1

**PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT**

PRECISIONSM RESOLUTION TARGETS

Contributions to Poststratification in Stratified Multi-Stage Samples

by

Wesley Thomas Yung, B.Sc., M.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of

Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University
Ottawa, Ontario
December, 1995

© Copyright
1996, Wesley Yung

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-13772-4

Canada

Name WESLEY YOUNG

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

STATISTICS

SUBJECT TERM

0463

SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0277
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0277
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language
General 0679
Ancient 0289
Linguistics 0290
Modern 0291
Literature
General 0401
Classical 0294
Comparative 0295
Medieval 0297
Modern 0298
African 0316
American 0591
Asian 0305
Canadian (English) 0352
Canadian (French) 0355
English 0593
Germanic 0311
Latin American 0312
Middle Eastern 0315
Romance 0313
Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
General 0318
Biblical Studies 0321
Clergy 0319
History of 0320
Philosophy of 0322
Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
Archaeology 0324
Cultural 0326
Physical 0327
Business Administration
General 0310
Accounting 0272
Banking 0770
Management 0454
Marketing 0338
Canadian Studies 0385
Economics
General 0501
Agricultural 0503
Commerce-Business 0505
Finance 0508
History 0509
Labor 0510
Theory 0511
Folklore 0358
Geography 0366
Gerontology 0351
History
General 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0228
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science
General 0615
International Law and Relations 0616
Public Administration 0617
Recruitment 0814
Social Work 0452
Sociology
General 0626
Criminology and Penology 0627
Demography 0938
Ethnic and Racial Studies 0631
Individual and Family Studies 0628
Industrial and Labor Relations 0629
Public and Social Welfare 0630
Social Structure and Development 0700
Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
General 0473
Agronomy 0285
Animal Culture and Nutrition 0475
Animal Pathology 0476
Food Science and Technology 0359
Forestry and Wildlife 0478
Plant Culture 0479
Plant Pathology 0480
Plant Physiology 0817
Range Management 0777
Wood Technology 0746
Biology
General 0306
Anatomy 0287
Biostatistics 0308
Botany 0309
Cell 0379
Ecology 0329
Entomology 0353
Genetics 0369
Limnology 0793
Microbiology 0410
Molecular 0307
Neuroscience 0317
Oceanography 0416
Physiology 0433
Radiation 0821
Veterinary Science 0778
Zoology 0472
Biophysics
General 0786
Medical 0760

EARTH SCIENCES

Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
General 0566
Audiology 0300
Chemotherapy 0992
Dentistry 0567
Education 0350
Hospital Management 0769
Human Development 0758
Immunology 0982
Medicine and Surgery 0564
Mental Health 0347
Nursing 0569
Nutrition 0570
Obstetrics and Gynecology 0380
Occupational Health and Therapy 0354
Ophthalmology 0381
Pathology 0571
Pharmacology 0419
Pharmacy 0572
Physical Therapy 0382
Public Health 0573
Radiology 0574
Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences
Chemistry
General 0485
Agricultural 0749
Analytical 0486
Biochemistry 0487
Inorganic 0488
Nuclear 0738
Organic 0490
Pharmaceutical 0491
Physical 0494
Polymer 0495
Radiation 0754
Mathematics 0405
Physics
General 0605
Acoustics 0986
Astronomy and Astrophysics 0606
Atmospheric Science 0608
Atomic 0748
Electronics and Electricity 0607
Elementary Particles and High Energy 0798
Fluid and Plasma 0759
Molecular 0609
Nuclear 0610
Optics 0752
Radiation 0756
Solid State 0611
Statistics 0463
Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
General 0537
Aerospace 0538
Agricultural 0539
Automotive 0540
Biomedical 0541
Chemical 0542
Civil 0543
Electronics and Electrical 0544
Heat and Thermodynamics 0348
Hydraulic 0545
Industrial 0546
Marine 0547
Materials Science 0794
Mechanical 0548
Metallurgy 0743
Mining 0551
Nuclear 0552
Packaging 0549
Petroleum 0765
Sanitary and Municipal 0554
System Science 0790
Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

The undersigned hereby recommend to
The Faculty of Graduate Studies and Research
acceptance of the thesis,

**“Contributions to Poststratification in
Stratified Multi-Stage Samples ”**

submitted by
Wesley Thomas Yung, B.Sc., M.Sc.

in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Dr. Vlasta Dlab, Chair
Department of Mathematics and Statistics

Dr. J.N.K. Rao, Thesis Supervisor

Dr. Jun Shao, External Examiner

Carleton University
February, 1996

Abstract

Poststratification is a common technique used in sample surveys to improve the efficiency of estimators. Although widely used, the majority of survey sampling research focuses on the basic expansion estimator and not the poststratified estimator. We systematically study the poststratified estimator beginning with the complete data case then proceeding to weighting adjustments for unit nonresponse and imputation for item nonresponse. To estimate the variance of the poststratified estimator in each situation, we consider the jackknife variance estimator and its linearized version.

We establish asymptotic normality of the poststratified estimator and asymptotic consistency of the corresponding jackknife variance estimator, in the complete data case, under relatively simple regularity conditions. These regularity conditions, similar to those of Scott and Wu (1981) for ratio and regression estimators, are weaker than those required by Krewski and Rao (1981) for smooth functions of means. To reduce computational costs, we linearize the jackknife variance estimator to obtain the linearized jackknife variance estimator which retains the desirable properties of the jackknife variance estimator. We show that the jackknife and linearized jackknife variance estimators are equivalent to higher order terms in the important special case of two sample clusters per stratum, thus generalizing the result of Rao and Wu (1985) to poststratification. Using a model-based approach, we extend Valliant's (1993) results for the one-way poststratified estimator, to the generalized regression estimator in the context of two-way poststratification.

Turning to the missing data case, we study the poststratified estimator un

der weighting adjustments for unit nonresponse with single and multiple weighting classes. In the single weighting class case, the poststratified estimator reduces to the situation where only the respondents are used for estimation. Asymptotic properties follow readily from the complete data case.

In the item nonresponse case, we consider the poststratified estimator under mean imputation and hot deck imputation. It is well known that treating the imputed values as true values can lead to severe underestimation of the variance if the nonresponse rate is appreciable. We use the Rao-Shao (1992) jackknife method to construct "correct" jackknife variance estimators and establish the asymptotic consistency of the proposed jackknife variance estimators.

The "correct" jackknife variance estimators are linearized in each case to produce linearized jackknife variance estimators. The linearized jackknife variance estimators are shown to be asymptotically equivalent to the jackknife variance estimators and are thus asymptotically consistent. Empirical results supporting this equivalence are presented as well as results on the finite sample performance of the jackknife variance estimators.

Acknowledgements

I would like to thank my supervisor, Dr. J.N.K. Rao whose guidance made this thesis possible. I would also like to thank my parents who always believed in me. Finally, I would like to thank my wife, Kimberley, who stood by me throughout this endeavour.

Contents

Acceptance Sheet	i
Abstract	ii
Acknowledgements	iv
Contents	v
List of Tables	viii
1 Introduction	1
1.1 Sample Design and Asymptotic Setup	3
1.1.1 Sample Design	3
1.1.2 Asymptotic Setup	4
1.2 Jackknife Variance Estimation	5
1.3 Full Response	7
1.3.1 Results of Chapters 2 and 3	8
1.4 Nonresponse	10
1.4.1 Unit Nonresponse	10
1.4.2 Results of Chapter 4	12
1.4.3 Imputation	13
1.4.4 Mean Imputation	15
1.4.5 Hot Deck Imputation	16

1.4.6	Results from Chapters 5 and 6	18
2	Poststratification and Jackknife Variance Estimation	20
2.1	Introduction	20
2.2	The Poststratified Estimator	22
2.3	Large Sample Properties	24
2.3.1	Basic Estimator	24
2.3.2	Poststratified Estimator	33
2.4	Jackknife Variance Estimation	41
3	Two-way Poststratification	54
3.1	Introduction	54
3.2	One-Way Poststratification	56
3.2.1	Model Bias and Prediction Variance	56
3.2.2	Variance Estimation	58
3.2.3	Second Order Analysis	66
3.3	Two-Way Poststratification	70
3.3.1	Model Bias and Prediction Variance	72
3.3.2	Variance Estimation	77
3.3.3	Large Sample Properties	86
3.3.4	Second Order Analysis	95
3.4	Simulation Study	99
4	Nonresponse Adjustments for Unit Nonresponse	107
4.1	Introduction	107
4.2	Poststratified Estimator	109
4.2.1	One Weighting Class	109

4.2.2	Multiple Weighting Classes	121
4.3	The Regression Estimator	128
4.3.1	One Weighting Class	128
4.3.2	Multiple Imputation Classes	138
4.4	Simulation Study	143
5	Deterministic Imputation	147
5.1	Introduction	147
5.2	Mean Imputation	149
5.2.1	One Imputation Class	149
5.2.2	Multiple Imputation Classes	168
5.3	Simulation Study	171
5.4	Appendix	175
6	Hot Deck Imputation	179
6.1	Introduction	179
6.2	Imputed Poststratified Estimator	181
6.3	Jackknife Variance Estimation	189
6.4	Simulation Study	195
6.5	Appendix	199
7	Conclusions and Future Research	202
7.1	Conclusions	202
7.2	Future Research	204
	References	206

List of Tables

3.1	FAMEX Average Expenditures for Married Couple Households	85
3.2	Unconditional Results for The Poststratified Estimator	101
3.3	Unconditional Results for The GREG Estimator	103
3.4	Conditional Relative Biases for The Poststratified Estimator	101
3.5	Conditional Error Rates for The Poststratified Estimator	105
3.6	Conditional Relative Biases for The GREG Estimator	105
3.7	Conditional Error Rates for The GREG Estimator	106
4.1	Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Weighting Adjustment for Unit Nonresponse (50 Strata Design)	144
4.2	Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Weighting Adjustments for Unit Nonresponse (100 Strata Design)	146
5.1	Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Mean Imputation (50 Strata Design)	172

5.2	Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Mean Imputation (100 Strata Design)	173
6.1	Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Hot Deck Imputation (50 Strata Design)	197
6.2	Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Hot Deck Imputation (100 Strata Design)	198

Chapter 1

Introduction

In complex large scale surveys, particularly household surveys, poststratification is commonly employed for two reasons: it ensures consistency with known population totals and it improves the efficiency of sample estimates. For simple random sampling, poststratification is almost as precise as proportional stratified random sampling provided the sample size in each poststratum is reasonably large (Cochran, 1977, p. 134). Therefore in multi-purpose surveys where stratification factors may be poorly correlated with a particular variable of interest, if auxiliary data is available poststratification can be used to increase the efficiency of the estimates.

The method of poststratification is described as follows: we assume the population can be divided into C nonoverlapping and exhaustive subsets or poststrata, with known population counts, based on factors such as age, sex, occupation or education level. After sample selection, the individual units are assigned to a poststratum according to these factors. The survey weights are then adjusted to force the estimated numbers of units in each poststratum to agree with the known population counts. These adjusted weights are then used in forming estimates of means or totals of variables collected in the survey.

A large portion of standard sampling theory is devoted to the estimation of the variance of estimators of a population total or mean. For nonlinear statistics, such as the poststratified estimator, the well known linearization (or Taylor expansion) method is often used (see Rao (1988) for detailed applications). With the increase in computing power resampling methods, such as the jackknife, balanced repeated replication (BRR) and the bootstrap, are becoming more popular. The jackknife, in particular, has become popular in survey sampling. In fact, Statistics Canada has adopted the jackknife method of variance estimation for stratified multi-stage surveys. An advantage of the linearization method is that it is applicable for all sampling designs, but involves a derivation of a separate standard error formula for each nonlinear statistic. On the other hand, resampling methods employ a single standard error formula for all statistics. However, the jackknife and BRR methods are strictly applicable to stratified multi-stage designs where the clusters are selected with replacement, or treated as such, within strata.

