representation that the estimator (5.81) with coefficient (5.83) gives the minimum variance among linear unbiased estimators of the form of $\hat{Q}' \lambda$. Therefore, the estimator $\hat{\mu}_{\text{BLUE}}$ is superior to estimators of type (5.75) with any other choice.
of \( \lambda \). If \( \Sigma \) is singular, the BLUE of \( \mu_Z \) is:

\[
\begin{align*}
\hat{\mu}_{\text{BLUE}}' &= \left( \begin{pmatrix} 1 & 0 \end{pmatrix} \Delta^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) \left( \begin{pmatrix} 1 & 0 \end{pmatrix} \Delta^{-1} \mathbf{Q} \right). \\
\Delta &= \Sigma + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} + \mathbf{I} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \\
\Sigma_{ii} &= \Sigma_{ii} + I, \\
\Sigma_{22} &= \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}.
\end{align*}
\]

Then (5.85) becomes

\[
\hat{\mu}_{\text{BLUE}}' = \begin{pmatrix} \hat{\mathbf{z}}' \\ \hat{\mathbf{y}}' \end{pmatrix} = \left( \hat{\mathbf{x}}_I - \hat{\mathbf{x}}_{III} \right) - \left( I + \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \right) \Sigma_{12} \Sigma_{22}^{-1} \left( \begin{pmatrix} \hat{\mathbf{x}}_I - \hat{\mathbf{x}}_{III} \\ \hat{\mathbf{y}}_I - \hat{\mathbf{y}}_{III} \end{pmatrix} \right)'
\]

\[
\lambda^{\text{BLUE-singular}} = \begin{pmatrix} I \\ -\Sigma_{22}^{-1} \Sigma_{21} \left( I + \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right) \end{pmatrix}.
\]

Note that the best choice of \( \lambda \) in (5.83) is a function of the covariance matrix \( \Sigma \) in (5.80). An estimator of \( \Sigma \) is given in section 5.2 based on cluster totals. Therefore, the best linear estimator in (5.81) is obtained from the regression based on cluster totals. The
three-phase estimator analogous to the one described in (5.32) but constructed using cluster
totals is the estimator (5.84) constructed with the estimated covariance matrix.

5.4.4 Superiority of Three-Phase Estimator

In section 5.3.3, we gave the three-phase estimator based on cluster totals. In section
5.3.2 we showed that the best linear estimator in the type of (5.75) is given with \( \lambda \) of
(5.83). In this section, we prove that the estimator given by (5.81) with \( \lambda \) of (5.83) is
equivalent to the three-phase estimator we proposed in (5.74). To simplify, we consider the
single stratum situation. Let

\[
\begin{align*}
\mathbf{a}_j &= N^{-1} \sum_{k=1}^{n_j} \bar{w}_{jk}^{(0,II)} (\mathbf{z}_{jk} - \tilde{Z}_{III}), \\
(\mathbf{b}_j, \mathbf{c}_j) &= N^{-1} \sum_{k=1}^{n_j^{(II)}} \left\{ w_{jk}^{(0,II)} (y_{jk} - \bar{Y}_{II}), \bar{w}_{jk}^{(0,II)} (y_{jk} - \bar{Y}_{III}) \right\}, \\
(\mathbf{d}_j, \mathbf{h}_j) &= N^{-1} \sum_{k=1}^{n_j^{(II)}} \left\{ w_{jk}^{(0,II)} (x_{jk} - \bar{X}_{II}), \bar{w}_{jk}^{(0,II)} (x_{jk} - \bar{X}_{III}) \right\},
\end{align*}
\]

for \( j = 1, \ldots, n^{(II)} \), where \( \bar{w}_{jk}^{(0,II)} \) is defined in (5.34), and \( w^{(0,II)} \) is as in (5.4). By
assumption,

\[
\sum_{j,k} w_{jk}^{(0,II)} = \sum_{j,k} \bar{w}_{jk}^{(0,II)} = N,
\]

we have

\[
(\bar{a}_., \bar{b}_., \bar{c}_., \bar{d}_., \bar{h}_.) = n^{(II)-1} \sum_{j=1}^{n^{(II)}} (\mathbf{a}_j, \mathbf{b}_j, \mathbf{c}_j, \mathbf{d}_j, \mathbf{h}_j) = 0.
\]

Therefore, the estimated covariance matrices for the means can be written as (assuming that
\( (n^{(II)} - 1)^{-1} n^{(III)} \approx 1) \):

\[
S_{ZZ33} = \hat{V} \left( \hat{\mathbf{Z}}_{III} \right) = \sum_{j=1}^{n^{(III)}} a'_j a_j ,
\]

\[
(S_{XX33}, S_{XX22}) = \left[ \hat{V} \left( \hat{\mathbf{X}}_{III} \right), \hat{V} \left( \hat{\mathbf{X}}_{II} \right) \right] = \sum_{j=1}^{n^{(III)}} [h'_j, h'_j, d'_j, d'_j].
\]

\[
(S_{XY32}, S_{XY33}) = \left[ \hat{\text{Cov}} \left( \hat{\mathbf{X}}_{III}, \hat{\mathbf{Y}}_{III} \right), \hat{\text{Cov}} \left( \hat{\mathbf{X}}_{III}, \hat{\mathbf{Y}}_{II} \right) \right]
\]

\[
= \sum_{j=1}^{n^{(III)}} [h'_j b_j, h'_j c_j]. \quad (5.93)
\]

The estimated covariance matrix \( \hat{\text{Cov}} \left( \hat{\mathbf{U}}_r, \hat{\mathbf{V}}_s \right) \) is denoted by \( S_{UVrs} \), where \( U, V \) may be \( X, Y, Z \), and \( r, s \) may be \( 2, 3, \) or \( II, III \).

**Result 5.1** The estimator \( \hat{\mu} \) three-phase cluster in (5.73) is equivalent to \( \hat{\mu}_{BLUE} \) in (5.84), when the covariance matrices in (5.83) are estimated by the matrices in (5.93).

**Proof.** In the phase II sample, the regression coefficient matrix of \( \mathbf{Y} \) regressed on \( \mathbf{X} \), based on cluster totals, is

\[
\hat{\beta}_{Y \cdot X} = \left( \sum_{j=1}^{n^{(III)}} d'_j d_j \right)^{-1} \left( \sum_{j=1}^{n^{(III)}} d'_j b_j \right) = S_{XX22}^{-1} S_{XY22}. \quad (5.94)
\]

Similarly, in the phase III sample, the regression coefficient matrix of \( \mathbf{Z} \) regressed on \( \mathbf{X} \) and \( \mathbf{Y} \) based on cluster totals is:

\[
\hat{\beta}_{Z \cdot XY} = \left( \sum_{j=1}^{n^{(III)}} (h_j, c_j)' (h_j, c_j) \right)^{-1} \left( \sum_{j=1}^{n^{(III)}} (h_j, c_j)' a_j \right)
\]

\[
= \left[ S_{XX33} \ S_{XY33} \right]^{-1} \left[ S_{XZ33} \right]. \quad (5.95)
\]
Write

\[ S_{33} = \begin{bmatrix} S_{XX33} & S_{XY33} \\ S_{YX33} & S_{YY33} \end{bmatrix}, \quad (5.96) \]

then

\[ S_{33}^{-1} = \begin{bmatrix} S_{XX33}^{-1} + S_{XX33}^{-1}S_{XY33}S_{33}^{-1}S_{YX33}S_{XX33}^{-1} & -S_{XX33}^{-1}S_{XY33}S_{33}^{-1} \\ -S_{33}^{-1}S_{YX33}S_{XX33}^{-1} & S_{33}^{-1} \end{bmatrix}, \quad (5.97) \]

where \( S_{33}^{-1} = S_{YY33} - S_{YX33}S_{XX33}^{-1}S_{XY33} \). Therefore, the three-phase estimator \( \hat{\mu}_{\text{three-phase cluster}} \) in (5.73) becomes

\[
\hat{\mu}_{\text{three-phase cluster}} = \hat{\mathbf{Z}}_{III} + \left[ \hat{X}_I - \hat{X}_{III}, \left( \hat{Y}_{II} + \left( \hat{X}_I - \hat{X}_{II} \right) S_{XX22}^{-1} S_{XY22} \right) - \hat{Y}_{III} \right] \begin{bmatrix} S_{33}^{-1} \\ S_{Y33}^{-1} \end{bmatrix} 
= \left[ \hat{\mathbf{Z}}_{III}, \hat{X}_I - \hat{X}_{III}, \hat{X}_I - \hat{X}_{II}, \hat{Y}_{II} - \hat{Y}_{III} \right] \begin{bmatrix} I \\ \tilde{\delta} \end{bmatrix}. \quad (5.98)
\]

where

\[
\tilde{\delta} = \begin{bmatrix} S_{XX33}^{-1}S_{XX33} + S_{XX33}^{-1}S_{XY33}S_{33}^{-1}S_{YY33}S_{XX33}^{-1} \left( S_{YY33}S_{XX33}^{-1}S_{XX33} - S_{Y33} \right) \\ -S_{XX22}^{-1}S_{XY22}S_{33}^{-1} \left( S_{Y33}S_{XX33}^{-1}S_{XX33} - S_{Y33} \right) \\ -S_{33}^{-1} \left( S_{Y33}S_{XX33}^{-1}S_{XX33} - S_{Y33} \right) \end{bmatrix}. \quad (5.99)
\]

From the linear model approach, as shown in (5.83), the regression coefficients are

\[
\delta = -\Sigma_{22}^{-1} \Sigma_{21}, \quad (5.100)
\]
where

\[
\Sigma_{21} = \begin{bmatrix}
\text{Cov} \left( \bar{X}_I - \bar{X}_{III}, \bar{X}_{III} \right) \\
\text{Cov} \left( \bar{X}_I - \bar{X}_{II}, \bar{Z}_{III} \right) \\
\text{Cov} \left( \bar{Y}_I - \bar{Y}_{III}, \bar{Z}_{III} \right)
\end{bmatrix} = \Sigma_{12}'.
\]

(5.101)

\[
\Sigma_{22} = \text{Var} \begin{bmatrix}
-\bar{X}_{III} \\
-\bar{X}_{II} \\
\bar{Y}_{III} - \bar{Y}_{III}
\end{bmatrix}.
\]

(5.102)

Let \( n_2 = n^{(II)}, n_3 = n^{(III)}, \)

\[
\Sigma_{21} = \begin{bmatrix}
-n_3^{-1} \Sigma_{XZ} \\
n_2^{-1} \Sigma_{XZ} \\
n_3^{-1} \Sigma_{YZ}
\end{bmatrix}.
\]

(5.103)

\[
\Sigma_{22} = \begin{bmatrix}
n_3^{-1} \Sigma_{XX} & n_2^{-1} \Sigma_{XX} & (n_3^{-1} - n_2^{-1}) \Sigma_{XY} \\
n_2^{-1} \Sigma_{XX} & n_2^{-1} \Sigma_{XX} & 0 \\
(n_3^{-1} - n_2^{-1}) \Sigma_{YY} & 0 & (n_3^{-1} - n_2^{-1}) \Sigma_{YY}
\end{bmatrix}.
\]

(5.104)

Define

\[
S_{22} = \begin{bmatrix}
n_2^{-1} \Sigma_{XX} & 0 \\
0 & (n_3^{-1} - n_2^{-1}) \Sigma_{YY}
\end{bmatrix},
\]

\[
S_{221} = S_{22} - n_2^{-1} \begin{bmatrix}
\Sigma_{XX} \\
(f^{-1} - 1) \Sigma_{XX}, (f^{-1} - 1) \Sigma_{XY}
\end{bmatrix} f \Sigma_{XX}^{-1} \left[ \Sigma_{XX}, (f^{-1} - 1) \Sigma_{XY} \right].
\]
\[ n_2^{-1} (1 - f) \begin{bmatrix} \Sigma_{xx} & -\Sigma_{xy} \\ -\Sigma_{yx} & f^{-1} \left\{ \Sigma_{yy} - (1 - f) \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \right\} \end{bmatrix} \] (5.105)

where \( f = n_3 n_2^{-1} \). Therefore, the inverse matrix of \( V_{22;1} \) is:

\[ S_{22;1}^{-1} = n_2 (1 - f)^{-1} \begin{bmatrix} \Sigma_{xx}^{-1} + f \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yx}^{-1} \Sigma_{xx} & f \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yx}^{-1} \\ f \Sigma_{yx}^{-1} \Sigma_{xy} \Sigma_{xx}^{-1} & \Sigma_{yx}^{-1} \end{bmatrix}, \] (5.106)

where

\[ \Sigma_{yx} = \Sigma_{yy} - \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}. \] (5.107)

Also, let

\[ S_{12} = \begin{bmatrix} n_2^{-1} \Sigma_{xx} & n_3^{-1} (1 - f) \Sigma_{xy} \end{bmatrix} = S'_{21}. \] (5.108)

Then

\[ \Sigma_{22} = \begin{bmatrix} n_3^{-1} \Sigma_{xx} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \] (5.109)

and

\[ \Sigma_{22}^{-1} = \begin{bmatrix} n_3 \Sigma_{xx}^{-1} + n_3 \Sigma_{xx}^{-1} S_{12} S_{22}^{-1} S_{21} \left( n_3 \Sigma_{xx}^{-1} \right) & -n_3 \Sigma_{xx}^{-1} S_{12} S_{22}^{-1} \\ -n_3 S_{22}^{-1} S_{21} \Sigma_{xx}^{-1} & S_{22}^{-1} \end{bmatrix}, \] (5.110)

where

\[ n_3 \Sigma_{xx}^{-1} + n_3^2 \Sigma_{xx}^{-1} S_{12} S_{22}^{-1} S_{21} \Sigma_{xx}^{-1} = n_3 (1 - f)^{-1} \Sigma_{xx}^{-1} + n_2 f (1 - f)^{-1} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yx}^{-1} \Sigma_{xx}^{-1}. \] (5.111)
and

\[-n_3 \Sigma_{XX}^{-1} S_{12} S_{22}^{-1} \]

\[= - \left[ f I, (1 - f) \Sigma_{XX}^{-1} \Sigma_{XY} \right] S_{22}^{-1} \]

\[= -n_3 (1 - f)^{-1} \left[ \Sigma_{XX}^{-1} + \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \Sigma_{YX}, \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \right]. \quad (5.112) \]

Therefore, by (5.103) and (5.110), the regression coefficient matrix in (5.100) is

\[\delta = -\Sigma_{22}^{-1} \Sigma_{21} = [\tau_1', \tau_2', \tau_3']' \quad (5.113)\]

where

\[\tau_1 = (1 - f)^{-1} \Sigma_{XX}^{-1} \Sigma_{XZ} + (1 - f)^{-1} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \Sigma_{XZ} - f (1 - f)^{-1} \Sigma_{XX}^{-1} \Sigma_{XZ} \]

\[+ f (1 - f)^{-1} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \Sigma_{XZ} - \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \Sigma_{XZ}, \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \Sigma_{XZ} \quad (5.114)\]

\[= \Sigma_{XX}^{-1} \Sigma_{XZ} + \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \left( \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XZ} - \Sigma_{YZ} \right) , \]

\[\tau_2 = - (1 - f)^{-1} \left[ \Sigma_{XX}^{-1} + \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \Sigma_{XX}^{-1} \right] \Sigma_{XZ} \]

\[+ (1 - f)^{-1} \left[ \Sigma_{XX}^{-1} + f \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \Sigma_{XX}^{-1} \right] \Sigma_{XZ} \]

\[+ \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \Sigma_{YZ} \quad (5.115)\]

\[= - \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YX}^{-1} \left( \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XZ} - \Sigma_{YZ} \right) \]

\[\tau_3 = - (1 - f)^{-1} \Sigma_{YX}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XZ} \]

\[+ (1 - f)^{-1} f \Sigma_{YX}^{-1} \Sigma_{YY} \Sigma_{XX}^{-1} \Sigma_{XZ} + \Sigma_{YX} \Sigma_{YZ} \quad (5.116)\]

\[= - \Sigma_{YX}^{-1} \left( \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XZ} - \Sigma_{YZ} \right) . \]
If we estimate the population covariance matrices in $\tau_1$, $\tau_2$, and $\tau_3$ by those in (5.93), then we will get

$$\hat{\delta} = \begin{bmatrix} \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \end{bmatrix} = \tilde{\delta}, \quad (5.117)$$

where $\tilde{\delta}$ is defined in (5.99). Hence,

$$\tilde{\mu}_{\text{three-phase cluster}} = \tilde{\mu}_{\text{BLUE}}.$$

5.5 Application to the SIPP Data

We compare regression weighting methods for the Panel 1987 data from SIPP. The phase I sample is the Current Population Survey. We assume that there is zero error for estimated means from the phase I sample. The phase II sample is the Panel 1987 Wave One sample. The sample size of the phase II sample is 30,766 individuals in 11,660 households. The phase III sample is the Panel 1987 longitudinal sample. The sample size of the phase III sample is 24,429 individuals in 9,776 households.

Equation (5.35) defines the weights used in calculating the covariances between means in different samples. The weights used for the phase III sample in these calculations depend on the way in which the phase III sample is selected from the phase II sample. In the SIPP situation the longitudinal sample is self selecting so that it is necessary to use a model for
the selection procedure. We model the phase III sample as a Poisson sample from the phase II sample.

The regression variables are based on the non-interview adjustment cells and on the Current Population Survey variables used by the Census Bureau to construct weights for the Panel 1987 longitudinal sample. The $X$-variables are the variables associated with the second-stage adjustment used by the Census Bureau. The second-stage adjustment variables are based on gender, age, race, family type, and household type. There are 97 $X$ variables in our analysis. The definitions of $X$ variables are given in Tables 5.2 - Table 5.5. In these tables, names of the indicator variables are shown in the cells which indicator variables represent. For example, in Table 5.2, $X_1$ is the indicator variable for the cell “White Male, Age of 0-1”, i.e.

$$X_1 = \begin{cases} 
1 & \text{if the individual is a white male and age of 0-1} \\
0 & \text{otherwise.}
\end{cases} \quad (5.118)$$

If there is no variable name in a cell (empty entry) in these tables, then there is no indicator variable generated for this cell, and this cell can be represented by combinations of other indicator variables. For example, there is no entry for the cell “Black-Female-Age-of-65-or-older” in Table 5.3. This is because this category can be represented by the value zero for all other indicator variables. This definition guarantees we will have a nonsingular design matrix in the regression.
Table 5.2. Definitions of $X$ variables for children based on race, gender, and age.

<table>
<thead>
<tr>
<th>Race</th>
<th>Gender</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6-7</td>
</tr>
<tr>
<td>White</td>
<td>Male</td>
<td>X1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X7</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>X8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X14</td>
</tr>
<tr>
<td>Non-White</td>
<td>Male</td>
<td>X15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X18</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>X19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X22</td>
</tr>
</tbody>
</table>

The $Y$ variables are indicator variables for the non-interview adjustment cells in the first stage adjustment procedure described in Waite (1990). The non-interview adjustment cells are formed using variables such as level of income, race, education, type of income, type of assets, labor force status, and employment status. There are 79 $Y$ variables for 80 categories used in our analysis. The definitions of $Y$ variables are given in Table 5.6 - Table 5.9. These tables are similar to Table 5.2 - Table 5.5. For example, in Table 5.6, $Y_2$ is an indicator variable for the category

$$Y_2 = \begin{cases} 
1 & \text{if white, non-Hispanic, with 12 or more years of education, unemployed,} \\
& \text{not on welfare, and living in low income household} \\
0 & \text{otherwise.}
\end{cases}$$

(5.119)

Again, in Table 5.6 - Table 5.9, if there is an empty entry for a cell (for example, in Table 5.9), then that cell can be represented by other indicator variables we have created.
Table 5.3. Definition of $X$ variables for adults based on race, gender, and age.

<table>
<thead>
<tr>
<th>Age</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>15</td>
<td>X23</td>
<td>X41</td>
</tr>
<tr>
<td>16-17</td>
<td>X24</td>
<td>X42</td>
</tr>
<tr>
<td>18-19</td>
<td>X25</td>
<td>X43</td>
</tr>
<tr>
<td>20-21</td>
<td>X26</td>
<td>X44</td>
</tr>
<tr>
<td>22-24</td>
<td>X27</td>
<td>X45</td>
</tr>
<tr>
<td>25-29</td>
<td>X28</td>
<td>X46</td>
</tr>
<tr>
<td>30-34</td>
<td>X29</td>
<td>X47</td>
</tr>
<tr>
<td>35-39</td>
<td>X30</td>
<td>X48</td>
</tr>
<tr>
<td>40-44</td>
<td>X31</td>
<td>X49</td>
</tr>
<tr>
<td>45-49</td>
<td>X32</td>
<td>X50</td>
</tr>
<tr>
<td>50-54</td>
<td>X33</td>
<td>X51</td>
</tr>
<tr>
<td>55-59</td>
<td>X34</td>
<td>X52</td>
</tr>
<tr>
<td>60-61</td>
<td>X35</td>
<td>X53</td>
</tr>
<tr>
<td>62-64</td>
<td>X36</td>
<td>X54</td>
</tr>
<tr>
<td>65-69</td>
<td>X37</td>
<td>X55</td>
</tr>
<tr>
<td>70-74</td>
<td>X38</td>
<td>X56</td>
</tr>
<tr>
<td>75-79</td>
<td>X39</td>
<td>X57</td>
</tr>
<tr>
<td>&gt;80</td>
<td>X40</td>
<td>X58</td>
</tr>
</tbody>
</table>
Table 5.4. Definition of $X$ variables for adults based on gender, race, and family type.