In this thesis we will study the poststratified estimator and the corresponding jackknife variance estimators under two situations; the full response case and the nonresponse case. In the nonresponse case, we distinguish between total, or unit, nonresponse and item nonresponse. In this Introduction we will review the relevant literature and summarize the main results of this thesis. We first describe our sample design and our asymptotic setup in Section 1.1. In Section 1.2, we review the jackknife method of variance estimation as it pertains to survey sampling. Section 1.3 reviews the poststratified estimator in the full response case. A model based approach, used in Chapter 3, is also discussed in Section 1.3. Results from Chapters 2 and 3 are summarized in Subsection 1.3.1. Section 1.4 and subsequent sections consider the poststratified estimator in the nonresponse case. Weighting adjustments

for unit nonresponse are reviewed in Subsection 1.4.1 with results from Chapter 4 summarized in Subsection 1.4.2. An overview of imputation methods and the Rao-Shao jackknife variance estimator under imputation are presented in Subsection 1.4.3. Mean imputation and hot deck imputation are discussed in Subsections 1.4.4 and 1.4.5 respectively. Results from Chapters 5 and 6 are summarized in Subsection 1.4.6.

1.1 Sample Design and Asymptotic Setup

1.1.1 Sample Design

We now describe the stratified multi-stage design which will be considered throughout this thesis. The population under consideration is stratified into L design strata with N_h primary sampling units (PSU's) or clusters in the h^{th} stratum. Within each stratum, $n_h \geq 2$ clusters are selected from stratum h , independently across strata. The first-stage clusters are usually selected without replacement to avoid the selection of the same cluster more than once. However, at the stage of variance estimation it is common practice to treat the sample as if the first-stage clusters are drawn with replacement, to simplify the calculations. The approximation leads to overestimation of the true variance, but the bias is small if the first-stage sampling fractions, n_h/N_h , are small.

Within the $(hi)^{th}$ sampled first-stage cluster, m_{hi} ultimate units (elements) are sampled according to some probability sampling method, $i = 1, \dots, n_h; h = 1, \dots, L$. Note that we do not specify the number of stages or the sampling methods beyond the first-stage. We simply assume that subsampling within sampled clusters is performed to ensure unbiased estimation of cluster totals, Y_{hi} , $i = 1, \dots, n_h; h =$

$1, \dots, L$.

Associated with the k^{th} sampled element in the i^{th} sampled cluster belonging to the h^{th} stratum is the variable of interest, y_{hik} , and the basic survey weight, w_{hik} . An estimator of the population total, Y , is of the form

$$\hat{Y} = \sum_{(hik) \in s} w_{hik} y_{hik}$$

where s denotes the sample of elements. Writing $\hat{Y}$ as

$$\hat{Y} = \sum_h \bar{r}_h,$$

with $r_{hi} = \sum_k (n_h w_{hik} y_{hik})$ and $\bar{r}_h = \sum_i r_{hi} / n_h$, we note that, under with replacement sampling of the clusters, the r_{hi} are independent and identically distributed random variables within each stratum. It therefore follows that an unbiased estimator of the variance of $\hat{Y}$ is given by

$$v_L(\hat{Y}) = \sum_h \frac{1}{n_h(n_h - 1)} \sum_i (r_{hi} - \bar{r}_h)^2.$$

Under without replacement sampling of the clusters, $v_L(\hat{Y})$ will overestimate the true variance of $\hat{Y}$.

1.1.2 Asymptotic Setup

Our asymptotic setup follows that of Krewski and Rao (1981). In their pioneering paper, Krewski and Rao (1981) suggested that rather than considering asymptotics at the stratum level, one should consider the behavior of estimators (aggregated over strata) as the number of strata, L , grows large. This setup includes the important special case where $n_h = 2$ clusters are selected within each stratum allowing maximum degree of stratification while still permitting valid variance estimation.

Under with replacement sampling, the clusters can be viewed as independent, non-identically distributed random variables and one can use results from main stream statistics on the sums of such random variables. We assume that there exist a sequence of finite populations $\{\mathfrak{U}_\nu\}$, with L_ν strata in $\mathfrak{U}_\nu$, and that our population of interest is one of the populations in the sequence. For simplicity, the subscript ν is suppressed and all limiting processes are understood to be as $L \rightarrow \infty$. Note that $n = \sum_h n_h \rightarrow \infty$ as $L \rightarrow \infty$.

1.2 Jackknife Variance Estimation

The jackknife was originally proposed as a method to reduce bias by Quenouille (1949) and for variance estimation and constructing approximate confidence intervals by Tukey (1958). We confine our review of the literature to stratified multi-stage designs. When the first-stage sampling is with replacement, the first-stage sample clusters are selected independently. Hence, when performing the jackknife we can delete a first stage cluster at a time. To illustrate the method, suppose that $\hat{\theta} = g(\hat{Y})$ denotes the estimator of some parameter $\theta = g(Y)$ where $g(\cdot)$ is a smooth function. Let $\hat{\theta}_{(gj)} = g(\hat{Y}_{(gj)})$ be the estimator of θ obtained from the sample after omitting the data from the j^{th} sampled cluster in the g^{th} stratum, ($j = 1, \dots, n_g; g = 1, \dots, L$), where

$$\hat{Y}_{(gj)} = \sum_{\substack{(hik) \in s, \\ h \neq g}} w_{hik} y_{hik} + \frac{n_g}{n_g - 1} \sum_{\substack{(gik) \in s, \\ i \neq j}} w_{gik} y_{gik}.$$

Note that $\hat{Y}_{(gj)}$ is obtained by changing the weights of the elements in the $(g)^{th}$ cluster, $i \neq j$, to $n_g w_{gik} / (n_g - 1)$, but retaining the original weights, w_{hik} $h \neq g$, for elements in a different stratum. A customary delete-1 cluster jackknife variance

estimator of $\hat{\theta}$ is given by

$$v_J(\hat{\theta}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (\hat{\theta}_{(gj)} - \hat{\theta})^2. \quad (1.2.1)$$

Asymptotic consistency of $v_J(\hat{\theta})$ was established by Krewski and Rao (1981) under the asymptotic setup described in Subsection 1.1.2. Two variants of $v_J(\hat{\theta})$ are obtained by replacing $\hat{\theta}$ in equation (1.2.1) with $\hat{\theta}_{(g\cdot)} = \sum_j \hat{\theta}_{(gj)}/n_g$ and $\hat{\theta}_{(\cdot\cdot)} = \sum_g \sum_j \hat{\theta}_{(gj)}/n$, where $n = \sum_g n_g$. However, these changes have little effect on the performance of the jackknife variance estimator. Rao and Wu (1985) made a second-order analysis of the BRR and the jackknife variance estimators when $\hat{\theta}$ is a smooth function of totals. Their main results on the jackknife are: (1) The different jackknife estimators are asymptotically equal to higher order terms, as the number of strata, L , increases. That is, they are equal up to the order $O_p(n^{-2})$. This result is supported by the empirical results of Kovar, Rao and Wu (1988). (2) In the important special case of $n_h = 2$ for all h , the linearization variance estimator, $v_L(\hat{\theta})$, and any jackknife variance estimator are asymptotically equal to higher order terms where

$$v_L(\hat{\theta}) = \sum_h \frac{1}{n_h(n_h - 1)} \sum_i (\hat{\theta}_{hi} - \hat{\theta}_h)^2$$

with $\hat{\theta}_{hi} = g(\hat{Y}_{hi})$, $\hat{\theta}_h = (1/n_h) \sum_i \hat{\theta}_{hi}$, and $\hat{Y}_{hi} = \sum_k (n_h w_{hik} y_{hik})$ are the estimated cluster totals. Thus in this special case, the choice of $v_J(\hat{\theta})$ or $v_L(\hat{\theta})$ should depend more on operational considerations than on statistical considerations.

More recently, Valliant (1993) considered the BRR and the jackknife variance estimators for the poststratified estimator. Using a model-based approach, he has shown that the BRR and the jackknife variance estimators are consistent for the model variance of the poststratified estimator. Valliant's work is reported in more detail in Chapter 3.

1.3 Full Response

In the full response case the poststratified estimator can be written as

$$\begin{aligned}\hat{Y}_{PS} &= \sum_c {}_cM \frac{{}_c\hat{Y}}{{}_c\hat{M}} \\ &= \sum_c \sum_{(hik) \in {}_cS} {}_cw_{hik} y_{hik}\end{aligned}$$

where ${}_cs$ denotes the sample in the c^{th} poststratum, ${}_cM$ are the known poststrata counts, ${}_c\hat{Y} = \sum_{(hik) \in {}_cS} w_{hik} y_{hik}$ and ${}_c\hat{M} = \sum_{(hik) \in {}_cS} w_{hik}$ are estimates of the poststrata total and counts respectively and ${}_cw_{hik} = {}_cM w_{hik} / {}_c\hat{M}$ are the poststratified adjusted weights ($c = 1, \dots, C$). To construct a jackknife variance estimator for $\hat{Y}_{PS}$, let $\hat{Y}_{PS(gj)}$ be the poststratified estimator based on the sample with the $(gj)^{th}$ cluster deleted (see Section 1.2). The jackknife variance estimator is

$$v_J(\hat{Y}_{PS}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (\hat{Y}_{PS(gj)} - \hat{Y}_{PS})^2.$$

Traditionally the variance of $\hat{Y}_{PS}$ is evaluated in an unconditional framework. That is, the sampling distribution is determined over all possible samples of size $m = \sum_h \sum_i m_{hi}$. Since the number of individuals falling into a poststratum is random, Holt and Smith (1979) argued “...that inferences should be made conditional on the achieved sample configuration.” This sentiment is echoed by Cox and Hinkley (1974), Smith (1976), Durbin (1969), Rao (1985) and Valliant (1993). Although conditioning is desirable, in principle, a design-based conditional theory may be intractable as noted by Rao (1985).

A popular alternative is the prediction or super population approach. The prediction approach, as described by Royall (1986), is:

My approach treats problems of finite population estimation as prediction problems, in which observations on sample units are used to make

predictions about nonsample units. The probability framework for these predictions is not the one usually employed in finite population inference - namely, the distribution created when a probability sampling plan decides which units will be observed. Instead I condition on the sample of units actually observed (so that s , s_i and u_i are held fixed) and use as the prediction framework a probability model that represents the y_{ij} and z_{ij} , both observed and unobserved, as realized values of random variables Y_{ij} and Z_{ij} .

Valliant (1993) used the prediction approach to determine whether several commonly used variance estimators estimate the conditional variance of the poststratified estimator of a finite population total.

Valliant considered four variance estimators, the customary linearization variance estimator, $v_L(\hat{Y}_{PS})$, an alternative linearization variance estimator, $v_{L2}(\hat{Y}_{PS})$, suggested by Rao (1985), the BRR variance estimator, $v_{BRR}(\hat{Y}_{PS})$, and the jackknife variance estimator, $v_J(\hat{Y}_{PS})$. After deriving the model variance, $V_m(\hat{Y}_{PS})$, of $\hat{Y}_{PS}$ Valliant has shown that $v_L(\hat{Y}_{PS})$ is not consistent for $V_m(\hat{Y}_{PS})$, while the remaining three do estimate $V_m(\hat{Y}_{PS})$ consistently. He has also shown that the poststratified estimator is model unbiased for the total Y .

1.3.1 Results of Chapters 2 and 3

Although the poststratified estimator and its corresponding jackknife variance estimator can be handled using existing asymptotic results for smooth functions of totals (Krewski and Rao, 1981), in Chapter 2 we establish asymptotic normality of $\hat{y}_{PS}$ and asymptotic consistency of $v_J(\hat{y}_{PS})$ using regularity conditions that are weaker than those assumed by Krewski and Rao (1981). These conditions are simi-

lar to those of Scott and Wu (1981) for ratio and regression estimators under simple random sampling. Also, the proofs in Chapter 2 are presented in detail and will be frequently referenced in subsequent chapters.

In Chapter 3, we extend Valliant's super population model to two-way poststratification. Two-way poststratification occurs when two poststratification variables are used but population counts are known only for the marginal poststrata and not for the cross categories. For instance, population counts may be known for age groups and race groups but not known for individual age-race groups. The two-way poststratification case can be handled by the generalized regression (GREG) estimator (see Huang and Fuller, 1978; Deville and Särndal, 1992). Using the model-based approach, we establish the model unbiasedness of the GREG estimator and the model consistency of the corresponding jackknife variance estimator and an alternative linearization variance estimator which is analogous to $v_{L2}(\hat{y}_{PS})$. We also show that the customary linearization variance estimator is not consistent for the model variance of the GREG estimator.

Also in Chapter 3, under the design-based framework, we establish the second order equivalence of the jackknife variance estimator and an alternative linearization variance estimator, obtained by linearizing the jackknife variance estimator, for both the poststratified estimator and the GREG estimator. This equivalence holds to a higher order in the special case of $n_h = 2$ for all h . This result extends the result of Rao and Wu (1985), who considered only the basic estimator, to the poststratified estimator and the GREG estimator. This theoretical result has motivated an empirical study at Statistics Canada, using actual data, in the hope of replacing the computationally expensive jackknife variance estimator by the faster and cheaper linearized jackknife variance estimator (see Chapter 3).

1.4 Nonresponse

In practice, even a carefully planned survey will suffer from nonresponse of some type. There are two major types of nonresponse as described by Kalton and Kasprzyk (1986):

It is common practice to distinguish between total (or unit) nonresponse, when none of the survey responses are available for a sampled element, and item nonresponse, when some but not all of the responses are available. Total nonresponse arises because of refusals, inability to participate, not-at-homes and untraced elements. Item nonresponse arises because of item refusals, "don't knows", omissions and answers deleted in editing.

The distinction between unit and item nonresponse is useful since different methods are used to compensate for the missing data. In general, weighting adjustments are used for unit nonresponse while imputation methods are used for item nonresponse.

Throughout this thesis we assume a uniform response mechanism, that is independent nonresponse across sample units and equal response probabilities. The latter assumption can be relaxed by dividing the sample into two or more weighting or imputation classes according to the values of an auxiliary variable observed for all units in the sample, and then assuming a uniform response mechanism within the weighting or imputation classes.

1.4.1 Unit Nonresponse

Unit nonresponse is usually compensated by weighting adjustments. In essence, weighting adjustment procedures increase the weights of the respondents so that they

represent the nonrespondents. In practice, the sample is usually divided into classes, called weighting classes, based on characteristics known for all sample elements, regardless of whether they respond or not. Separate weighting adjustments are then applied within each weighting class. We note that the population counts in each weighting class are unknown, unlike the poststrata counts.

We illustrate the weighting adjustment method in the simple situation where poststrata do not exist and we have a single weighting class. The sample, s , is divided into respondents, s_r , and nonrespondents, s_m . The nonresponse adjustment is

$$d = \frac{\sum_{(hik) \in s} w_{hik}}{\sum_{(hik) \in s_r} w_{hik}}.$$

That is, the nonresponse adjustment is simply the inverse of the weighted response rate. Using this nonresponse adjustment, the basic estimator of the total is

$$\begin{aligned} \hat{Y} &= \sum_{(hik) \in s_r} dw_{hik} y_{hik} \\ &= \sum_{(hik) \in s_r} w_{hik}^N y_{hik} \end{aligned}$$

where $w_{hik}^N = dw_{hik}$ is the nonresponse adjusted weight. In the case of multiple weighting classes, the adjustment is performed independently within each class.

When poststrata exist, it is common practice to use the poststrata as weighting classes and to use known poststrata counts to adjust the respondent weights (see Jones and Chromy, 1982, Kalton and Kasprzyk, 1986). A more interesting situation occurs when the weighting classes cut across the poststrata. In this case the weighting adjustment is usually performed first and then the poststratification adjustment is performed using the adjusted weights (see Flyer, Rust and Morganstein, 1989). This situation is addressed in Chapter 4. Also in Chapter 4, we consider the GREG

estimator in the presence of unit nonresponse with weighting classes cutting across the poststrata.

1.4.2 Results of Chapter 4

In Chapter 4 we investigate the properties of the poststratified estimator and the GREG estimator in the presence of unit nonresponse with a single weighting class and multiple weighting classes. In Subsection 4.2.1 we show that the poststratified estimator under a single weighting class reduces to the poststratified estimator based on the respondents only. We establish the asymptotic normality of the poststratified estimator and the asymptotic consistency of the corresponding jackknife variance estimator thus allowing the use of normal theory confidence intervals. A linearization variance estimator is proposed which is asymptotically equal to the jackknife variance estimator. This equivalence is confirmed by empirical results presented in Section 4.4.

Subsection 4.2.2 studies the poststratified estimator under multiple weighting classes. The poststratified estimator is defined and a jackknife variance estimator is constructed. For simplicity, asymptotic properties are not presented but we note that normality and consistency follows along the lines of the single weighting class case. The jackknife variance estimator is linearized to obtain a linearization variance estimator. This linearized jackknife variance estimator is asymptotically equivalent to the jackknife variance estimator but is computationally easier and can be used with existing linearization variance estimation software with only slight modifications.

In Section 4.3, the results for the poststratified estimator are extended to the GREG estimator in the context of two-way poststratification. Linearized jackknife

variance estimators are also obtained.

1.4.3 Imputation

Most surveys use some form of imputation for item nonresponse. Some common imputation methods are: (see Kalton and Kasprzyk, 1986)

- (a) Overall mean imputation
- (b) Class mean imputation
- (c) Ratio imputation
- (d) Regression imputation
- (e) Nearest neighbor imputation
- (f) Random overall imputation
- (g) Random imputation within classes
- (h) Sequential hot deck imputation
- (i) Hierarchical hot deck imputation.

The imputation procedures listed above are usually divided into two major groups :

(1) Deterministic imputation which includes procedures (a)-(e). The imputed values are deterministic, given the sample of respondents and any auxiliary information on the nonrespondents. (2) Hot deck imputation which includes procedures (f)-(i). In hot deck imputation, missing responses are replaced by values selected randomly from the respondents.

Imputation is usually used for item nonresponse to complete the data set in order that full response procedures can be used. This leads to valid point estimates but if the imputed values are treated as true values, variance estimates computed with standard formulas can be severe underestimates of the true variance. Rao and Shao (1992) proposed new jackknife variance estimators under hot deck imputation which are design consistent under a uniform response mechanism. Rao (1996) used this new jackknife method to handle deterministic and stochastic ratio and regression imputation.

We now give the motivation behind the new jackknife variance method, proposed by Rao and Shao (1992). Let y_{hik}^* be the imputed values for the nonrespondents under a specified imputation procedure. The imputed estimator of Y is then given by

$$\hat{Y}_I = \sum_{(hik) \in s_r} w_{hik} y_{hik} + \sum_{(hik) \in s_m} w_{hik} y_{hik}^*$$

where s_r denotes the total sample of respondents and s_m denotes the total sample of nonrespondents to item y . Before calculating the jackknife variance estimator we need to adjust the imputed values each time a cluster is deleted. Under mean imputation (Subsection 1.4.4), we adjust y_{hik}^* by the amount $y_{hik(gj)}^* - y_{hik}^*$, where $y_{hik(gj)}^*$ is the value one would impute for the $(hik)^{th}$ nonrespondent if the $(gj)^{th}$ cluster is deleted. That is we replace the imputed values with

$$\begin{aligned} z_{hik(gj)}^* &= y_{hik}^* + y_{hik(gj)}^* - y_{hik}^* \\ &= y_{hik(gj)}^*. \end{aligned}$$

We see that the adjustment simply adjusts the imputed values to reflect the deletion of the $(gj)^{th}$ cluster. We now apply the standard jackknife variance estimator except we use the adjusted values, $z_{hik(gj)}^*$, instead of the imputed values, y_{hik}^* .

In the case of hot deck imputation, we adjust the imputed values, y_{hik}^* , by an average amount $E_{\bullet}^{(gj)} y_{hik}^* - E_{\bullet} y_{hik}^*$, where $E_{\bullet}$ denotes expectation with respect to the imputation given the donor set and $E_{\bullet}^{(gj)}$ denotes the expectation with respect to the imputation when the donor set is modified by excluding the respondents in the $(gj)^{th}$ cluster.

We note that to implement this jackknife procedure, the imputed complete dataset must carry identification flags to locate the imputed and non-imputed values.

1.4.4 Mean Imputation

To illustrate the mean imputation procedure, suppose we have a stratified multi stage design with a single imputation class. In the traditional mean imputation procedure, for the nonrespondents we impute the mean of the respondents,

$$\hat{y}_r = \frac{\sum_{(hik) \in s_r} w_{hik} y_{hik}}{\sum_{(hik) \in s_r} w_{hik}},$$

where s_r denotes the respondents in the sample. After imputation poststratification can be performed. We propose a slightly different mean imputation procedure. Before imputing, calculate the poststratified adjusted weights and use these adjusted weights to calculate the mean of the respondents. That is, for the nonrespondents we impute

$$\hat{y}_r = \frac{\sum_c \sum_{(hik) \in c s_r} {}_c w_{hik} y_{hik}}{\sum_c \sum_{(hik) \in c s_r} {}_c w_{hik}},$$

where ${}_c w_{hik}$ are the poststratified adjusted weights and ${}_c s_r$ denotes the respondents in the c^{th} poststratum. This poststratified mean is a more efficient estimator of the respondent mean and should therefore lead to more efficient estimation.

1.4.5 Hot Deck Imputation

Hot deck imputation is commonly employed for item nonresponse because of the following advantages: (1) It preserves the distribution of the item values. (2) Results obtained from different analyses are consistent with one another. (3) It permits the use of the same survey weight for all items. The simplest form of hot deck imputation uses a simple random sample of size n_m selected with replacement from the sample respondents, s_r , to an item y and the associated item values are used as "donors", where n_m is the number of nonrespondents to an item y .

Regardless of the imputation procedure, using standard variance estimation procedures will lead to serious underestimation of the true variance. Rubin (1978) proposed multiple imputation to account for the inflation in the variance due to imputation. In this method, assuming simple random sampling, $t \geq 2$ independent imputations are performed to form t complete datasets. From each dataset a point estimate, $\bar{y}_{il}$, and a variance estimate, s_{il}^2 , are calculated. Rubin proposed the average of the t point estimates as the final estimate,

$$\bar{y}_I = \frac{1}{t} \sum_i \bar{y}_{il}$$

with estimated variance,

$$v_m = \frac{1}{t} \sum_i s_{il}^2 + \frac{t+1}{t} \left(\frac{1}{t-1} \sum_i (\bar{y}_{il} - \bar{y}_I)^2 \right). \quad (1.4.2)$$

Rubin and Schenker (1986) proposed the following "proper" imputation scheme, called the approximate Bayesian bootstrap (ABB), as an alternative to the usual simple random hot deck: For each imputation l ,

- 1) draw a simple random sample with replacement of size n_r from the n_r elements of s_r

- 2) draw a simple random sample with replacement of size n_m from the sample drawn above to use as donors.

They showed that performing the imputation as described above generates appropriate between imputation variability.

However, multiple imputation does have its limitations as noted by Rao (1996):

First, the costs of storage and processing multiple datasets are significantly increased relative to those under single imputation, especially in large scale surveys. Secondly, general ABB methods for generating "proper" imputations that accommodate issues of clustering, stratification and weighting to compensate for unequal probabilities of selection are not currently available (Fay, 1993); for example, in the common situations of imputation cutting across sample clusters. Thirdly, imputations are sometimes deterministic, i.e., determined uniquely given a sample of donors and auxiliary data, due to the desire to fill in the best possible value (Särndal, 1992); for example, ratio and nearest neighbour imputations. Finally, a small number of imputations, M , may result in a low level of precision for the multiple imputation variance estimator (1.4.2) since the last term in (1.4.2) is based on $M - 1$ degrees of freedom (Kalton and Kasprzyk, 1986).

To address these concerns, Rao and Shao (1992) proposed a modification to the standard design-based stratified jackknife variance formula to provide suitable variance estimates for data with single imputation (see Subsection 1.4.3). They showed that, under a particular weighted hot deck and the uniform response mechanism, the proposed jackknife variance estimator is asymptotically consistent. The weighted

hot deck they considered involves selecting donors $(gjl)\epsilon s_r$ with replacement with probabilities $w_{gjl}/\sum_s w_{hik}$ and using $y_{hik}^* = y_{gjl}$.

In Chapter 6, we extend the results of Rao and Shao (1992) to the poststratified estimator under a different hot deck imputation. We perform the poststratification before the imputation and use the poststratified weights in the imputation procedure. This is, we select donors $(gjl)\epsilon s_r$ with replacement with probabilities ${}_c w_{gjl}/\sum_c \sum_{s_r} {}_c w_{hik}$ where ${}_c w_{hik}$ are the poststratified adjusted weights.