<table>
<thead>
<tr>
<th>Household</th>
<th>Living with Relative</th>
<th>Not Living with Relative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Other</td>
</tr>
</tbody>
</table>
|           | Householder-
|           | No Spouse
|           | Present - With
|           | Own Children        | Other                    |
|           |                      | Female                  | Male                    |
| Male      | White                | X76                     | X77                     |
|           | Black                | X85                     | X86                     |
| Female    | White                | X80                     | X81                     |
|           | Black                | X89                     | X90                     |
|           |                      | X82                     | X91                     |

<table>
<thead>
<tr>
<th>Not Household</th>
<th>Related to Householder</th>
<th>Not Related to Householder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spouse of Householder</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>or Related Subfamily</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>White</td>
<td>X78</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>X87</td>
</tr>
<tr>
<td>Female</td>
<td>White</td>
<td>X83</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>X92</td>
</tr>
</tbody>
</table>

Table 5.5. Definition of $X$ variables for Hispanic based on age.

<table>
<thead>
<tr>
<th>Age</th>
<th>&lt; 15</th>
<th>15 ~ 24</th>
<th>25 ~ 44</th>
<th>&gt; 44</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>X94</td>
<td>X95</td>
<td>X96</td>
<td>X97</td>
</tr>
</tbody>
</table>
Table 5.6. Definition of $Y$ variables for low income class (average monthly household income < $1,200).

<table>
<thead>
<tr>
<th>Race</th>
<th>Education</th>
<th>Type of Income in Households</th>
<th>Labor Force Status</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Unemployment Benefits, No Welfare</td>
<td>In</td>
<td>Not</td>
</tr>
<tr>
<td>White, Non Hispanic</td>
<td>&lt; 12 yrs</td>
<td>Y1</td>
<td>Y3</td>
<td>Y7</td>
</tr>
<tr>
<td></td>
<td>12-15 yrs</td>
<td>Y2</td>
<td>Y4</td>
<td>Y8</td>
</tr>
<tr>
<td></td>
<td>16+ yrs</td>
<td>Y2</td>
<td>Y4</td>
<td>Y8</td>
</tr>
<tr>
<td>Other</td>
<td>&lt; 12 yrs</td>
<td>Y1</td>
<td>Y5</td>
<td>Y9</td>
</tr>
<tr>
<td></td>
<td>&gt; 12 yrs</td>
<td>Y1</td>
<td>Y6</td>
<td>Y10</td>
</tr>
</tbody>
</table>

Table 5.7. Definition of $Y$ variables for middle income class ($1,200 < average monthly household incomes < $5,000)

<table>
<thead>
<tr>
<th>Race</th>
<th>Education</th>
<th>Employment Status</th>
<th>Assets</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bonds</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Type of Income</td>
<td>Welfare</td>
<td>Labor Force</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unemployment</td>
<td>In</td>
<td>Not</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Benefits, No Welfare</td>
<td>No Welfare</td>
<td></td>
</tr>
<tr>
<td>White, Non-Hispanic</td>
<td>&lt; 12 yrs</td>
<td>Y21</td>
<td>Y25</td>
<td>Y26</td>
</tr>
<tr>
<td></td>
<td>12-14 yrs</td>
<td>Y22</td>
<td>Y25</td>
<td>Y26</td>
</tr>
<tr>
<td></td>
<td>16+ yrs</td>
<td>Y23</td>
<td>Y25</td>
<td>Y26</td>
</tr>
<tr>
<td>Other</td>
<td>&lt; 12 yrs</td>
<td>Y24</td>
<td>Y25</td>
<td>Y26</td>
</tr>
<tr>
<td></td>
<td>&gt; 12 yrs</td>
<td>Y24</td>
<td>Y25</td>
<td>Y26</td>
</tr>
</tbody>
</table>
Table 5.8. Definitions of $V$ variables for middle income class ($1,200 \leq \text{average monthly household income} \leq 4,000$), not self employed and type of income other.

<table>
<thead>
<tr>
<th>Race</th>
<th>Education</th>
<th>Employment Status</th>
<th>Assets</th>
<th>Type of Income</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Self-Employed</td>
<td>Bonds</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>White, Non Hispanic</td>
<td>&lt;12 yrs.</td>
<td>Y39</td>
<td>Y50</td>
<td>Y56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12-15 yrs.</td>
<td>Y40</td>
<td>Y51</td>
<td>Y57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16+ yrs.</td>
<td>Y41</td>
<td>Y52</td>
<td>Y58</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>&lt;12 yrs.</td>
<td>Y42</td>
<td>Y53</td>
<td>Y59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12-15 yrs.</td>
<td>Y43</td>
<td>Y54</td>
<td>Y59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16+ yrs.</td>
<td>Y44</td>
<td>Y55</td>
<td>Y60</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.9. Definition of $Y$ variables for high income class (average household income > $4,000$).

<table>
<thead>
<tr>
<th>Race</th>
<th>Education</th>
<th>Employment Status</th>
<th>Assets</th>
<th>Type of Income</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Self-Employed</td>
<td>Bonds</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>White, Non Hispanic</td>
<td>&lt;12 yrs.</td>
<td>Y61</td>
<td>Y64</td>
<td>Y69</td>
<td>Y73</td>
</tr>
<tr>
<td></td>
<td>12-15 yrs.</td>
<td>Y62</td>
<td>Y65</td>
<td>Y70</td>
<td>Y74</td>
</tr>
<tr>
<td></td>
<td>16+ yrs.</td>
<td>Y63</td>
<td>Y66</td>
<td>Y71</td>
<td>Y75</td>
</tr>
<tr>
<td>Other</td>
<td>&lt;12 yrs.</td>
<td>Y61</td>
<td>Y67</td>
<td>Y72</td>
<td>Y76</td>
</tr>
<tr>
<td></td>
<td>12-15 yrs.</td>
<td>Y61</td>
<td>Y67</td>
<td>Y72</td>
<td>Y76</td>
</tr>
<tr>
<td></td>
<td>16+ yrs.</td>
<td>Y61</td>
<td>Y68</td>
<td>Y72</td>
<td>Y77</td>
</tr>
</tbody>
</table>
The $Z$ variables used in our analysis are Personal Income, Personal Earnings, Family Income, Family Earnings, Family Property Income, Family Means Tested Transfers, Family Other Income, Household Earnings, Household Property Income, Household Means Tested Transfers, and Household Other Income. All variables are recorded for January 1987 and for January 1989. For example, Personal Income for January 1987 was the total income of the person in January of 1987. Family income for January 1989 is the total income of the family with which the interviewed person lived in January 1989. Similarly, Household Earnings for January 1987 is the total earnings of the household in which the interviewed person lived in January 1987. The Census Bureau defines family and household differently. The household is the sample unit for the SIPP. A household may have more than one family. The terms income, earnings, property income, means-tested income transfers and "other income" are different sources of income for individuals and households.

5.5.1 Regressions Based on Individuals

Estimated standard errors for the three schemes using regressions with individuals as observations are compared in Table 5.10. The estimated means of the $Z$ variables are listed in the column "Estimate". These estimates were calculated using the three-phase estimator. Estimates of the means computed by other schemes are omitted to simplify the table. The estimated standard errors for the means from scheme #1 are listed under the column “s.e. #1”, where scheme #1 is the three-phase estimator. The variances were calculated using the
Table 5.10. Estimated means for SIPP panel 1987 data and estimated ratios of standard errors for the three schemes using individuals to estimate regression coefficients.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Estimate</th>
<th>s.e. #1 ($)</th>
<th>s.e. #2</th>
<th>s.e. #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 87 Personal Income</td>
<td>981.8</td>
<td>7.76</td>
<td>1.012</td>
<td>1.003</td>
</tr>
<tr>
<td>Jan 89 Personal Income</td>
<td>1037.7</td>
<td>7.57</td>
<td>1.008</td>
<td>1.002</td>
</tr>
<tr>
<td>Jan 87 Personal Earnings</td>
<td>755.0</td>
<td>7.15</td>
<td>1.012</td>
<td>1.003</td>
</tr>
<tr>
<td>Jan 89 Personal Earnings</td>
<td>791.5</td>
<td>6.69</td>
<td>1.008</td>
<td>1.002</td>
</tr>
<tr>
<td>Jan 87 Family Income</td>
<td>2743.6</td>
<td>24.10</td>
<td>0.993</td>
<td>1.001</td>
</tr>
<tr>
<td>Jan 89 Family Income</td>
<td>2849.0</td>
<td>23.72</td>
<td>0.994</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 87 Family Earnings</td>
<td>2246.7</td>
<td>23.55</td>
<td>0.990</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 89 Family Earnings</td>
<td>2313.1</td>
<td>21.60</td>
<td>0.992</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 87 Family Property Income</td>
<td>150.5</td>
<td>5.28</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 89 Family Property Income</td>
<td>153.4</td>
<td>5.12</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 87 Family Means Tested Transfers</td>
<td>31.1</td>
<td>1.76</td>
<td>0.994</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 89 Family Means Tested Transfers</td>
<td>29.2</td>
<td>1.75</td>
<td>1.000</td>
<td>1.003</td>
</tr>
<tr>
<td>Jan 87 Family Other Income</td>
<td>315.3</td>
<td>5.64</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 89 Family Other Income</td>
<td>353.3</td>
<td>8.54</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 87 Household Income</td>
<td>2818.9</td>
<td>24.43</td>
<td>0.993</td>
<td>1.001</td>
</tr>
<tr>
<td>Jan 89 Household Income</td>
<td>2922.7</td>
<td>23.77</td>
<td>0.994</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 87 Household Earnings</td>
<td>2311.0</td>
<td>23.86</td>
<td>0.990</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 89 Household Earnings</td>
<td>2364.4</td>
<td>21.78</td>
<td>0.992</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 87 Household Property Income</td>
<td>152.2</td>
<td>5.29</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 89 Household Property Income</td>
<td>154.9</td>
<td>5.14</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 87 Household Means Tested Transfers</td>
<td>32.9</td>
<td>1.91</td>
<td>0.994</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 89 Household Means Tested Transfers</td>
<td>30.2</td>
<td>1.78</td>
<td>1.000</td>
<td>1.003</td>
</tr>
<tr>
<td>Jan 87 Household Other Income</td>
<td>322.8</td>
<td>5.86</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 89 Household Other Income</td>
<td>360.6</td>
<td>8.62</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>Jan 87 Labor Force (%)</td>
<td>46.0</td>
<td>0.23</td>
<td>1.021</td>
<td>1.045</td>
</tr>
<tr>
<td>Jan 89 Labor Force (%)</td>
<td>47.0</td>
<td>0.24</td>
<td>1.008</td>
<td>1.015</td>
</tr>
</tbody>
</table>

s.e. = standard error.
methods described in Section 5.4. The ratios of estimated standard errors from other schemes to the standard error from scheme #1 are also listed in the table. The differences among the standard errors for the three schemes are small. Because the phase III sample is about 80% of the phase II sample, there must be very large differences in the regression correlations to produce noticeable differences among the standard errors.