1.4.6 Results from Chapters 5 and 6

In Chapter 5 we study the poststratified estimator under mean imputation using the poststratified weights. In Subsection 5.2.1, we consider the poststratified estimator under a single imputation class and establish asymptotic normality of the estimator. A jackknife variance estimator is constructed and asymptotic consistency is established. These results establish the validity of normal theory confidence intervals.

Subsection 5.2.2 examines the poststratified estimator under mean imputation with multiple imputation classes. Asymptotic properties of the poststratified estimator and the corresponding jackknife variance estimators are not presented but we note that normality and consistency follows along the lines of the single imputation class case. Linearized jackknife variance estimators, which are asymptotically equal to the jackknife variance estimators, are derived for both the single imputation class and multiple imputation class cases. In the single imputation class case, this theoretical equivalence is supported by empirical results presented in Section 5.3.

In Chapter 6, we extend the results of Rao and Shao (1992) to the poststratified estimator under a weighted hot deck using the poststratified weights. We present results from the single imputation class case only, but the extension to multiple im-

putation classes is fairly straightforward. We note that hot deck imputation is the stochastic counterpart of mean imputation as the poststratified estimator reduces to the poststratified estimator under mean imputation plus a stochastic component from the imputation procedure. Asymptotic normality of the poststratified estimator is established in Section 6.2.

For variance estimation, we construct a jackknife variance estimator in Section 6.3 and establish the asymptotic consistency of the jackknife variance estimator. The consistency of the jackknife and the asymptotic normality of the poststratified estimator imply that normal theory confidence intervals are valid. A linearized jackknife variance estimator is presented which is asymptotically equivalent to the jackknife and is thus asymptotically consistent for the variance of the poststratified estimator. Results of a limited simulation study showing this equivalence are presented in Section 6.4.

Chapter 2

Poststratification and Jackknife Variance Estimation

2.1 Introduction

In complex large-scale surveys, poststratification is commonly used to increase the efficiency of estimators. Survey weights are adjusted so that the estimated number of units in each poststratum is equal to or will benchmark to known population counts such as projected census counts. These adjusted weights are then used in estimating totals or means of variables collected in the survey. Although producing more efficient estimators is important, an administrative advantage of the poststratified estimator is the benchmarking property. This is attractive to agencies, such as Statistics Canada, which produce a large number of statistics, since it ensures agreement between their surveys. In their ground breaking paper, Holt and Smith (1979) showed that the poststratified estimator has "built in" protection against what they termed unfavourable sample configurations in the sense that if a sample is badly balanced for some characteristic, such as age, the poststratified estimator automatically corrects for this. A description of the poststratification method, paraphrased

from Holt and Smith (1979) is as follows:

Responses from individuals may vary by age, sex, education and other similar factors, none of which are available for stratification purposes at the design stage. However, a population census may provide information on these variables at aggregate levels which can be used at the estimation stage. After sample selection, individuals can be placed into poststrata according to the factors mentioned above. The known population totals can be used to adjust the design weights to produce an estimate of the population total of the poststrata. The poststrata totals are then summed to yield an estimate of the full population total. This technique is particularly useful in multi-purpose surveys where the stratification variables may be poorly correlated with secondary variables.

Since poststrata identifiers for individuals are unknown at the design stage, the sample size in each poststratum is a random variable. Thus, inferences can be made either unconditionally (i.e. across all possible realizations of the poststrata sample sizes) or conditionally on the estimated counts in the poststrata. Durbin (1969), argues that the achieved sample size should be treated as an ancillary statistic and inferences should be made conditional on the estimated poststrata counts. This sentiment is echoed by Cox and Hinkley (1974), Smith (1976), Holt and Smith (1979), Rao (1985) and more recently by Valliant (1993).

This chapter will be concerned with the unconditional properties of the post-stratified estimator. Conditional properties, in particular conditionally valid variance estimators, will be discussed in Chapter 3. In Section 2.2, the poststratified estimator will be introduced. Section 2.3 will investigate the large sample properties

of the poststratified estimator including asymptotic consistency and normality. Also in Section 2.3, the asymptotic consistency of the customary linearization variance estimator will be shown. In Section 2.4 the jackknife variance estimator will be introduced and its asymptotic consistency will be established.

2.2 The Poststratified Estimator

Suppose we have a stratified multi-stage design with L design strata and N_h clusters or primary sampling units (PSU's) in stratum h . Cluster (hi) contains M_{hi} ultimate units (elements) with the total number of elements in stratum h being $M_h = \sum_{i=1}^{N_h} M_{hi}$ and the total number of elements in the population $M = \sum_h M_h$. A first stage sample of $n_h \geq 2$ clusters is sampled from each stratum independently across stratum and m_{hi} elements are sampled if the $(hi)^{th}$ cluster is selected. We will assume that subsampling within clusters is performed to ensure unbiased estimation of cluster totals Y_{hi} , $h = 1, \dots, L$; $i = 1, \dots, n_h$. For simplicity, we will index the elements in a first-stage cluster using a single index, i.e. element (hik) is the k^{th} sampled element in the i^{th} sampled cluster of the h^{th} stratum, $h = 1, \dots, L$, $i = 1, \dots, N_h$ and $k = 1, \dots, M_{hi}$. We will also assume that the number of clusters sampled in each stratum, n_h , is bounded for all strata.

From the specification of the sampling design, basic weights $w_{hik} (> 0)$ attached to the $(hik)^{th}$ sample element are obtained. We will let s denote the total sample of elements and y_{hik} be the variable of interest associated with the $(hik)^{th}$ sample element. Using these basic weights, an unbiased estimator of the total, Y , is

$$\hat{Y} = \sum_{(hik) \in s} w_{hik} y_{hik},$$

and will be denoted as the basic expansion estimator. Krewski and Rao (1981)

showed the asymptotic normality for smooth functions of the basic expansion estimator and the consistency of variance estimators obtained using the linearization, jackknife and balanced repeated replication (BRR) methods in stratified sampling. Their asymptotic setup assumed that there is a sequence of populations $\{\rho_\nu, \nu = 1, 2, \dots\}$, where population ρ_ν contains L_ν strata, and the population under consideration is a member of this sequence of populations. Note that L, N_h, w_{hik}, y_{hik} , etc described above depend on ν , but ν is omitted for simplicity. They assumed that n_h is bounded and that $L \rightarrow \infty$ as $\nu \rightarrow \infty$. All limiting processes will be understood to be as $\nu \rightarrow \infty$.

Next we introduce the notion of one-way poststratification. Suppose that each element in the population belongs to a poststratum which can cut across design strata. The total number of elements in the c^{th} poststratum is ${}_cM$ and is assumed to be known, e.g. projected census counts. We will use the prescript notation to denote the poststrata. To define the poststratified estimator of Y , let ${}_cw_{hik}$ be the poststratified weight

$${}_cw_{hik} = \frac{{}_cM}{{}_c\hat{M}} w_{hik}$$

where ${}_c\hat{M}$ is an estimator of ${}_cM$, and is given by

$${}_c\hat{M} = \sum_{(hik)cs} w_{hik} {}_c\delta_{hik}$$

and ${}_c\delta_{hik}$ is the poststratum indicator variable defined as

$${}_c\delta_{hik} = \begin{cases} 1 & \text{if the } (hik)^{th} \text{ unit is in the } c^{th} \text{ poststratum} \\ 0 & \text{otherwise.} \end{cases}$$

The poststratified estimator is then given as

$$\hat{Y}_{PS} = \sum_c \sum_{(hik)cs} {}_cw_{hik} y_{hik} {}_c\delta_{hik} \quad (2.2.1)$$

$$= \sum_c \frac{{}_cM}{{}_c\hat{M}} {}_c\hat{Y},$$

where ${}_c\hat{Y} = \sum_{(hik) \in s} w_{hik} y_{hik} {}_c\delta_{hik}$ is an estimator of the poststratum total ${}_cY$. If the poststratified weights, ${}_cw_{hik}$, are used to estimate a poststratum total ${}_dM$, then substituting

$${}_d\delta_{hik} = \begin{cases} 1 & \text{if the } (hik)^{th} \text{ unit is in the } d^{th} \text{ poststratum} \\ 0 & \text{otherwise} \end{cases}$$

for y_{hik} in (2.2.1), we have

$$\begin{aligned} {}_d\hat{M}_{PS} &= \sum_c \sum_{(hik) \in s} {}_cw_{hik} {}_d\delta_{hik} {}_c\delta_{hik} \\ &= \sum_{(hik) \in d_s} {}_d\hat{w}_{hik} \\ &= \frac{{}_dM}{{}_d\hat{M}} {}_d\hat{M} = {}_dM, \end{aligned} \tag{2.2.2}$$

thus showing agreement with the known poststratum total ${}_dM$.

2.3 Large Sample Properties

2.3.1 Basic Estimator

Before investigating the large sample properties of the poststratified estimator, we will first extend some results of Krewski and Rao (1981) to stratified multi-stage sampling. For developing asymptotic theory, it is more convenient to work with the basic expansion estimator of the mean than that of the total. Letting $\hat{w}_{hik} = w_{hik}/M$ be the normalized basic weight, the basic expansion estimator of the mean $\bar{Y}$ is

$$\bar{y} = \sum_{(hik) \in s} \hat{w}_{hik} y_{hik}.$$

The estimator $\bar{y}$ is assumed to be (design) unbiased for Y . We will assume that the clusters are sampled with replacement and that further subsampling is done independently within sampled clusters. In practice, clusters are sampled without replacement, but the bias due to assuming with replacement sampling will be negligible if the first stage sampling fractions are small. Under this assumption, the normalized cluster totals, $y_{hi} = \sum_k w_{hik} y_{hik}$, are independent identically distributed random variables for each h , and independent across strata. Thus, we can make use of the following Central Limit Theorem and Law of Large Numbers for independent nonidentically distributed random variables $\{X_t\}_{t=1}^T$, with $E(X_t) = \mu_t$ and $V(X_t) = \sigma_t^2$.

Lemma 2.1 (*Central Limit Theorem*)

If (1) $\frac{1}{T} \sum_{t=1}^T \sigma_t^2 \rightarrow \sigma^2 (> 0)$ and (2) $\frac{1}{T} \sum_{t=1}^T E|X_t - \mu_t|^{2+\delta} = O(1)$ as $T \rightarrow \infty$ for some $\delta > 0$, then

$$\sqrt{T}(\bar{X} - \bar{\mu}) \longrightarrow_d N(0, \sigma^2)$$

where $\bar{X} = (1/T) \sum_t X_t$, $\bar{\mu} = (1/T) \sum_t E(X_t)$, and $\longrightarrow_d$ denotes convergence in distribution.

Lemma 2.2 (*Law of Large Numbers*)

Let $\{X_n, n = 1, 2, \dots\}$ be a sequence of independent random variables. If

$$\frac{1}{n} \sum_{k=1}^n E|X_k|^{1+\delta} = O(1)$$

as $n \rightarrow \infty$ for some $\delta > 0$, then

$$\frac{1}{n} \left(\sum_k X_k - \sum_k E(X_k) \right) \longrightarrow_p 0$$

where $\longrightarrow_p$ denotes convergence in probability.

We will also make use of the following lemmas.

Lemma 2.3 (*Slutsky's Theorem*)

Let $\{Z_n\}$, and $\{U_n\}$ be sequences of random variables and Z be a random variable. If $Z_n \rightarrow_d Z$ and $U_n \rightarrow_p u$ (a constant) then

$$(a) \quad Z_n + U_n \rightarrow_d Z + u$$

$$(b) \quad Z_n U_n \rightarrow_d uZ$$

$$(c) \quad Z_n/U_n \rightarrow_d Z/u \text{ if } u \neq 0$$

where $\rightarrow_d$ denotes convergence in distribution and $\rightarrow_p$ denotes convergence in probability.

Lemma 2.4 (*Kreuski, 1977*)

Let $X_1, \dots, X_n$ be arbitrary random variables. Then for $r > 1$,

$$E \left| \frac{1}{n} \sum_{i=1}^n X_i \right|^r \leq \frac{1}{n} \sum_i E |X_i|^r.$$

Lemma 2.5 (*Kreuski, 1977*)

Let $X_1, \dots, X_n$ be arbitrary random variables. Then for $0 < r < s$ and $w_i \geq 0$ ($i = 1, \dots, n$) with $\sum_i w_i = 1$,

$$\sum_i w_i E |X_i|^r \leq \left(\sum_i w_i E |X_i|^s \right)^{\frac{r}{s}}.$$

The regularity conditions in Kreuski and Rao (1981) can be replaced by the following conditions.

(1) $n_T^{1+\delta} \sum_h \sum_i E |y_{hi} - E(y_{hi})|^{2+\delta} = O(1)$ for some $\delta > 0$, where $n_T = \sum_h n_h \rightarrow \infty$ as $L \rightarrow \infty$, and $y_{hi} = \sum_k \tilde{w}_{hik} y_{hik}$.

C2 $n_T V(\bar{y}) \rightarrow \sigma^2$ a positive constant as $n_T \rightarrow \infty$, where $y = \sum_h \sum_i y_{hi}$.

Condition C1 is a standard Liapounov type condition on the $2+\delta$ absolute moments of y_{hi} , while C2 assumes that the limit of the variance of y exists when multiplied by n_T .

To establish consistency of $\bar{y}$, it is enough to have

$$C1' \quad n_T^\delta \sum_h \sum_i E|y_{hi} - E(y_{hi})|^{1+\delta} = O(1) \text{ for some } \delta > 0$$

which is weaker than C1. Note that C1 implies C1' as shown in Lemma 2.6.

Lemma 2.6 *Condition C1 implies C1'.*

Proof

Let

$$X_{hi} = n_T (y_{hi} - E(y_{hi})).$$

Applying Lemma 2.5 with $r = 1 + \delta$ and $s = 2 + \delta$ gives

$$\frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{1+\delta} \leq \left(\frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{2+\delta} \right)^{\frac{1+\delta}{2+\delta}}.$$

But

$$\frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{1+\delta} = n_T^\delta \sum_h \sum_i E|y_{hi} - E(y_{hi})|^{1+\delta}$$

and

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{2+\delta} &= n_T^{1+\delta} \sum_h \sum_i E|y_{hi} - E(y_{hi})|^{2+\delta} \\ &= O(1) \end{aligned}$$

by C1. Therefore,

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{1+\delta} &\leq \left(n_T^{1+\delta} \sum_h \sum_i E|y_{hi} - E(y_{hi})|^{2+\delta} \right)^{\frac{1+\delta}{2+\delta}} \\ &= O(1). \end{aligned}$$

or

$$n_T^\delta \sum_h \sum_i E|y_{hi} - E(y_{hi})|^{1+\delta} = O(1)$$

■

Asymptotic consistency of $\bar{y}$ is established in the following theorem.

Theorem 2.1 *Under C1', $\bar{y}$ is consistent for the population mean $\bar{Y}$, i.e.,*

$$\bar{y} \longrightarrow_p \bar{Y}.$$

Proof

Let

$$X_{hi} = n_T (y_{hi} - E(y_{hi})).$$

then under C1',

$$\frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{1+\delta} = O(1).$$

Therefore by Lemma 2.2

$$\frac{1}{n_T} \sum_h \sum_i X_{hi} \longrightarrow_p \frac{1}{n_T} \sum_h \sum_i E(X_{hi})$$

or

$$\frac{1}{n_T} \sum_h \sum_i X_{hi} = \bar{y} - \bar{Y} \longrightarrow_p 0$$

since $(1/n_T) \sum_h \sum_i E(X_{hi}) = 0$.

■

Theorem 2.2 below establishes the asymptotic normality of $\bar{y}$.

Theorem 2.2 *Under C1 and C2*

$$\sqrt{n_T}(\bar{y} - \bar{Y}) \longrightarrow_d N(0, \sigma^2).$$

Proof

Let

$$X_{hi} = n_T(y_{hi} - E(y_{hi}))$$

with $E(X_{hi}) = 0$. Under C2

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i V(X_{hi}) &= n_T \sum_h \sum_i V(y_{hi}) = n_T V(y) \\ &\rightarrow \sigma^2 \end{aligned}$$

satisfying condition 1 of Lemma 2.1. Now by C1,

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{2+\delta} &= n_T^{1+\delta} \sum_h \sum_i E|y_{hi} - E(y_{hi})|^{2+\delta} \\ &= O(1) \end{aligned}$$

satisfying condition 2 of Lemma 2.1. Therefore,

$$(\sqrt{n_T}/n_T) \sum_h \sum_i X_{hi} \longrightarrow_d N(0, \sigma^2)$$

or

$$\sqrt{n_T}(y - Y) \longrightarrow_d N(0, \sigma^2)$$

noting that $\sum_h \sum_i E(y_{hi}) = Y$. ■

A commonly used unbiased estimator of the variance of y is given by

$$v_L(\bar{y}) = \sum_h \frac{n_h}{n_h - 1} \sum_i (y_{hi} - \bar{y}_h)^2 \quad (2.3.1)$$

with $\bar{y}_h = (1/n_h) \sum_i y_{hi}$. Using Lemma 2.2, we establish the consistency of $v_L(y)$ in the following theorem.

Theorem 2.3 *Under C1 and C2,*

$$n_T(v_L(y) - V(y)) \longrightarrow_p 0.$$

Proof

By adding and subtracting $E(y_{hi})$, we can rewrite (2.3.1) as

$$v_L(y) = \sum_h \frac{n_h}{n_h - 1} \sum_i (y_{hi} - E(y_{hi}))^2 - \sum_h \frac{n_h^2}{n_h - 1} (\bar{y}_h - E(y_{hi}))^2. \quad (2.3.2)$$

Considering the first term of (2.3.2), let

$$X_{hi} = n_T \left(\frac{n_h}{n_h - 1} \right)^{1/2} (y_{hi} - E(y_{hi}))$$

and note that $n_h \geq 2$ implies

$$\frac{1}{n_h - 1} \leq \frac{2}{n_h}.$$

Thus,

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}^2|^{1+\epsilon/2} &= n_T^{1+\delta} \sum_h \sum_i \left(\frac{n_h}{n_h - 1} \right)^{1+\epsilon/2} E|y_{hi} - E(y_{hi})|^{2+\delta} \\ &\leq 2^{1+\epsilon/2} n_T^{1+\epsilon} \sum_h \sum_i E|y_{hi} - E(y_{hi})|^{2+\epsilon} \\ &= O(1), \end{aligned}$$

by (1). Applying Lemma 2.2 gives

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i X_{hi}^2 &= n_T \sum_h \frac{n_h}{n_h - 1} \sum_i (y_{hi} - E(y_{hi}))^2 \\ &\rightarrow_p n_T \sum_h \frac{n_h^2}{n_h - 1} \sigma_h^2 \end{aligned} \quad (2.3.3)$$

where $\sigma_h^2 = V(y_{hi})$.

Working with the second term of (2.3.2), let

$$X_h = (Ln_T)^{1/2} \frac{n_h}{(n_h - 1)^{1/2}} (\bar{y}_h - E(y_{hi}))$$

and let $0 < \delta'/2 \leq \min(\delta/2, 1/2)$. Also, noting that $L < n_T$, we have

$$\begin{aligned} \frac{1}{L} \sum_h E|X_h^2|^{1+\delta'/2} &= \frac{1}{L} \sum_h L^{1+\delta'/2} n_T^{1+\delta'/2} \frac{n_h^{2+\delta'}}{(n_h - 1)^{1+\delta'/2}} E|\bar{y}_h - E(y_{hi})|^{2+\delta'} \\ &\leq \left(\frac{L}{n_T} \right)^{\delta'/2} 2^{1+\delta'/2} n_T^{1+\delta'} \sum_h n_h^{1+\delta'/2} E|\bar{y}_h - E(y_{hi})|^{2+\delta'} \\ &= O(1) \end{aligned}$$

if we can show that

$$n_T^{1+\delta'} \sum_h n_h^{1+\delta'/2} E|\bar{y}_h - E(y_{h_i})|^{2+\delta'} = O(1). \quad (2.3.4)$$

Using the Schwarz inequality we have

$$\begin{aligned} E|\bar{y}_h - E(y_{h_i})|^{2+\delta'} &= E|\bar{y}_h - E(y_{h_i})| |\bar{y}_h - E(y_{h_i})|^{1+\delta'} \\ &\leq (E|\bar{y}_h - E(y_{h_i})|^2 E|\bar{y}_h - E(y_{h_i})|^{2+2\delta'})^{1/2} \\ &= (\sigma_h^2/n_h)^{1/2} (E|\bar{y}_h - E(y_{h_i})|^{2+2\delta'})^{1/2} \\ &\leq (\sigma_h^2/n_h)^{1/2} \left(\frac{1}{n_h} \sum_i E|y_{h_i} - E(y_{h_i})|^{2+2\delta'} \right)^{1/2} \end{aligned} \quad (2.3.5)$$

where we have used the Lemma 2.4 to obtain the last inequality. Using expression (2.3.5) in the left hand side of (2.3.4), we have

$$\begin{aligned} n_T^{1+\delta'} \sum_h n_h^{1+\delta'/2} E|\bar{y}_h - E(y_{h_i})|^{2+\delta'} &\leq n_T^{1+\delta'} \sum_h n_h^{1+\delta'/2} (\sigma_h^2/n_h)^{1/2} \left(\frac{1}{n_h} \sum_i E|y_{h_i} - E(y_{h_i})|^{2+\delta'} \right)^{1/2} \\ &\leq n_T^{1+\delta'} \sum_h n_h \sigma_h \left(\frac{1}{n_h} \sum_i E|y_{h_i} - E(y_{h_i})|^{2+\delta'} \right)^{1/2} \\ &= \sum_h n_T^{1/2} n_h^{1/2} \sigma_h n_T^{1/2+\delta'} n_h^{1/2} \left(\frac{1}{n_h} \sum_i E|y_{h_i} - E(y_{h_i})|^{2+\delta'} \right)^{1/2} \\ &\leq \left(n_T \sum_h n_h \sigma_h^2 \right)^{1/2} \left(n_T^{1+2\delta'} \sum_h \sum_i E|y_{h_i} - E(y_{h_i})|^{2+2\delta'} \right)^{1/2} \\ &= O(1) \end{aligned}$$

where we have used C1, C2 and the fact that $\delta'/2 \leq 1/2$. Note that $V(y) = \sum_h n_h \sigma_h^2$.

Thus by Lemma 2.2,

$$\frac{1}{L} \sum_h X_h^2 \longrightarrow_p \frac{1}{L} \sum_h E(X_h^2)$$

or

$$n_T \sum_h \frac{n_h^2}{n_h - 1} (\bar{y}_h - E(y_{hi}))^2 \rightarrow_p n_T \sum_h \frac{n_h}{n_h - 1} \sigma_h^2. \quad (2.3.6)$$

Combining (2.3.3) and (2.3.6) gives

$$\begin{aligned} n_T v_L(y) &= n_T \sum_h \frac{n_h}{n_h - 1} \sum_i (y_{hi} - E(y_{hi}))^2 - n_T \sum_h \frac{n_h^2}{n_h - 1} (\bar{y}_h - E(y_{hi}))^2 \\ &\rightarrow_p n_T \sum_h \frac{n_h^2}{n_h - 1} \sigma_h^2 - n_T \sum_h \frac{n_h}{n_h - 1} \sigma_h^2 = n_T V(\bar{y}). \end{aligned}$$

That is,

$$n_T (v_L(\bar{y}) - V(\bar{y})) \rightarrow_p 0.$$

■

Recently, with the increase in computing power, resampling techniques for variance estimation have become popular. One such technique is jackknife variance estimation. To calculate a jackknife variance estimator from a stratified multi-stage sample, one simply deletes a cluster, adjusts the weights of the remaining clusters in the strata and then recalculates the estimator based on this pseudo-replicate. By repeating this for all clusters in the sample in turn one obtains n_T pseudo-replicate estimators. The jackknife variance estimator uses the variation present in the pseudo-replicate estimators to approximate the variance of the original estimator.