If the regression coefficients were computed using cluster totals, the three-phase estimator would always dominate the other two estimators to the degree of accuracy employed in the Taylor approximations. Because the regression coefficients are computed using individuals as observations, it is possible for the estimated standard deviations for schemes two and three to be less than the estimated standard deviation for three-phase estimation.

The results are mildly surprising in that procedure two, the approximation to the current Census Bureau procedure, performs marginally better than the other two procedures. It must be realized that these are estimated variances and, in particular, that the ratio of the variance of the phase II sample to the variance of the phase III sample is estimated. There may be a hidden bias in that the variables used in the analysis are those selected by the Census Bureau.

The ratio of the standard error of the mean of $Z$ in the phase III sample (estimator (5.62)) to the standard error of the three-phase estimator, the ratio of the standard error of the regression estimator using only $X$-variables (estimator (5.63)) to the standard error
of the three-phase estimator, and the ratio of the standard error of the two-phase estimator using only \( Y \)-variables (estimator (5.64)) to the standard error of the three-phase estimator are given in Table 5.11. As expected, each of these three procedures is uniformly inferior to the three-phase estimator.

Also, the regression procedure using \( X \)-variables is uniformly superior to the two-phase estimator using only \( Y \)-variables. The gains from using the \( Y \)-variables in addition to the \( X \)-variables ranges from 0.2% for January '89 “Other Income” to 7.0% for January '87 Labor Force Status.

### 5.5.2 Regressions Based on Cluster Totals

We also investigated the standard errors of the regression estimator in which the regression coefficient matrices are based on the estimated covariance matrices calculated from cluster totals. The variances we will compute are based on the pseudo clusters of the public Research File. These variances may be different from the variances based on the true PSU totals. However, we are not able to compute the variances based on the true cluster totals because the identification for primary sampling units is not available to the public due to confidentiality.

The three-phase estimator and scheme two are compared. For each mean estimator, the regression coefficient matrices are such that the Taylor approximation to the variance of the mean estimator is a minimum for that particular procedure. The two schemes are compared in Table 5.12. Since the covariance matrices are calculated based on cluster totals
Table 5.11. Estimated ratios of standard errors for alternative estimators using individuals to estimate regression coefficients.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean s.e.</th>
<th>Reg. X s.e.</th>
<th>2-Ph. Y s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3-Ph. s.e.</td>
<td>3-Ph. s.e.</td>
<td>3-Ph. s.e.</td>
</tr>
<tr>
<td>Jan 87 Personal Income</td>
<td>1.240</td>
<td>1.044</td>
<td>1.174</td>
</tr>
<tr>
<td>Jan 89 Personal Income</td>
<td>1.234</td>
<td>1.036</td>
<td>1.171</td>
</tr>
<tr>
<td>Jan 87 Personal Earnings</td>
<td>1.247</td>
<td>1.044</td>
<td>1.173</td>
</tr>
<tr>
<td>Jan 89 Personal Earnings</td>
<td>1.272</td>
<td>1.036</td>
<td>1.200</td>
</tr>
<tr>
<td>Jan 87 Family Income</td>
<td>1.138</td>
<td>1.046</td>
<td>1.084</td>
</tr>
<tr>
<td>Jan 89 Family Income</td>
<td>1.125</td>
<td>1.034</td>
<td>1.082</td>
</tr>
<tr>
<td>Jan 87 Family Earnings</td>
<td>1.161</td>
<td>1.042</td>
<td>1.107</td>
</tr>
<tr>
<td>Jan 89 Family Earnings</td>
<td>1.174</td>
<td>1.035</td>
<td>1.125</td>
</tr>
<tr>
<td>Jan 87 Family Property Income</td>
<td>1.053</td>
<td>1.016</td>
<td>1.035</td>
</tr>
<tr>
<td>Jan 89 Family Property Income</td>
<td>1.048</td>
<td>1.011</td>
<td>1.033</td>
</tr>
<tr>
<td>Jan 87 Family Means Tested Transfers</td>
<td>1.073</td>
<td>1.016</td>
<td>1.051</td>
</tr>
<tr>
<td>Jan 89 Family Means Tested Transfers</td>
<td>1.055</td>
<td>1.019</td>
<td>1.032</td>
</tr>
<tr>
<td>Jan 87 Family Other Income</td>
<td>1.185</td>
<td>1.006</td>
<td>1.166</td>
</tr>
<tr>
<td>Jan 89 Family Other Income</td>
<td>1.085</td>
<td>1.002</td>
<td>1.076</td>
</tr>
<tr>
<td>Jan 87 Household Income</td>
<td>1.128</td>
<td>1.046</td>
<td>1.073</td>
</tr>
<tr>
<td>Jan 89 Household Income</td>
<td>1.121</td>
<td>1.035</td>
<td>1.076</td>
</tr>
<tr>
<td>Jan 87 Household Earnings</td>
<td>1.154</td>
<td>1.042</td>
<td>1.099</td>
</tr>
<tr>
<td>Jan 89 Household Earnings</td>
<td>1.170</td>
<td>1.035</td>
<td>1.121</td>
</tr>
<tr>
<td>Jan 87 Household Property Income</td>
<td>1.053</td>
<td>1.016</td>
<td>1.035</td>
</tr>
<tr>
<td>Jan 89 Household Property Income</td>
<td>1.048</td>
<td>1.012</td>
<td>1.033</td>
</tr>
<tr>
<td>Jan 87 Household Means Tested Transfers</td>
<td>1.066</td>
<td>1.015</td>
<td>1.045</td>
</tr>
<tr>
<td>Jan 89 Household Means Tested Transfers</td>
<td>1.054</td>
<td>1.019</td>
<td>1.031</td>
</tr>
<tr>
<td>Jan 87 Household Other Income</td>
<td>1.178</td>
<td>1.006</td>
<td>1.160</td>
</tr>
<tr>
<td>Jan 89 Household Other Income</td>
<td>1.084</td>
<td>1.002</td>
<td>1.076</td>
</tr>
<tr>
<td>Jan 87 Labor Force</td>
<td>1.533</td>
<td>1.070</td>
<td>1.407</td>
</tr>
<tr>
<td>Jan 89 Labor Force</td>
<td>1.445</td>
<td>1.028</td>
<td>1.361</td>
</tr>
</tbody>
</table>

s.e. = standard error.
Table 5.12. Estimated means SIPP panel 1987 data and estimated ratio of standard errors for two schemes using cluster totals to estimate regression coefficients.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Estimate Scheme #1 ($)</th>
<th>s.e. Scheme #1</th>
<th>s.e. #2 s.e. #1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 87 Personal Income</td>
<td>986.0</td>
<td>7.66</td>
<td>1.007</td>
</tr>
<tr>
<td>Jan 89 Personal Income</td>
<td>1043.6</td>
<td>7.45</td>
<td>1.006</td>
</tr>
<tr>
<td>Jan 87 Personal Earnings</td>
<td>761.4</td>
<td>7.03</td>
<td>1.008</td>
</tr>
<tr>
<td>Jan 89 Personal Earnings</td>
<td>799.7</td>
<td>6.55</td>
<td>1.006</td>
</tr>
<tr>
<td>Jan 87 Family Income</td>
<td>2708.1</td>
<td>23.22</td>
<td>1.011</td>
</tr>
<tr>
<td>Jan 89 Family Income</td>
<td>2824.6</td>
<td>22.98</td>
<td>1.008</td>
</tr>
<tr>
<td>Jan 87 Family Earnings</td>
<td>2217.4</td>
<td>22.58</td>
<td>1.011</td>
</tr>
<tr>
<td>Jan 89 Family Earnings</td>
<td>2296.6</td>
<td>20.77</td>
<td>1.009</td>
</tr>
<tr>
<td>Jan 87 Family Property Income</td>
<td>148.8</td>
<td>5.24</td>
<td>1.002</td>
</tr>
<tr>
<td>Jan 89 Family Property Income</td>
<td>152.4</td>
<td>5.09</td>
<td>1.002</td>
</tr>
<tr>
<td>Jan 87 Family Means Tested Transfers</td>
<td>31.6</td>
<td>1.67</td>
<td>1.007</td>
</tr>
<tr>
<td>Jan 89 Family Means Tested Transfers</td>
<td>29.3</td>
<td>1.69</td>
<td>1.009</td>
</tr>
<tr>
<td>Jan 87 Family Other Income</td>
<td>310.3</td>
<td>5.52</td>
<td>1.003</td>
</tr>
<tr>
<td>Jan 89 Family Other Income</td>
<td>346.3</td>
<td>8.46</td>
<td>1.002</td>
</tr>
<tr>
<td>Jan 87 Household Income</td>
<td>2775.9</td>
<td>23.52</td>
<td>1.011</td>
</tr>
<tr>
<td>Jan 89 Household Income</td>
<td>2896.2</td>
<td>23.05</td>
<td>1.008</td>
</tr>
<tr>
<td>Jan 87 Household Earnings</td>
<td>2275.6</td>
<td>22.87</td>
<td>1.011</td>
</tr>
<tr>
<td>Jan 89 Household Earnings</td>
<td>2345.9</td>
<td>20.96</td>
<td>1.009</td>
</tr>
<tr>
<td>Jan 87 Household Property Income</td>
<td>150.5</td>
<td>5.25</td>
<td>1.002</td>
</tr>
<tr>
<td>Jan 89 Household Property Income</td>
<td>153.9</td>
<td>5.11</td>
<td>1.002</td>
</tr>
<tr>
<td>Jan 87 Household Means Tested Transfers</td>
<td>32.8</td>
<td>1.81</td>
<td>1.007</td>
</tr>
<tr>
<td>Jan 89 Household Means Tested Transfers</td>
<td>30.3</td>
<td>1.72</td>
<td>1.009</td>
</tr>
<tr>
<td>Jan 87 Household Other Income</td>
<td>316.9</td>
<td>5.73</td>
<td>1.003</td>
</tr>
<tr>
<td>Jan 89 Household Other Income</td>
<td>353.5</td>
<td>8.53</td>
<td>1.002</td>
</tr>
<tr>
<td>Jan 87 Labor Force (0.1%)</td>
<td>464.9</td>
<td>2.18</td>
<td>1.036</td>
</tr>
<tr>
<td>Jan 89 Labor Force (0.1%)</td>
<td>474.4</td>
<td>2.39</td>
<td>1.014</td>
</tr>
</tbody>
</table>

s.e. = standard error.
and the regression coefficients are estimated with these coefficient matrices, the three-phase estimator is uniformly superior to the census-type procedure. The ratios of the standard errors of scheme two to those of the three-phase estimator are similar to the corresponding ratios in Table 5.10. The largest ratio is 1.036 for the January '87 labor force.

Table 5.13 contains the ratios of the estimated standard errors of the three estimators defined in equations (5.62), (5.63), and (5.64) to the standard error of the three-phase estimator based on cluster totals. All ratios are greater than one because the three-phase estimator is uniformly superior to the regression on \( X \) and the two-phase estimator based on \( Y \), to the degree of approximation used in computing the standard errors.

Table 5.14 contains the ratios of the standard error of the three-phase estimator calculated using cluster totals to the standard error of the three-phase estimator calculated using individuals as observations. The standard errors of the three-phase estimator, using cluster totals, are uniformly smaller than those using individuals as observations in the regression. This agrees with the theory which uses regression coefficients computed from cluster totals to minimize the variance of the estimator. The ratios are generally larger for characteristics associated for families. For example, the ratio is 1.045 for January '87 Family Income and is 1.018 for January '87 Personal Income. These ratios suggest that the use of cluster totals in estimation is worth serious consideration. A reduction of five percent in the standard error is equivalent to a ten percent increase in the sample size.
Table 5.13. Ratios of estimated standard errors for alternative estimators using cluster totals to estimate regression coefficients.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean s.e.</th>
<th>Reg. X s.e.</th>
<th>2-Ph. Y s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3-Ph. s.e.</td>
<td>3-Ph. s.e.</td>
<td>3-Ph. s.e.</td>
</tr>
<tr>
<td>Jan 87 Personal Income</td>
<td>1.262</td>
<td>1.054</td>
<td>1.186</td>
</tr>
<tr>
<td>Jan 89 Personal Income</td>
<td>1.258</td>
<td>1.044</td>
<td>1.184</td>
</tr>
<tr>
<td>Jan 87 Personal Earnings</td>
<td>1.275</td>
<td>1.055</td>
<td>1.88</td>
</tr>
<tr>
<td>Jan 89 Personal Earnings</td>
<td>1.302</td>
<td>1.044</td>
<td>1.218</td>
</tr>
<tr>
<td>Jan 87 Family Income</td>
<td>1.189</td>
<td>1.061</td>
<td>1.121</td>
</tr>
<tr>
<td>Jan 89 Family Income</td>
<td>1.166</td>
<td>1.045</td>
<td>1.111</td>
</tr>
<tr>
<td>Jan 87 Family Earnings</td>
<td>1.219</td>
<td>1.057</td>
<td>1.148</td>
</tr>
<tr>
<td>Jan 89 Family Earnings</td>
<td>1.227</td>
<td>1.046</td>
<td>1.163</td>
</tr>
<tr>
<td>Jan 87 Family Property Income</td>
<td>1.060</td>
<td>1.019</td>
<td>1.038</td>
</tr>
<tr>
<td>Jan 89 Family Property Income</td>
<td>1.055</td>
<td>1.015</td>
<td>1.036</td>
</tr>
<tr>
<td>Jan 87 Family Means Tested Transfers</td>
<td>1.129</td>
<td>1.023</td>
<td>1.099</td>
</tr>
<tr>
<td>Jan 89 Family Means Tested Transfers</td>
<td>1.091</td>
<td>1.025</td>
<td>1.063</td>
</tr>
<tr>
<td>Jan 87 Family Other Income</td>
<td>1.211</td>
<td>1.012</td>
<td>1.182</td>
</tr>
<tr>
<td>Jan 89 Family Other Income</td>
<td>1.094</td>
<td>1.005</td>
<td>1.082</td>
</tr>
<tr>
<td>Jan 87 Household Income</td>
<td>1.179</td>
<td>1.062</td>
<td>1.110</td>
</tr>
<tr>
<td>Jan 89 Household Income</td>
<td>1.160</td>
<td>1.046</td>
<td>1.104</td>
</tr>
<tr>
<td>Jan 87 Household Earnings</td>
<td>1.213</td>
<td>1.058</td>
<td>1.141</td>
</tr>
<tr>
<td>Jan 89 Household Earnings</td>
<td>1.222</td>
<td>1.047</td>
<td>1.157</td>
</tr>
<tr>
<td>Jan 87 Household Property Income</td>
<td>1.060</td>
<td>1.020</td>
<td>1.038</td>
</tr>
<tr>
<td>Jan 89 Household Property Income</td>
<td>1.055</td>
<td>1.015</td>
<td>1.036</td>
</tr>
<tr>
<td>Jan 87 Household Means Tested Transfers</td>
<td>1.122</td>
<td>1.022</td>
<td>1.094</td>
</tr>
<tr>
<td>Jan 89 Household Means Tested Transfers</td>
<td>1.091</td>
<td>1.025</td>
<td>1.062</td>
</tr>
<tr>
<td>Jan 87 Household Other Income</td>
<td>1.204</td>
<td>1.012</td>
<td>1.177</td>
</tr>
<tr>
<td>Jan 89 Household Other Income</td>
<td>1.095</td>
<td>1.005</td>
<td>1.083</td>
</tr>
<tr>
<td>Jan 87 Labor Force</td>
<td>1.600</td>
<td>1.092</td>
<td>1.447</td>
</tr>
<tr>
<td>Jan 89 Labor Force</td>
<td>1.481</td>
<td>1.037</td>
<td>1.382</td>
</tr>
</tbody>
</table>

s.e. = standard error.
Table 5.14. Ratios of standard errors of three-phase estimator using cluster totals to those of three-phase estimators using individuals to estimate regression coefficients.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>s.e. Using Cluster Totals</th>
<th>s.e. Using Individual Cluster Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 87 Personal Income</td>
<td>7.66</td>
<td>1.013</td>
</tr>
<tr>
<td>Jan 89 Personal Income</td>
<td>7.45</td>
<td>1.016</td>
</tr>
<tr>
<td>Jan 87 Personal Earnings</td>
<td>7.03</td>
<td>1.017</td>
</tr>
<tr>
<td>Jan 89 Personal Earnings</td>
<td>6.55</td>
<td>1.021</td>
</tr>
<tr>
<td>Jan 87 Family Income</td>
<td>23.22</td>
<td>1.038</td>
</tr>
<tr>
<td>Jan 89 Family Income</td>
<td>22.98</td>
<td>1.032</td>
</tr>
<tr>
<td>Jan 87 Family Earnings</td>
<td>22.58</td>
<td>1.043</td>
</tr>
<tr>
<td>Jan 89 Family Earnings</td>
<td>20.77</td>
<td>1.040</td>
</tr>
<tr>
<td>Jan 87 Family Property Income</td>
<td>5.24</td>
<td>1.008</td>
</tr>
<tr>
<td>Jan 89 Family Property Income</td>
<td>5.09</td>
<td>1.006</td>
</tr>
<tr>
<td>Jan 87 Family Means Nested Transfers</td>
<td>1.67</td>
<td>1.054</td>
</tr>
<tr>
<td>Jan 89 Family Means Nested Transfers</td>
<td>1.69</td>
<td>1.036</td>
</tr>
<tr>
<td>Jan 87 Family Other Income</td>
<td>5.52</td>
<td>1.022</td>
</tr>
<tr>
<td>Jan 89 Family Other Income</td>
<td>8.46</td>
<td>1.009</td>
</tr>
<tr>
<td>Jan 87 Household Income</td>
<td>23.52</td>
<td>1.039</td>
</tr>
<tr>
<td>Jan 89 Household Income</td>
<td>23.05</td>
<td>1.031</td>
</tr>
<tr>
<td>Jan 87 Household Earnings</td>
<td>22.87</td>
<td>1.043</td>
</tr>
<tr>
<td>Jan 89 Household Earnings</td>
<td>20.96</td>
<td>1.039</td>
</tr>
<tr>
<td>Jan 87 Household Property Income</td>
<td>5.25</td>
<td>1.008</td>
</tr>
<tr>
<td>Jan 89 Household Property Income</td>
<td>5.11</td>
<td>1.006</td>
</tr>
<tr>
<td>Jan 87 Household Means Tested Transfers</td>
<td>1.81</td>
<td>1.055</td>
</tr>
<tr>
<td>Jan 89 Household Means Tested Transfers</td>
<td>1.72</td>
<td>1.035</td>
</tr>
<tr>
<td>Jan 87 Household Other Income</td>
<td>5.73</td>
<td>1.023</td>
</tr>
<tr>
<td>Jan 89 Household Other Income</td>
<td>8.53</td>
<td>1.011</td>
</tr>
<tr>
<td>Jan 87 Labor Force (0.1%)</td>
<td>2.18</td>
<td>1.045</td>
</tr>
<tr>
<td>Jan 89 Labor Force (0.1%)</td>
<td>2.39</td>
<td>1.025</td>
</tr>
</tbody>
</table>

s.e. = standard error.
5.5.3 Response Probability Adjustment for Regression Weights

In section 4.5, we discussed the regression estimators with the adjustment for regression weights by the estimated response probabilities. For SIPP data, we use two sets of initial weights in the phase III sample: the weights $w_{ij}^{(0,III)}$ defined in (5.5) and the weights $w_{ijk}^{(0,III)}$ defined in (5.7) which are adjusted by estimated response probabilities. We will calculate the three-phase estimators using these two sets of initial weights, and compare the differences.

The results of the comparison are presented in Table 5.15. The column “Mean with adj.” shows the three-phase estimates for characteristics using initial weights with the response probability adjustment, $w_{ij}^{(0,III)}$ in (5.7). The column “Mean without adj.” is computed without the adjustment of response probability. The column “t-test” gives the t-statistics for testing the effects of nonresponse probability adjustment on the mean estimators. The t-statistics are computed as follows: we calculate the difference between the final weights with and without adjustment for the response probabilities, then we use this difference of weights as the weight to calculate the variance of the difference of the means in the column “Mean with adj.” and “Mean without adj.”