To construct the jackknife variance estimator, let $\bar{y}_{(gj)}$ be the basic expansion estimator when the $(gj)^{th}$ cluster has been deleted. This is defined as

$$y_{(gj)} = \sum_{\substack{(hik) \in s, \\ h \neq g}} \hat{w}_{hik} y_{hik} + \frac{n_g}{n_g - 1} \sum_{\substack{(gik) \in s, \\ i \neq j}} \hat{w}_{gik} y_{gik}. \quad (2.3.7)$$

The jackknife variance estimator is then given by

$$v_J(\bar{y}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (\bar{y}_{(gj)} - \bar{y})^2. \quad (2.3.8)$$

Consistency of (2.3.8) is straight forward since in this case, the jackknife variance estimator reduces to $v_L(\bar{y})$. To see this, consider $\bar{y}_{(gj)}$. For fixed (gj) , adding and subtracting $\sum_{(gik) \in s} \hat{w}_{gik} y_{gik}$ to equation (2.3.7) gives

$$y_{(gj)} = \sum_{(hik) \in s} \hat{w}_{hik} y_{hik} + \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} y_{gik} - \sum_{(gjk) \in s} \hat{w}_{gjk} y_{gjk} \right)$$

or

$$\bar{y}_{(gj)} - \bar{y} = \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} y_{gik} - \sum_{(gjk) \in s} \hat{w}_{gjk} y_{gjk} \right).$$

The notation $\sum_{(gik) \in s}$ indicates that we sum over the indices i and k , while keeping the index g fixed. Likewise, $\sum_{(gjk) \in s}$ indicates that we sum over the index k while keeping g and j fixed. For the remainder of this thesis, when g or j appear in a summation, they will be considered as fixed. Putting this expression into (2.3.8) gives

$$\begin{aligned} v_J(\bar{y}) &= \sum_g \frac{n_g - 1}{n_g} \sum_j \left[\frac{n_g}{n_g - 1} (\bar{y}_g - y_{gj}) \right]^2 \\ &= \sum_g \frac{n_g}{n_g - 1} \sum_j (y_{gj} - y_g)^2 \end{aligned}$$

which is precisely $v_L(\bar{y})$. Therefore, as a result of Theorem 2.3, we have the following corollary.

Corollary 2.1 *Under C1 and C2,*

$$n_T (v_J(\bar{y}) - V(\bar{y})) \rightarrow_p 0.$$

2.3.2 Poststratified Estimator

We now turn to the poststratified estimator of the population mean,

$$\begin{aligned} \bar{y}_{PS} &= \frac{\hat{Y}_{PS}}{M} \\ &= \sum_c \frac{cM}{c\hat{M}} \sum_{(hik) \in s} \hat{w}_{hik} y_{hik} c\delta_{hik}. \end{aligned}$$

Writing Y as $\bar{Y} = \sum_c ({}_cM/M) {}_c\bar{Y}$, we have

$$\begin{aligned}
 y_{rs} - \bar{Y} &= \sum_c \frac{{}_cM}{M} \sum_{(hik)_{cs}} \hat{w}_{hik} (y_{hik} - {}_c\bar{Y}) {}_c\delta_{h,k} \\
 &= \sum_{(hik)_{cs}} \hat{w}_{hik} (y_{hik} - \sum_c {}_c\bar{Y} {}_c\delta_{hik}) + \sum_c \left(\frac{{}_cM}{M} - 1 \right) \sum_{(hik)_{cs}} \hat{w}_{hik} (y_{hik} - {}_c\bar{Y}) {}_c\delta_{hik} \\
 &= \sum_{(hik)_{cs}} \hat{w}_{hik} \epsilon_{hik} + \sum_c \left(\frac{{}_cM}{M} - 1 \right) \sum_{(hik)_{cs}} \hat{w}_{hik} {}_c\epsilon_{hik} {}_c\delta_{hik} \\
 &= \epsilon + \sum_c \left(\frac{{}_cM}{M} - 1 \right) {}_c\bar{\epsilon}
 \end{aligned}$$

where $\epsilon_{hik} = y_{hik} - \sum_c {}_c\bar{Y} {}_c\delta_{hik}$ and ${}_c\epsilon_{hik} = y_{hik} - {}_c\bar{Y}$ are residuals, $\bar{\epsilon} = \sum_{(hik)_{cs}} \hat{w}_{hik} \epsilon_{hik}$ and ${}_c\bar{\epsilon} = \sum_{(hik)_{cs}} \hat{w}_{hik} {}_c\epsilon_{hik} {}_c\delta_{hik}$.

To account for the poststratification we will use the following regularity conditions.

(C3) $n_T^{1+\delta} \sum_h \sum_i E[|\epsilon_{hi} - E(\epsilon_{hi})|^{2+\delta}] = O(1)$ for some $\delta > 0$, where $n_T = \sum_h n_h \rightarrow \infty$ as $L \rightarrow \infty$, and $\epsilon_{hi} = \sum_k \hat{w}_{hik} \epsilon_{hik}$.

(C4) $n_T \max_{h,i,k} m_{hi} \hat{w}_{hik} = O(1)$

(C5) $n_T V({}_c\epsilon) = O(1)$ for all c .

Assumption C3 is the same as C1 except the y_{hik} 's have been replaced by the residuals, ϵ_{hik} . Assumption C4 means that no survey weight is disproportionately large.

It is of interest to see what C4 reduces to in some special cases. If the design is a stratified simple random sampling design, then $\hat{w}_{hik} = N_h/(Nn_h)$ with $N = \sum_h N_h$ and C4 reduces to $n_T \max_h N_h/(Nn_h) = O(1)$, an assumption used by Krewski and Rao (1981). If the design is a stratified two-stage design with simple random sampling within each cluster, then $\hat{w}_{hik} = M_{hi}/(Mn_h m_{hi} p_{hi})$ and C4 reduces to

$n_T \max_{hi} M_{hi} / (M n_h p_{hi}) = O(1)$, where p_{hi} is the probability of selecting the $(hi)^{th}$ cluster and m_{hi} is the number of elements sampled in the $(hi)^{th}$ sampled cluster.

We will also assume that ${}_c M / M \rightarrow {}_c \lambda > 0$ as $L \rightarrow \infty$. This assumption is not unreasonable as the poststratified estimator is not defined if a poststratum is empty. Before proceeding, we will prove the following lemma.

Lemma 2.7 Let ${}_c \hat{M} = \sum_{(hik) \in S} \hat{w}_{hik} {}_c \delta_{hik}$. Then under C'_4 , we have

$$n_T V({}_c \hat{M}) = O(1)$$

and ${}_c M / {}_c \hat{M} \xrightarrow{p} 1$ for all c .

Proof

We have

$$\begin{aligned} n_T V({}_c \hat{M}) &= n_T V \left(\sum_{(hik) \in S} \hat{w}_{hik} {}_c \delta_{hik} \right) \\ &= n_T \sum_h \sum_i V \left(\sum_k \hat{w}_{hik} {}_c \delta_{hik} \right) \\ &\leq n_T \sum_h \sum_i E \left(\sum_k \hat{w}_{hik} {}_c \delta_{hik} \right)^2 \\ &\leq n_T \sum_h \sum_i E \left(\sum_k \hat{w}_{hik}^2 \right) \left(\sum_k {}_c \delta_{hik}^2 \right) \\ &\leq n_T \sum_h \sum_i m_{hi} E \left(\sum_k \hat{w}_{hik}^2 \right) \\ &\leq n_T \max_{h,i,k} m_{hi} \hat{w}_{i+k} E \left(\sum_{(hik) \in S} \hat{w}_{hik} \right) \\ &= n_T \max_{h,i,k} m_{hi} \hat{w}_{hik} \\ &= O(1) \end{aligned}$$

since $E \left(\sum_{(hik) \in S} \hat{w}_{hik} \right) = 1$.

As a consequence we have,

$$P \left\{ |{}_c\hat{M} - {}_cM/M| \geq \epsilon \right\} \leq \frac{1}{\epsilon^2} V({}_c\hat{M}) \rightarrow 0$$

or

$${}_c\hat{M} \longrightarrow_p {}_cM/M.$$

Since the function $1/x$ is continuous for $x \neq 0$, we have

$$\frac{1}{{}_c\hat{M}} \longrightarrow_p \frac{1}{{}_cM/M} = \frac{M}{{}_cM}$$

and,

$$\begin{aligned} \frac{{}_cM}{{}_c\hat{M}} &= \frac{{}_cM/M}{{}_c\hat{M}} \\ &\longrightarrow_p \frac{{}_cM}{M} \frac{M}{{}_cM} = 1 \end{aligned}$$

i.e.

$$\frac{{}_cM}{{}_c\hat{M}} \longrightarrow_p 1.$$

■

To show consistency of $\bar{y}_{PS}$, it is sufficient to have

$$(C3') \quad n_T^\delta \sum_h \sum_i E |\epsilon_{hi} - E(\epsilon_{hi})|^{1+\delta} = O(1) \text{ for some } \delta > 0$$

which is weaker than C3. Replacing the y_{hi} with ϵ_{hi} in the proof of Lemma 2.6, we can show that C3 implies C3'. Asymptotic consistency of $\bar{y}_{PS}$ is shown in the following theorem.

Theorem 2.4 *Under C3', C4 and C5, $\bar{y}_{PS}$ is consistent for $\bar{Y}$, i.e.*

$$\bar{y}_{PS} \longrightarrow_p \bar{Y}.$$

Proof

Recall that

$$\hat{y}_{PS} - \bar{Y} = \bar{c} + \sum_c \left(\frac{cM}{c\hat{M}} - 1 \right) c\epsilon. \quad (2.3.9)$$

We need to show that both terms in (2.3.9) converge to 0 in probability. Considering the first term in (2.3.9), to apply Lemma 2.2 let

$$X_{hi} = n_T(\epsilon_{hi} - E(\epsilon_{hi})),$$

then by C3' we get

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{1+\epsilon} &= n_T^{\frac{1}{2}} \sum_h \sum_i E|\epsilon_{hi} - E(\epsilon_{hi})|^{1+\epsilon} \\ &= O(1). \end{aligned}$$

Thus,

$$\frac{1}{n_T} \sum_h \sum_i X_{hi} \rightarrow_p \frac{1}{n_T} \sum_h \sum_i E(X_{hi}) = 0.$$

Now,

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i X_{hi} &= \sum_h \sum_i (\epsilon_{hi} - E(\epsilon_{hi})) \\ &= \bar{c} - E(\bar{c}) \end{aligned}$$

and

$$\begin{aligned} E(\bar{c}) &= E \left(\sum_{(hik) \in s} \hat{w}_{hik} y_{hik} - \sum_c cY_c \delta_{hik} \right) \\ &= E(y) - \sum_c cY_c E \left(\sum_{(hik) \in s} \hat{w}_{hik} \delta_{hik} \right) \\ &= \bar{Y} - \sum_c cY_c \frac{cM}{M} \\ &= \bar{Y} - \bar{Y} = 0 \end{aligned}$$

since $E({}_c\tilde{M}) = {}_cM/M$. Therefore we have,

$$\bar{\epsilon} \longrightarrow_p 0.$$

Next consider the second term in (2.3.9). By Lemma 2.7, we have

$$\frac{{}_cM}{{}_c\tilde{M}} - 1 \longrightarrow_p 0$$

and under C5 we have, noting that $E({}_c\bar{\epsilon}) = 0$,

$$P\{|{}_c\bar{\epsilon}| > \epsilon\} \leq \frac{V({}_c\bar{\epsilon})}{\epsilon^2} \rightarrow 0$$

or ${}_c\epsilon \longrightarrow_p 0$. Thus

$$\left(\frac{{}_cM}{{}_c\tilde{M}} - 1\right) {}_c\bar{\epsilon} \longrightarrow_p 0$$

for all c . Therefore, we have

$$\begin{aligned} y_{PS} - \bar{Y} &= \frac{1}{n_T} \sum_h \sum_i X_{hi} + \sum_c \left(\frac{{}_cM}{{}_c\tilde{M}} - 1\right) {}_c\bar{\epsilon} \\ &\longrightarrow_p 0 \end{aligned}$$

or

$$\bar{y}_{PS} \longrightarrow_p \bar{Y}.$$

■

Before establishing the asymptotic normality of $\bar{y}_{PS}$, we need the additional condition

C6 $n_TV(\epsilon) \rightarrow \sigma_\epsilon^2$, a positive constant as $n_T \rightarrow \infty$ where $\bar{\epsilon} = \sum_{(hik)\epsilon s} \tilde{w}_{hik} \epsilon_{hik}$.

We also need the following lemma.