The effect of response probability adjustment is significant for “Labor Force” but is not significant for other characteristics. This may be due to the fact that the regression variables have produced adjustments equivalent to the response probability adjustment for most variables. Table 5.15 also presents the estimated standard errors for three-phase estimators.
Table 5.15. Comparison between three-phase estimators with and without response probability adjustment.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean with adj. ($)</th>
<th>Mean without adj. ($)</th>
<th>t-test</th>
<th>s.e. with adj.</th>
<th>s.e. without adj.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 87 Personal Income</td>
<td>971.4</td>
<td>986.0</td>
<td>-1.37</td>
<td>7.49</td>
<td>7.57</td>
</tr>
<tr>
<td>Jan 89 Personal Income</td>
<td>1031.2</td>
<td>1042.5</td>
<td>-1.07</td>
<td>7.42</td>
<td>7.49</td>
</tr>
<tr>
<td>Jan 87 Personal Earnings</td>
<td>753.2</td>
<td>757.2</td>
<td>-0.41</td>
<td>6.90</td>
<td>6.93</td>
</tr>
<tr>
<td>Jan 89 Personal Earnings</td>
<td>794.4</td>
<td>795.3</td>
<td>-0.10</td>
<td>6.54</td>
<td>6.57</td>
</tr>
<tr>
<td>Jan 87 Family Income</td>
<td>2716.7</td>
<td>2725.7</td>
<td>-0.28</td>
<td>23.11</td>
<td>22.97</td>
</tr>
<tr>
<td>Jan 89 Family Income</td>
<td>2818.4</td>
<td>2842.0</td>
<td>-0.72</td>
<td>22.96</td>
<td>23.08</td>
</tr>
<tr>
<td>Jan 87 Family Earnings</td>
<td>2231.3</td>
<td>2227.9</td>
<td>0.11</td>
<td>22.48</td>
<td>22.30</td>
</tr>
<tr>
<td>Jan 89 Family Earnings</td>
<td>2297.1</td>
<td>2307.1</td>
<td>-0.34</td>
<td>20.74</td>
<td>20.78</td>
</tr>
<tr>
<td>Jan 87 Family Property Income</td>
<td>144.8</td>
<td>152.6</td>
<td>-1.05</td>
<td>5.18</td>
<td>5.35</td>
</tr>
<tr>
<td>Jan 89 Family Property Income</td>
<td>147.9</td>
<td>155.0</td>
<td>-0.98</td>
<td>5.06</td>
<td>5.22</td>
</tr>
<tr>
<td>Jan 87 Family MTT</td>
<td>32.7</td>
<td>31.0</td>
<td>-0.73</td>
<td>1.67</td>
<td>1.62</td>
</tr>
<tr>
<td>Jan 89 Family MTT</td>
<td>30.3</td>
<td>29.2</td>
<td>0.47</td>
<td>1.66</td>
<td>1.62</td>
</tr>
<tr>
<td>Jan 87 Family Other Income</td>
<td>307.9</td>
<td>314.1</td>
<td>-0.79</td>
<td>5.50</td>
<td>5.60</td>
</tr>
<tr>
<td>Jan 89 Family Other Income</td>
<td>343.1</td>
<td>350.6</td>
<td>-0.62</td>
<td>8.48</td>
<td>8.71</td>
</tr>
<tr>
<td>Jan 87 Household Income</td>
<td>2805.8</td>
<td>2793.0</td>
<td>0.39</td>
<td>23.45</td>
<td>23.26</td>
</tr>
<tr>
<td>Jan 89 Household Income</td>
<td>2899.3</td>
<td>2913.2</td>
<td>-0.43</td>
<td>23.06</td>
<td>23.14</td>
</tr>
<tr>
<td>Jan 87 Household Earnings</td>
<td>2307.5</td>
<td>2285.4</td>
<td>0.69</td>
<td>22.82</td>
<td>22.59</td>
</tr>
<tr>
<td>Jan 89 Household Earnings</td>
<td>2354.5</td>
<td>2357.0</td>
<td>-0.08</td>
<td>20.98</td>
<td>20.97</td>
</tr>
<tr>
<td>Jan 87 Household Property Income</td>
<td>146.8</td>
<td>154.4</td>
<td>-1.02</td>
<td>5.19</td>
<td>5.36</td>
</tr>
<tr>
<td>Jan 89 Household Property Income</td>
<td>149.6</td>
<td>156.5</td>
<td>-0.95</td>
<td>5.68</td>
<td>5.24</td>
</tr>
<tr>
<td>Jan 87 Household MTT</td>
<td>34.6</td>
<td>32.4</td>
<td>0.87</td>
<td>1.83</td>
<td>1.75</td>
</tr>
<tr>
<td>Jan 89 Household MTT</td>
<td>31.5</td>
<td>30.1</td>
<td>0.59</td>
<td>1.70</td>
<td>1.65</td>
</tr>
<tr>
<td>Jan 87 Household Other Income</td>
<td>316.9</td>
<td>320.8</td>
<td>-0.48</td>
<td>5.74</td>
<td>5.78</td>
</tr>
<tr>
<td>Jan 89 Household Other Income</td>
<td>350.9</td>
<td>357.9</td>
<td>-0.57</td>
<td>8.55</td>
<td>8.77</td>
</tr>
<tr>
<td>Jan 87 Labor Force (0.1%)</td>
<td>466.4</td>
<td>458.4</td>
<td>2.60</td>
<td>2.20</td>
<td>2.15</td>
</tr>
<tr>
<td>Jan 89 Labor Force (0.1%)</td>
<td>478.8</td>
<td>469.4</td>
<td>2.77</td>
<td>2.45</td>
<td>2.35</td>
</tr>
</tbody>
</table>

s.e. = standard error.

MTT = Means Tested Transfers
using two sets of initial weights, \( w_{ijk}^{(0,III)} \) and \( w_{ij}^{(0,III)} \) in (5.7) and (5.5). The column “s.e. with adj.” gives estimated standard errors calculated by weights \( w_{ij}^{(0,III)} \) which are adjusted by estimated response probabilities, and the column “s.e. without adj.” presents estimated standard errors calculated by weights \( w_{ij}^{(0,III)} \) without the adjustment from response probabilities. There is very little difference between the two sets of estimated standard errors for three-phase estimators.
APPENDIX A: NONNEGATIVE REGRESSION ESTIMATION FOR SAMPLE SURVEY DATA

A.1 Introduction

In some situations, data available from outside sources can be used to improve the estimates derived from a sample. An estimation procedure that incorporates the known means (or totals) of auxiliary variables into the estimation procedure is regression estimation. We consider regression estimation given that the population means of continuous (and) or categorical data for several characteristics of the population are available.

Suppose that the population means \((\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)\) of \(p\) auxiliary variables \((X_1, X_2, \ldots, X_p)\) are known. Let a sample of \(n\) observations be available and suppose that we have a weight \(v_i\) for each observation such that \(v_i > 0\) for all \(i\), \(\sum_{i=1}^{n} v_i = n\), and \(n^{-1} \sum_{i=1}^{n} v_i y_i\) is unbiased for the population mean of \(Y\). All summations in this paper are over the sample elements and in the sequel we omit the range of summation. Let

\[
X = \begin{pmatrix}
X_{11} & X_{12} & \cdots & X_{1p} \\
X_{21} & X_{22} & \cdots & X_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
X_{n1} & X_{n2} & \cdots & X_{np}
\end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \pi^{(0)} = n^{-1} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.
\]

Then a regression estimator of the population mean \(\bar{Y}\) is

\[
\bar{y}_{GL} = \bar{y}_w + (\bar{X} - \bar{X}_w)' b_G = w'y
\]
where
\[
\begin{align*}
\bar{X} &= (\bar{X}_1, \ldots, \bar{X}_p)' \\
\bar{X}_w &= X' \pi^{(0)} \\
\bar{y}_w &= \pi^{(0)'}y \\
b_G &= (X'Q'VQX)^{-1} [X'Q'Vy] \\
Q &= (I - n^{-1}J_{(n)}J_{(n)}') \\
w &= (w_1, \ldots, w_n)' = \pi^{(0)} + VQX(X'Q'VQX)^{-1}(\bar{X} - \bar{X}_w)
\end{align*}
\]

The weights have the property that
\[
w'X = \bar{X}'.
\] (A.4)

In the single-\(X\) case with all \(v_i = 1\), the regression estimator (A.1) reduces to the ordinary regression estimator \(\bar{y}_i\).

The regression weight \(w_i\) for the \(i\)-th sample observation given in (A.3) may be negative. Such weights could then produce negative estimates of population means known to be positive. Fuller (1968) used grouping methods to insure positive regression weights \(w_i\). Husain (1969) used a quadratic programming approach to obtain positive regression weights \(w_i\). We describe a computer algorithm designed to provide nonnegative weights \(w_i\) for a regression estimator of the form \(\sum w_i y_i\). The procedure builds upon that developed by Huang (1978) and described in Huang and Fuller (1978).
A.2 The Computer Algorithm for Regression M-Weights

Because the weight function $w_i$ for the regression estimator $\bar{y}_{GL}$ is a linear function of $X_i$, where $X_i$ is the $i$-th row of $X$, it is possible for some $w_i$ to be negative. The objective of the computer algorithm is to produce weights that are positive and that fall within a specified range. The algorithm is iterative. At each step, the algorithm produces weights that are a smooth, continuous, monotone increasing function of the weights of the previous step. Also at each step, the weights satisfy equality (A.1).

The weights are checked against user supplied criteria at each step. One user supplied parameter, denoted by $M$, is, approximately, the maximum fraction by which any weight can deviate from the previous weight at any iteration of the program. Thus, if one sets $M = 0.75$, at each step the program attempts to construct weights such that all weights are between 0.25 and 1.75 of the original weight. Note that choosing $M = 1$ is equivalent to requiring positive weights. If $M$ is not supplied, a default value of 0.8 is in the program.

In addition, the weights are checked against bound criteria. An upper bound $U_M$ and a lower bound $L_M$ are provided by the user (the default bounds are 0 and 100), where $0 \leq L_M < 1 < U_M$. The bounds are expressed as a multiple of the average weight. At each step, the program checks to see if there is any weight such that the ratio of the weight to the average of weights falls outside the bounds. If the criteria are not met, the program moves to another iteration.
It will not always be possible to construct weights satisfying the specified restriction. If the sample is such that the restriction cannot be met, the program will produce weights "approximating" the criterion. In the single $x$ case, when the criterion cannot be satisfied, there will be two weights, one for those greater than the population mean, and one for those less than the population mean.

An iteration of the program consists of computing the generalized least squares weights, where an adjusting factor $g_i$ is applied to each observation. The $g_i$ is a "bell" shaped function of the distance (in a suitable metric) that the observation is from the population mean.

In addition to the array of observations and the population means, two initial factors, $v_i$ and $g_i^{(0)}$, $i = 1, 2, ..., n$, are required to initiate the computations. The $v_i$ are typically inversely proportional to the probabilities of selection. The default values for $g_i^{(0)}$ are $g_i^{(0)} = 1$. For stratified samples or data with unequal variances, the user may choose other values. (See Huang (1978), and Goebel (1976)). The program input includes the sample size $n$, the population size $N$, the parameter $M$, the maximum number of iterations permitted $LI$, the upper bound of the ratios of weights to the average weight $U_M$, and the lower bound of the ratios, $L_M$. It is required that $0 < L_M < 1 < U_M$.

To describe the algorithm, let

$$Z_{ij} = X_{ij} - \bar{X}_j, \quad \text{(A.5)}$$
The algorithm for computing modified regression weights for the estimated mean is composed of the following steps:

1. Set $Q = 0$. The initial generalized regression weights, denoted by $w^{(0)}$, are

$$w^{(0)} = \left[1 + n \bar{u}_w^{(0)}\right]^{-1} P^{(0)} \left(n^{-1} J(u) + u^{(0)}\right) = (w_1^{(0)}, \ldots, w_n^{(0)})',$$  

(A.12)

where

$$u^{(0)} = G^{(0)} Z \left(A^{(0)}\right)^+ \left(\bar{X} - \bar{X}_w\right) = (u_1^{(0)}, \ldots, u_n^{(0)})'.$$  

(A.13)

$\left(A^{(0)}\right)^+$ is a symmetric generalized inverse of $A^{(0)}$ and

$$n \bar{u}_w^{(0)} = \max \left\{J'_n, P^{(0)} u^{(0)}, n^{-1} - 1\right\}.$$  

(A.14)
2. Check if $|n u_i^{(\alpha)}| > M$ for each $i$.

   If $|n u_i^{(\alpha)}| > M$ for some $i$, check the signs of the $u_i^{(\alpha)}$.

   If all $u_i^{(\alpha)}$ are of the same sign, the program terminates and prints, "Check population totals, if data are correct, nonnegative weights cannot be constructed."

   If all $u_i^{(\alpha)}$ are not of the same sign, set $\alpha = \alpha + 1$.

   If $LI$ iterations haven't been completed, go to step (3). If $LI$ iterations have been completed, output the weights $w^{(\alpha)}$.

   If $|n u_i^{(\alpha)}| \leq M$ for all $i$, check if $L_M < \frac{n w_i^{(\alpha)}}{\sum_w^{(\alpha)}} < U_M$.

   If this inequality holds for all $i$, then output the weight vector $w^{(\alpha)}$.

   If this inequality fails for any $i$, then set $\alpha = \alpha + 1$.

   If $LI$ iterations haven't been completed, go to step (3). If $LI$ iterations have been completed, output the weight vector $w^{(\alpha)}$.

   If requested, also output the estimate

   $$\bar{y}_{GL}^{(\alpha)} = w^{(\alpha)'y}$$  \hspace{1cm} \text{(A.15)}

3. Define

   $$e_i^{(\alpha)} = |M^{-1} n u_i^{(\alpha-1)}|$$  \hspace{1cm} \text{(A.16)}
\[
c_i^{(a)} = \begin{cases} 
1 & 0 \leq e_i^{(a)} \leq \frac{1}{2} \\
1 - 0.8 \left(e_i^{(a)} - 0.5\right)^2 & \frac{1}{2} < e_i^{(a)} \leq 1 \\
0.8 \left(e_i^{(a)}\right)^{-1} & e_i^{(a)} > 1
\end{cases} \tag{A.17}
\]

\[
a_i^{(a)} = c_i^{(a)} u_i^{(a-1)} \tag{A.18}
\]

\[
\tilde{a}_w^{(a)} = n^{-1} \sum_{i=1}^{n} P_i^{(a-1)} a_i^{(a)} \tag{A.19}
\]

\[
b_i^{(a)} = \begin{cases} 
d_i^{(a)} - \tilde{a}_w^{(a)} & \text{if } d_i^{(a)} - \tilde{a}_w^{(a)} > -0.8n^{-1} \\
-0.8n^{-1} & \text{if } d_i^{(a)} - \tilde{a}_w^{(a)} \leq -0.8n^{-1}
\end{cases} \tag{A.20}
\]

\[
\tilde{b}_w^{(a)} = n^{-1} \sum_{i=1}^{n} q_i^{(a-1)} b_i^{(a)} \tag{A.21}
\]

\[
q_i^{(a)} = P_i^{(a-1)} \left[n^{-1} + b_i^{(a)}\right] \left[1 + n \tilde{b}_w^{(a)}\right]^{-1} \tag{A.22}
\]

\[
g_i^{(a)} = (M - n c_i^{(a)} u_i^{(a-1)}) M^{-1} = 1 - c_i^{(a)} c_i^{(a)} \tag{A.23}
\]

\[
L_B = \begin{cases} 
n^{-1} [L_M + 0.07] & \text{if } 0 \leq L_M \leq 0.8 \\
n^{-1} [1 - 0.6 (1 - L_M)] & \text{if } 0.8 < L_M < 1
\end{cases} \tag{A.24}
\]
Compute the regression weights

\[ w^{(a)} = \left[ 1 + n\tilde{w}^{(a)} \right]^{-1} P^{(a)} \left( n^{-1}J(n) + u^{(a)} \right) = \left( w_1^{(a)} , \ldots , w_n^{(a)} \right)' , \]  

where

\[ u^{(a)} = G^{(a)}Z \left( A^{(a)} \right)^+ \left( \tilde{X} - \tilde{X}_w^{(a)} \right) = \left( u_1^{(a)} , \ldots , u_n^{(a)} \right)' , \]  

\[ A^{(a)} = ZP^{(a)}H^{(a)}Z , \]  

\[ H^{(a)} = \prod_{t=0}^{t=a} G^{(a)} , \]  

\[ n\tilde{w}^{(a)} = \max \{ \pi^{(a)}u^{(a)} , n^{-1} - 1 \} . \]
The regression weights $w_i^{(a)}$, $i = 1, 2, ..., n$ satisfy

$$\sum_{i=1}^{n} w_i^{(a)} X_{ij} = \bar{X}_j,$$

$$\sum_{i=1}^{n} w_i^{(a)} = 1,$$

provided $n \bar{u}_w^{(a)} > n^{-1} - 1$. Go to Step 2.

Note that $M$ must be chosen with some care. If $M$ is chosen very small, it will be impossible to find weights to meet the restriction. For most practical problems, $\frac{1}{2} < M < (N - n) N^{-1}$ seems reasonable.

If the data set is such that the input requirements are met, the regression weights, $\{w_i^{(a)}\}$ have the following properties:

$$w_i^{(a)} \geq 0 \text{ for } i = 1, 2, ..., n,$$  \hfill (A.36)

$$\sum_{i=1}^{n} w_i^{(a)} X_{ij} = \bar{X}_j,$$  \hfill (A.37)

$$\sum_{i=1}^{n} w_i^{(a)} = 1,$$  \hfill (A.38)

$$\max_i n^{-1} w_i^{(a)} \leq U_M,$$  \hfill (A.39)

$$\min_i n^{-1} w_i^{(a)} \geq L_M.$$  \hfill (A.40)

The regression weights computer algorithm will produce positive weights under certain data configurations of the auxiliary variables. In the single $x$ case, the weights $w_i^{(a)}$ will be positive if and only if there exists at least one $x_i$ greater than the population mean $\bar{X}$ and one $x_i$ less than $\bar{X}$. 
A.3 Asymptotic Properties of the Regression Type Estimator

In this section, we present the asymptotic properties of the regression estimator defined by the regression weight generation program. For simplicity, we give results only for the single $x$ case.

Lemma A1 (Huang and Fuller (1978)) Let $(x_i, y_i)$ be a sequence of independent identically distributed random vectors with mean vector $(\mu_x, \mu_y)$, and finite covariance matrix with $\sigma^2_x > 0$. Let

$$I_n(x_i) = \begin{cases} 1 & \text{if } |x_i - \mu_x| < (n\lambda)^{1/2}, \\ 0 & \text{otherwise}, \end{cases}$$  \hspace{1cm} (A.41)

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \mu_x) y_i}{\sum_{i=1}^n (x_i - \mu_x)^2},$$  \hspace{1cm} (A.42)

$$\tilde{\beta} = \frac{\sum_{i=1}^n I_n(x_i) (x_i - \mu_x) y_i}{\sum_{i=1}^n I_n(x_i) (x_i - \mu_x)^2},$$  \hspace{1cm} (A.43)

where $\lambda > 0$ is a fixed constant. If the denominator in $\hat{\beta}$ or $\tilde{\beta}$ is zero, the estimator is defined to be zero. Then:

$$\left( \hat{\beta} - \tilde{\beta} \right) \overset{a.s.}{\longrightarrow} 0.$$ 

**Proof.** We assume, without loss of generality, that $\mu_x = 0$. Using

$$P(x_k^2 I_k(x_k) \neq x_k^2) = P(x_k^2 \geq k\lambda) = P(x_1^2 \geq k\lambda), \hspace{1cm} (A.44)$$
Lemma 4 of Tucker (1967), p. 123), and the assumption that $\sigma_x^2$ is finite, we have

$$\sum_{k=0}^{\infty} P\left(x_k^2 \geq k\lambda\right) \leq \left[E\left(x_k^2\right) + \lambda\right] \lambda^{-1} < \infty. \quad (A.45)$$

By the Borel-Cantelli lemma,

$$P\left\{x_k^2 I_k (x_k) \neq x_k^2 \ i.o.\right\} = 0. \quad (A.46)$$

Also,

$$\lim_{n \to \infty} \sum_{k=1}^{n} P\left(x_k^2 I_k (x_k) \neq x_k^2\right) = \lim_{n \to \infty} \sum_{k=1}^{n} P\left(x_k^2 \geq n\lambda\right) \leq \sum_{k=1}^{\infty} P\left(x_k^2 \geq k\lambda\right) < \infty.$$ 

Therefore, by the Borel-Cantelli lemma, $P\left(x_k^2 I_k (x_k) \neq x_k^2 \ i.o.\right) = 0$, and by Theorem 5.2.1. of Chung (1974, p. 108), we have

$$n^{-1} \sum_{k=1}^{n} \left[x_k^2 I_n (x_k) - x_k^2\right] \xrightarrow{a.s.} 0. \quad (A.47)$$

Hence,

$$n^{-1} \sum_{k=1}^{n} x_k^2 I_n (x_k) \xrightarrow{a.s.} \sigma_x^2.$$ 

By a similar argument, we can prove

$$n^{-1} \sum_{i=1}^{n} x_i y_i I_n \xrightarrow{a.s.} \sigma_{xy}.$$ 

Therefore, $\tilde{\beta}_t \xrightarrow{a.s.} \beta$. 
Theorem A.1. (Huang an Fuller (1978)). Let \((x_i, y_i)\) satisfy the hypothesis of Lemma A1. and let

\[
g(d_i) = \begin{cases} 
  1 & 0 \leq d_i < \frac{1}{2}, \\
  1 - 0.8 \left( d_i - \frac{1}{2} \right)^2 & \frac{1}{2} \leq d_i \leq 1, \\
  0.8d_i^{-1} & d_i > 1,
\end{cases}
\]

\[
\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \mu_x) y_i}{\sum_{i=1}^{n} (x_i - \mu_x)^2},
\]

\[
\tilde{\beta}_g = \frac{\sum_{i=1}^{n} g(d_i) (x_i - \mu_x) y_i}{\sum_{i=1}^{n} g(d_i) (x_i - \mu_x)^2},
\]

where

\[
d_i = s_x^{-2} |(\mu_x - \bar{x}) (x_i - \mu_x)|,
\]

\[
\bar{x} = n^{-1} \sum_{i=1}^{n} x_i,
\]

\[
s_x^2 = n^{-1} \sum_{i=1}^{n} (x_i - \mu_x)^2.
\]

Then, \(\hat{\beta} \xrightarrow{a.s.} \beta\).