Lemma 2.8 *Let C_ν and Z_ν be random variables. Then if $C_\nu \longrightarrow_p 0$ and $\{Z_\nu\}_\nu$ is stochastically bounded, then $C_\nu Z_\nu \longrightarrow_p 0$ as $\nu \rightarrow \infty$.*

Note that a sequence, $\{Z_\nu\}$, is stochastically bounded if for each $\epsilon > 0$ there exists an a such that

$$P\{|Z_\nu| > a\} < \epsilon$$

for all ν sufficiently large. A useful sufficient condition for $\{Z_\nu\}_\nu$ to be stochastically bounded is the uniform boundedness of $\{E|Z_\nu|\}_\nu$ which follows from the Markov inequality. We now establish the asymptotic normality of y_{PS} .

Theorem 2.5 *Under C3, C4, C5 and C6,*

$$\sqrt{n_T}(y_{PS} - Y) \longrightarrow_d N(0, \sigma_\epsilon^2).$$

Proof

Again we will consider the two terms of (2.3.9) separately. To apply Lemma 2.1 to the first term, let

$$X_{ht} = n_T(\epsilon_{ht} - E(\epsilon_{ht})).$$

Under C6,

$$\frac{1}{n_T} \sum_h \sum_i V(X_{ht}) \longrightarrow \sigma_\epsilon^2,$$

satisfying condition 1 of Lemma 2.1. Now C3 gives

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{ht}|^{2+\delta} &= n_T^{1+\delta} \sum_h \sum_i E|\epsilon_{ht} - E(\epsilon_{ht})|^{2+\delta} \\ &= O(1) \end{aligned}$$

satisfying condition 2 of Lemma 2.1. Therefore,

$$\sqrt{n_T} \left(\frac{1}{n_T} \sum_h \sum_i X_{ht} \right) = \sqrt{n_T} \bar{\epsilon} \longrightarrow_d N(0, \sigma_\epsilon^2).$$

Condition C5 implies that $\sqrt{n_T}(\epsilon \bar{\epsilon} - E(\epsilon \bar{\epsilon})) = \sqrt{n_T} \epsilon \bar{\epsilon} = O_p(1)$. Lemma 2.8 then implies

$$\sqrt{n_T} \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \epsilon \bar{\epsilon} \longrightarrow_p 0$$

since ${}_cM/\dot{c}\dot{M} - 1 \rightarrow_p 0$ by Lemma 2.7. Thus, from Lemma 2.1 and Lemma 2.3,

$$\begin{aligned}\sqrt{n_T}(\bar{y}_{PS} - \bar{Y}) &= \sqrt{n_T} \frac{1}{n_T} \sum_h \sum_i X_{hi} + \sqrt{n_T} \sum_c \left(\frac{{}_cM}{\dot{c}\dot{M}} - 1 \right) {}_c\bar{e} \\ &\rightarrow_d N(0, \sigma_c^2).\end{aligned}$$

■

One may note that the poststratified estimator can be regarded as a smooth function of totals and then appeal to asymptotic results for such functions (Krewski and Rao, 1981), but the regularity conditions are not as simple as C3 to C6. Scott and Wu (1981) used regularity conditions similar to C3 to C6 in the context of ratio and regression estimation under simple random sampling.

A customary linearization variance estimator of $V(\bar{y}_{PS})$ is given as

$$v_L(\bar{y}_{PS}) = \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi}^* - \bar{\epsilon}_h^*)^2, \quad (2.3.10)$$

where $\epsilon_{hi}^* = \sum_k \hat{w}_{hik} \epsilon_{hik}^*$, $\bar{\epsilon}_h^* = \frac{1}{n_h} \sum_i \epsilon_{hi}^*$ and ϵ_{hik}^* is the estimated residual defined as $\epsilon_{hik}^* = y_{hik} - \sum_c {}_c\bar{y} {}_c\delta_{hik}$ with ${}_c\bar{y} = {}_c\bar{Y}/{}_c\dot{M}$. Consistency of $v_L(\bar{y}_{PS})$ is now shown in the following theorem.

Theorem 2.6 *Under C3, C4, C5 and C6*

$$n_T(v_L(\bar{y}_{PS}) - V(\bar{y}_{PS})) \rightarrow_p 0.$$

Proof

By adding and subtracting $\sum_c {}_c\bar{Y} {}_c\delta_{hik}$ to ϵ_{hik}^* , we get

$$\begin{aligned}\epsilon_{hik}^* &= y_{hik} - \sum_c {}_c\bar{y} {}_c\delta_{hik} \\ &= \epsilon_{hik} - \sum_c ({}_c\bar{y} - {}_c\bar{Y}) {}_c\delta_{hik}\end{aligned}$$

where $\epsilon_{hik} = y_{hik} - \sum_c \bar{Y}_c \delta_{hik}$. Now consider ${}_c\bar{y} - {}_c\bar{Y}$

$$\begin{aligned} {}_c\bar{y} - {}_c\bar{Y} &= \frac{{}_c\bar{Y}}{{}_c\bar{M}} - {}_c\bar{Y} \\ &= \frac{{}_c\bar{Y} - {}_c\bar{Y} {}_c\bar{M}}{{}_c\bar{M}} \\ &= \frac{\sum_{(hik)s} \hat{w}_{hik} \delta_{hik} (y_{hik} - {}_c\bar{Y})}{{}_c\bar{M}} \\ &= \frac{{}_c\epsilon}{{}_c\bar{M}}. \end{aligned}$$

Therefore,

$$\begin{aligned} \epsilon_{hi}^* &= \epsilon_{hi} - \sum_c \frac{{}_c\epsilon}{{}_c\bar{M}} \sum_k \hat{w}_{hik} \delta_{hik} \\ \bar{\epsilon}_h^* &= \bar{\epsilon}_h - \sum_c \frac{{}_c\epsilon}{{}_c\bar{M}} \frac{1}{n_h} \sum_i \sum_k \hat{w}_{hik} \delta_{hik} \quad \text{and} \\ \epsilon_{hi}^* - \bar{\epsilon}_h^* &= \epsilon_{hi} - \bar{\epsilon}_h - \sum_c \frac{{}_c\epsilon}{{}_c\bar{M}} ({}_c\hat{M}_{hi} - {}_c\hat{M}_h) \end{aligned}$$

where ${}_c\hat{M}_{hi} = \sum_k \hat{w}_{hik} \delta_{hik}$ and ${}_c\hat{M}_h = (1/n_h) \sum_i {}_c\hat{M}_{hi}$.

Putting the expression for $(\epsilon_{hi}^* - \bar{\epsilon}_h^*)$ in (2.3.10) gives

$$\begin{aligned} v_L(\bar{y}_{PS}) &= \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi}^* - \bar{\epsilon}_h^*)^2 \\ &= \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\epsilon_{hi} - \bar{\epsilon}_h - \sum_c \frac{{}_c\epsilon}{{}_c\bar{M}} ({}_c\hat{M}_{hi} - {}_c\hat{M}_h) \right]^2 \\ &= \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi} - \bar{\epsilon}_h)^2 + \\ &\quad \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \frac{{}_c\epsilon}{{}_c\bar{M}} ({}_c\hat{M}_{hi} - {}_c\hat{M}_h) \right]^2 - \\ &\quad 2 \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \frac{{}_c\epsilon}{{}_c\bar{M}} ({}_c\hat{M}_{hi} - {}_c\hat{M}_h) \right] (\epsilon_{hi} - \bar{\epsilon}_h) \\ &= v_L^*(\bar{y}_{PS}) + v_1 + v_2 \end{aligned}$$

where

$$v_L^*(\bar{y}_{PS}) = \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi} - \bar{\epsilon}_h)^2$$

$$\begin{aligned}
v_1 &= \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \frac{c\bar{\epsilon}}{c\hat{M}} \left({}_c\hat{M}_{hi} - {}_c\hat{M}_h \right) \right]^2 \quad \text{and} \\
v_2 &= -2 \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \frac{c\bar{\epsilon}}{c\hat{M}} \left({}_c\hat{M}_{hi} - {}_c\hat{M}_h \right) \right] (\epsilon_{hi} - \bar{\epsilon}_h).
\end{aligned}$$

To show consistency of $v_L(\hat{y}_{PS})$, we will show

$$n_T (v_L^*(\hat{y}_{PS}) - V(\hat{y}_{PS})) \longrightarrow_p 0,$$

$$n_T v_1 \longrightarrow_p 0 \quad \text{and} \quad n_T v_2 \longrightarrow_p 0.$$

Note that $v_L^*(\hat{y}_{PS})$ is not a true variance estimator as it involves the unknown residuals, ϵ_{hik} . We can rewrite $v_L^*(\hat{y}_{PS})$ as

$$v_L^*(\hat{y}_{PS}) = \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi} - E(\epsilon_{hi}))^2 - \sum_h \frac{n_h^2}{n_h - 1} (\bar{\epsilon}_h - E(\epsilon_{hi}))^2.$$

Let

$$X_{hi} = n_T (n_h / (n_h - 1))^{1/2} (\epsilon_{hi} - E(\epsilon_{hi}))$$

and note that $n_h \geq 2$ implies

$$\frac{1}{n_h - 1} \leq \frac{2}{n_h}.$$

Thus,

$$\begin{aligned}
\frac{1}{n_I} \sum_h \sum_i E |X_{hi}^2|^{1+\delta/2} &= n_T^{1+\delta} \sum_h \sum_i \left(\frac{n_h}{n_h - 1} \right)^{1+\delta/2} E |\epsilon_{hi} - E(\epsilon_{hi})|^{2+\delta} \\
&\leq 2^{1+\delta/2} n_T^{1+\delta} \sum_h \sum_i E |\epsilon_{hi} - E(\epsilon_{hi})|^{2+\delta} \\
&= O(1),
\end{aligned}$$

by C3 and Lemma 2.2 gives

$$n_T \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi} - E(\epsilon_{hi}))^2 \longrightarrow_p n_T \sum_h \frac{n_h^2}{n_h - 1} \sigma_{\epsilon h}^2$$

where $\sigma_{\epsilon_h}^2 = V(\epsilon_h)$. Next, let

$$X_h = (Ln_T)^{1/2} \frac{n_h}{(n_h - 1)^{1/2}} (\epsilon_h - E(\epsilon_h))$$

and $0 < \delta'/2 \leq \min(\delta/2, 1/2)$. Noting that $L < n_I$, we follow the proof of theorem 2.3 with y_{ht} replaced by ϵ_{ht} to get

$$n_T \sum_h \frac{n_h^2}{n_h - 1} (\epsilon_h - E(\epsilon_h))^2 \longrightarrow_p n_T \sum_h \frac{n_h}{n_h - 1} \sigma_{\epsilon_h}^2.$$

Combining these two results and noting that $V(y_{PS}) \approx V(\epsilon)$ gives,

$$n_T (v_L^*(\hat{y}_{PS}) - V(y_{PS})) \longrightarrow_p 0.$$

Now it remains to show $n_T v_1 \longrightarrow_p 0$ and $n_T v_2 \longrightarrow_p 0$. Consider $n_T v_1$,

$$\begin{aligned} n_T v_1 &= n_T \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \frac{c\epsilon}{c\hat{M}} (\hat{M}_{hi} - \hat{M}_h) \right]^2 \\ &\leq n_T \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \left(\frac{c\epsilon}{c\hat{M}} \right)^2 \right] \left[\sum_i (\hat{M}_{hi} - \hat{M}_h)^2 \right] \\ &= \left[\sum_c c\epsilon^2 \left(\frac{c\hat{M}}{c\hat{M}} \right)^2 \left(\frac{\hat{M}}{c\hat{M}} \right)^2 \right] \left[\sum_c n_T \sum_h \frac{n_h}{n_h - 1} \sum_i (\hat{M}_{hi} - \hat{M}_h)^2 \right] \\ &= \left[\sum_c c\bar{\epsilon}^2 \left(\frac{c\hat{M}}{c\hat{M}} \right)^2 \left(\frac{\hat{M}}{c\hat{M}} \right)^2 \right] \left[\sum_c n_T v_L(c\hat{M}) \right]. \end{aligned}$$