Proof. Without loss of generality, we assume that \(\mu_x = 0\). We note that \(d_i\) can be rewritten as

\[
d_i = \frac{-n^{1/2} \bar{x}}{\sigma_x} \left| \frac{x_i}{n^{1/2} \sigma_x} \right|.
\]
Clearly,

\[
\left[ (\sigma_x^2)^{-1} n \right]^{1/2} \bar{x} = O_p(1),
\]

\[
(\sigma_x^2)^{-1} \frac{s_x^2}{n} = O_p(1).
\]

and

\[
(\sigma_x^2 n)^{-1/2} x_i = O_p\left( n^{-1/2} \right).
\]

Therefore, given \( \epsilon > 0 \), there exists an \( N_\epsilon \) and an \( M_\epsilon > 0 \) such that for \( n > N_\epsilon \)

\[
P\left\{ n^{1/2} \left( s_x^2 \right)^{-1} \sigma_x \bar{x} \leq M_\epsilon \right\} \geq 1 - \epsilon.
\]

If

\[
\left\{ \left[ n (\sigma_x^2)^{-1} \right]^{1/2} \bar{x} \left( s_x^2 \right)^{-1} \sigma_x^2 \leq M_\epsilon, \right\}
\]

and

\[
(n \sigma_x^2)^{-1/2} x_i < (2M_\epsilon)^{-1},
\]

we have \( d_i < \frac{1}{2} \), and \( g(d_i) = 1 \). Setting

\[
\lambda^{1/2} = (2M_\epsilon)^{-1} \sigma_x,
\]

the probability is greater than \( 1 - \epsilon \) that \( g(d_i) = 1 \) when \( I_n(x_i) = 1 \), where \( I_n(x_i) \) was defined in Lemma A.1. Now \( x_i^2 \geq 0 \), and \( 1 \geq g(d_i) > 0 \). Hence,

\[
P\left\{ n^{-1} \sum_{i=1}^{n} [1 - g(d_i)] x_i^2 \leq n^{-1} \sum_{i=1}^{n} [1 - I_n(x_i)] x_i^2 \right\} \geq 1 - \epsilon
for $n > N$. By a similar argument,

$$P \left\{ n^{-1} \sum_{i=1}^{n} |1 - g(d_i)| |x_i y_i| \leq n^{-1} \sum_{i=1}^{n} |1 - I_n(x_i)| |x_i y_i| \right\} \geq 1 - \epsilon.$$ 

Since

$$\left| n^{-1} \sum_{i=1}^{n} (1 - g(d_i)) x_i y_i \right| \leq n^{-1} \sum_{i=1}^{n} (1 - g(d_i)) |x_i y_i|,$$

and since, by Lemma A.1,

$$n^{-1} \sum_{i=1}^{n} (1 - I_n(x_i)) x_i^2 \overset{a.s.}{\longrightarrow} 0,$$

and

$$n^{-1} \sum_{i=1}^{n} (1 - I_n(x_i)) |x_i y_i| \overset{a.s.}{\longrightarrow} 0.$$

we have $\tilde{\beta}_g \overset{a.s.}{\longrightarrow} \beta.$

In Theorem A.1, we used a function $g(d_i)$ closely related to the function $g(d_i)$ of the computer algorithm. Note that the theorem holds for any function that is one for all $d_i < k$, where $k$ is a positive constant.

Therefore, the asymptotic distribution of the regression estimator constructed with the computer algorithm is the same as that of the ordinary regression estimator $\tilde{y}_{1w}$.

### A.4 Guidelines for Using the Weight Generating Program

Associated with the algorithm in Section 2, a package of programs written in the FORTRAN language is available. There are 3 files in the package: weight.for, weight.exe, and weight.dat.
weight.for: The FORTRAN source program for generating the nonnegative weights. This file is only a reference file for users that are going to use the file “weight.exe” directly.

weight.exe: The execute file which has been compiled. Users can use their own FORTRAN compiler to compile the file “weight.for” to create this file.

weight.dat: The file needs to be edited to use the program. When users execute “weight.exe”, the program will read data and parameters from this file.

STEPS TO USE THIS PACKAGE

Step 1. Use a text editor (for example: emacs, edit etc.) to open the file “weight.dat”.

The file “weight.dat” gives the template of data files required by the program. In order to execute the program successfully, we need to create a data file using the format provided by “weight.dat.” In editing the file “weight.dat,” we replace those blank lines following the instruction lines using the format offered at the end of each instruction line, or leave them blank if it is acceptable. The formats such as I5, F10.5, go by the syntax of the FORTRAN language. Do this till the line (F15.2) is reached.
Step 2. Enter totals or means of the independent variables \((X_1, X_2, ..., X_p)\) following the line

\((F15.2)\)

Each total or mean occupies one line.

Step 3. Input ID, independent variables, initial weights etc., starting with a line which describes the formats of the data. For example, it might look like

\((12, \, 1F3.0, \, 1F2.0, \, 1F1.0)\)

01100011
02100021
03100031
04100041
...

Step 4. Edit the line at the end of the file which describes the format of the output for ID and weights generated by the program. If the format \((1X, (17, \, 1X, \, F15.2))\) satisfies the requirement, leave the last line unchanged.

Step 5. Save the edited file "weight.dat".

Step 6. Execute the program by command "weight.exe".

Step 7. After the program is finished, the result will be saved in the file "weight.res".
A.5 An Example

As an example, we construct weights for a simple random sample of size 15 selected from a population with size 1,500. The parameters for the regression $M$-weights program in this example are: $\bar{X}_1 = 8000/1500$, $\bar{X}_2 = 900/1500$, $N = 1500$, $n = 15$, $r_1 = 100$, $g_1^{(o)} = 1$, $M = 0.9$, $L_M = 0.15$, $U_M = 4.5$. The weights meet the restriction after 7 iterations.

(Note: The regression $M$-weights for estimating the total are $w_i^{(\alpha)} = NW_i^{(\alpha)}$, $i = 1, ..., n$.)

The edited file "weight.dat" after Step 5 in Section 4 is:

Problem Identifier (2A8)

EXAMPLE DATA - 3/17/93

Enter number of observation in the sample (15)

15

Enter the population size (F20.10)

1500.

Max. fractional deviation for weights to differ from average (F20.10)

(If the following line is left blank, the default value of .8 will be supplied.)

0.90

Enter the lower relative bound for weight (F20.10)

(If the following line is left blank, then the default of 0 will be supplied.)

0.15
Enter the upper relative bound for weight (f20.10)

(If the following line is left blank, then the default of 100 will be supplied.)

4.5

Number of independent variables (I5)

2

Number of dependent variables (I5)

0

Unit number for input in "open" or "read" statements (I5)

5

Unit number for output in "write" statements (I5)

6

If an optional output device is specified, record MMGP=number

of weights (with corresponding identifications) that are to be

included on each record. Leave blank if the output is to be written only on line printer

(I5)

1

Enter 1 if population totals are to be input.

Leave BLANK if population means will be supplied. (I5)

1

Enter 1 if initial weights for each observation are to be input.
Leave BLANK if each initial weight is 1. (I5)

1

If stratified sample weights are to be input, enter 1.
Leave BLANK if the data come from a simple random sample. (I5)

Enter 1 if weights for the means are desired.
Leave BLANK if weights for the totals are desired. (I5)

Enter the maximum number of iterations allowable.
Leave BLANK for default value of 7. (I5)

10

Enter 1 if real weights are desired.
Leave BLANK if integer weights are desired. (I5)

1

Enter 1 if ID numbers augmented by a group number are desired.
Leave BLANK if otherwise. (I5)

(F15.2)
8000.
900.
(I2, 1F3.0, 1F2.0, 1F1.0)
01100011
02100021
03100031
04100041
05100051
06100061
07100071
08100081
09100091
10100101
11100000
12100000
13100000
14100000
15100000

(1X, (17, 1X, F15.2))

The output for this data (saved in "weight.res") is:

problem identification— TEST - 5/3/91
sample size = 15
population size = 1500.000
number of \( x \) vectors = 2
number of \( y \) vectors = 0
the lower relative bound for weight \( L_m \) = 0.15
the upper relative bound for weight \( U_m \) = 4.50

Population totals:

1 8000.00000 2 900.00000

Population means:

1 5.333333 2 0.600000

The computed weights along with their identification:

<table>
<thead>
<tr>
<th>id</th>
<th>weights</th>
<th>id</th>
<th>weights</th>
<th>id</th>
<th>weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.18594</td>
<td>2</td>
<td>15.13473</td>
<td>3</td>
<td>15.12907</td>
</tr>
<tr>
<td>4</td>
<td>15.15531</td>
<td>5</td>
<td>15.13482</td>
<td>6</td>
<td>15.21100</td>
</tr>
<tr>
<td>7</td>
<td>15.17268</td>
<td>8</td>
<td>15.19006</td>
<td>9</td>
<td>332.99722</td>
</tr>
<tr>
<td>10</td>
<td>445.68922</td>
<td>11</td>
<td>120.00001</td>
<td>12</td>
<td>120.00001</td>
</tr>
<tr>
<td>13</td>
<td>120.00001</td>
<td>14</td>
<td>120.00001</td>
<td>15</td>
<td>120.00001</td>
</tr>
</tbody>
</table>
The 2 population totals for $x$ are:

8000.00000  900.00000

The 2 estimated totals for $x$ are:

8000.00036  900.00004

number of iterations needed = 7

the critical value, $m$, = 0.90000

the sum of the weights = 1500.00007

1  15.19
2  15.13
3  15.13
4  15.16
5  15.13
6  15.21
7  15.17
8  15.19
9  333.00
10  445.69
11  120.00


ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my major professor Dr. Wayne A. Fuller for his immense contributions and guidance at every facet of the research. I should not have been able to complete this research without his constant help.

I would like to thank other members of my committee: Dr. Krishna Athreya, Dr. F. Jay Breidt, Dr. Wolfgang H. Kliemann, Dr. Soumendra N. Lahiri, and Dr. Kenneth J. Koehler for their useful suggestions to my dissertation.

Thanks are due to Dr. Dean L. Isaacson, Dr. Paul N. Hinz, Dr. Alicia L. Carriquiry, Dr. Roy D. Hickman, and Dr. Sarah M. Nusser for their support and help throughout my graduate study at Iowa State University.

I would also like to thank Sandie Smith, Helen Nelson, Rose Ann Anderson, and the entire faculty and staff of the Statistics Department, particularly members in the Survey Section, for making my stay here enjoyable, and for all the help they provided.

I thank Judy Shafer for typing the manuscript patiently and wonderfully in a timely manner.

I am grateful to Dr. Robert Rodriguez, Dr. Donna Watts, Dr. Alan Eaton, and the SAS Institute for their constant support at the final stage of my dissertation work.

The research was partly supported by The Liberal Arts and Science College at Iowa State University, and by the joint statistical agreement with the U.S. Bureau of the Census, U.S. Department of Commerce.