Now, $c\bar{\epsilon} \longrightarrow_p 0$, $c\hat{M}/c\hat{M} \longrightarrow_p 1$, and $\hat{M}/c\hat{M} \rightarrow 1/c\lambda < \infty$ for all c . Noting that $n_T v_L(c\hat{M})$ is stochastically bounded for all c since

$$\begin{aligned} E(|n_T v_L(c\hat{M})|) &= n_T E(v_L(c\hat{M})) \\ &= n_T V(c\hat{M}) = O(1) \end{aligned}$$

by Lemma 2.7, Lemma 2.8 implies

$$n_T v_1 \longrightarrow_p 0.$$

From the Schwarz inequality, we have

$$n_T v_2 \longrightarrow_p 0.$$

Therefore,

$$n_T v_L(\bar{y}_{PS}) = n_T v_L^*(\bar{y}_{PS}) + n_T v_1 + n_T v_2 \longrightarrow_p n_T V(\bar{y}_{PS})$$

or

$$n_T (v_L(\bar{y}_{PS}) - V(\bar{y}_{PS})) \longrightarrow_p 0.$$

■

2.4 Jackknife Variance Estimation

In order to construct a jackknife variance estimator, we need to compute $\hat{Y}_{PS(gj)}$, the poststratified estimator obtained after omitting the $(gj)^{th}$ sampled cluster. To define $\hat{Y}_{PS(gj)}$, we will first define the jackknife basic weights $w_{hik(gj)}$,

$$w_{hik(gj)} = \begin{cases} 0 & \text{if } (hi) = (gj) \\ \frac{n_g}{n_g - 1} w_{gik} & \text{if } h = g, i \neq j \\ w_{hik} & \text{if } h \neq g \end{cases} \quad (2.4.1)$$

and the estimator of the c^{th} poststratum count with the $(gj)^{th}$ sampled cluster omitted as

$${}_c \hat{M}_{(gj)} = \sum_{(hik) \in s} w_{hik(gj)} {}_c \delta_{hik}.$$

The poststratified adjusted jackknife weight is then defined as

$${}_c u_{hik(gj)} = w_{hik(gj)} \frac{{}_c M}{{}_c \hat{M}_{(gj)}}$$

and the associated estimator of Y as

$$\hat{Y}_{PS(gj)} = \sum_c \sum_{(hik) \in s} {}_c w_{hik(gj)} y_{hik} {}_c \delta_{hik}. \quad (2.4.2)$$

Finally the jackknife variance estimator of $\hat{Y}_{PS}$ is defined as

$$v_J(\hat{Y}_{PS}) = \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\hat{Y}_{PS(gj)} - \hat{Y}_{PS} \right)^2. \quad (2.4.3)$$

Before showing the consistency of the jackknife variance estimator, we need the following result.

Lemma 2.9 *Under C4,*

$$\max_{h,i} |{}_c \hat{M}_{(hi)} - {}_c \hat{M}| \longrightarrow_p 0$$

for all c .

Proof

Let $\epsilon > 0$ be given. Then

$$\begin{aligned} P \left\{ \max_{h,i} |{}_c \hat{M}_{(hi)} - {}_c \hat{M}| > \epsilon \right\} &= P \left\{ \text{at least one } |{}_c \hat{M}_{(hi)} - {}_c \hat{M}| > \epsilon \right\} \\ &\leq \sum_h \sum_i P \left\{ |{}_c \hat{M}_{(hi)} - {}_c \hat{M}| > \epsilon \right\} \\ &\leq \sum_i \sum_h \frac{1}{\epsilon^2} E \left({}_c \hat{M}_{(hi)} - {}_c \hat{M} \right)^2. \end{aligned}$$

Noting that

$${}_c \hat{M}_{(hi)} - {}_c \hat{M} = \frac{n_h}{n_h - 1} \left(\frac{1}{n_h} \sum_{(gik) \in s} \tilde{w}_{gik} {}_c \delta_{gik} - \sum_{(gjk) \in s} \tilde{w}_{gjk} {}_c \delta_{gjk} \right)$$

we have

$$P \left\{ \max_{h,i} |{}_c \hat{M}_{(hi)} - {}_c \hat{M}| > \epsilon \right\}$$

$$\begin{aligned}
&\leq \frac{1}{\epsilon^2} \sum_h \left(\frac{n_h}{n_h - 1} \right)^2 \sum_i E \left(\frac{1}{n_h} \sum_{(gik)\epsilon s} \tilde{w}_{gik} \epsilon \delta_{gik} - \sum_{(gjk)\epsilon s} \tilde{w}_{gjk} \epsilon \delta_{gjk} \right)^2 \\
&\leq \frac{1}{\epsilon^2} 2E \left[\sum_h \frac{n_h}{n_h - 1} \sum_i \left(\frac{1}{n_h} \sum_{(gik)\epsilon s} \tilde{w}_{gik} \epsilon \delta_{gik} - \sum_{(gjk)\epsilon s} \tilde{w}_{gjk} \epsilon \delta_{gjk} \right)^2 \right] \\
&= \frac{1}{\epsilon^2} 2V({}_\epsilon \hat{M}) \rightarrow 0.
\end{aligned}$$

Since ϵ was arbitrary, we have

$$\max_{h,i} |{}_c \hat{M}_{(hi)} - {}_\epsilon \hat{M}| \rightarrow_p 0.$$

We now establish the consistency of $v_J(\bar{y}_{PS})$ in the following theorem.

Theorem 2.7 *Under conditions C3, C4, C5 and C6*

$$n_T(v_J(\bar{y}_{PS}) - V(\bar{y}_{PS})) \rightarrow_p 0.$$

Proof

Writing $\bar{y}_{PS(gj)}$ as

$$\bar{y}_{PS(gj)} = \sum_c \frac{{}_c M}{M} \frac{{}_c \hat{Y}_{(gj)}}{{}_c \hat{M}_{(gj)}}$$

with

$${}_c \hat{Y}_{(gj)} = \sum_{\substack{(hik)\epsilon s \\ h \neq g}} \tilde{w}_{hik} y_{hik} \epsilon \delta_{hik} + \frac{n_g}{n_g - 1} \sum_{\substack{(gik)\epsilon s \\ i \neq j}} \tilde{w}_{gik} y_{gik} \epsilon \delta_{gik}$$

and

$${}_c \hat{M}_{(gj)} = \sum_{\substack{(hik)\epsilon s \\ h \neq g}} \tilde{w}_{hik} \epsilon \delta_{hik} + \frac{n_g}{n_g - 1} \sum_{\substack{(gik)\epsilon s \\ i \neq j}} \tilde{w}_{gik} \epsilon \delta_{gik}$$

we can write $\bar{y}_{PS(gj)} - \bar{y}_{PS}$ as

$$\bar{y}_{PS(gj)} - \bar{y}_{PS} = \sum_c \frac{{}_c M}{M} \left(\frac{{}_c \hat{Y}_{(gj)}}{{}_c \hat{M}_{(gj)}} - \frac{{}_c \hat{Y}}{{}_c \hat{M}} \right). \quad (2.4.4)$$

Expanding the function ${}_c\hat{M}^{-1}$ about the point ${}_c\hat{M}$ gives

$$\frac{1}{{}_c\hat{M}_{(gj)}} = \frac{1}{{}_c\hat{M}} - \frac{1}{{}_c\hat{x}_{gj}^2}({}_c\hat{M}_{(gj)} - {}_c\hat{M})$$

where ${}_c\hat{x}_{gj}$ lies between ${}_c\hat{M}_{(gj)}$ and ${}_c\hat{M}$. Let ${}_c\hat{y} = {}_c\hat{Y}/{}_c\hat{M}$, then

$$\begin{aligned} \frac{{}_c\hat{Y}_{(gj)}}{{}_c\hat{M}_{(gj)}} - \frac{{}_c\hat{Y}}{{}_c\hat{M}} &= \frac{1}{{}_c\hat{M}_{(gj)}}({}_c\hat{Y}_{(gj)} - {}_c\hat{y}{}_c\hat{M}_{(gj)}) \\ &= \frac{1}{{}_c\hat{M}_{(gj)}}[{}_c\hat{Y}_{(gj)} - {}_c\hat{Y} - {}_c\hat{y}({}_c\hat{M}_{(gj)} - {}_c\hat{M})] \\ &= \frac{1}{{}_c\hat{M}_{(gj)}} \frac{n_g}{n_g - 1}({}_c\hat{\epsilon}_g^* - {}_c\hat{\epsilon}_{gj}^*) \end{aligned} \quad (2.4.5)$$

where ${}_c\hat{\epsilon}_{gj}^* = \sum_k \hat{w}_{gjk} {}_c\hat{\epsilon}_{hik}^*$, ${}_c\hat{\epsilon}_g^* = (1/n_g) \sum_j {}_c\hat{\epsilon}_{gj}^*$ and ${}_c\hat{\epsilon}_{hik}^* = (y_{hik} - {}_c\hat{y}) {}_c\hat{\delta}_{hik}$ is the estimated residual. The estimated residual, ${}_c\hat{\epsilon}_{hik}^*$, can be written as

$${}_c\hat{\epsilon}_{hik}^* = {}_c\hat{\epsilon}_{hik} - \frac{{}_c\hat{\epsilon}}{{}_c\hat{M}} {}_c\hat{\delta}_{hik}$$

which gives

$${}_c\hat{\epsilon}_{hi}^* - {}_c\hat{\epsilon}_h^* = {}_c\hat{\epsilon}_{hi} - {}_c\hat{\epsilon}_h - \frac{{}_c\hat{\epsilon}}{{}_c\hat{M}}({}_c\hat{M}_{hi} - {}_c\hat{M}_h) \quad (2.4.6)$$

with ${}_c\hat{M}_{hi} = \sum_k \hat{w}_{hik} {}_c\hat{\delta}_{hik}$ and ${}_c\hat{M}_h = (1/n_h) \sum_i {}_c\hat{M}_{hi}$. Putting (2.4.6) into (2.4.5) gives

$$\begin{aligned} \frac{{}_c\hat{Y}_{(gj)}}{{}_c\hat{M}_{(gj)}} - \frac{{}_c\hat{Y}}{{}_c\hat{M}} &= \frac{1}{{}_c\hat{M}_{(gj)}} \frac{n_g}{n_g - 1} \left({}_c\hat{\epsilon}_g - {}_c\hat{\epsilon}_{gj} - \frac{{}_c\hat{\epsilon}}{{}_c\hat{M}}({}_c\hat{M}_{gj} - {}_c\hat{M}_g) \right) \\ &= \frac{1}{{}_c\hat{M}} \frac{n_g}{n_g - 1} \left({}_c\hat{\epsilon}_g - {}_c\hat{\epsilon}_{gj} - \frac{{}_c\hat{\epsilon}}{{}_c\hat{M}}({}_c\hat{M}_{gj} - {}_c\hat{M}_g) \right) - \\ &\quad \frac{1}{{}_c\hat{x}_{gj}^2} \frac{n_g}{n_g - 1} \left({}_c\hat{\epsilon}_g - {}_c\hat{\epsilon}_{gj} - \frac{{}_c\hat{\epsilon}}{{}_c\hat{M}}({}_c\hat{M}_{gj} - {}_c\hat{M}_g) \right) ({}_c\hat{M}_{(gj)} - {}_c\hat{M}) \\ &= \frac{1}{{}_c\hat{M}} \frac{n_g}{n_g - 1} ({}_c\hat{\epsilon}_g - {}_c\hat{\epsilon}_{gj}) - \frac{1}{{}_c\hat{x}_{gj}^2} \frac{n_g}{n_g - 1} ({}_c\hat{\epsilon}_g - {}_c\hat{\epsilon}_{gj}) ({}_c\hat{M}_{(gj)} - {}_c\hat{M}) - \\ &\quad \frac{1}{{}_c\hat{M}} \frac{n_g}{n_g - 1} \frac{{}_c\hat{\epsilon}}{{}_c\hat{M}} ({}_c\hat{M}_{gj} - {}_c\hat{M}_g) + \frac{1}{{}_c\hat{x}_{gj}^2} \frac{n_g}{n_g - 1} \frac{{}_c\hat{\epsilon}}{{}_c\hat{M}} ({}_c\hat{M}_{gj} - {}_c\hat{M}_g) ({}_c\hat{M}_{(gj)} - {}_c\hat{M}). \end{aligned}$$

Putting this expression in (2.4.4) gives

$$\begin{aligned} y_{PS(gj)} - \bar{y}_{PS} &= \sum_c \frac{{}_cM}{M} \frac{n_g}{n_g - 1} \left[\frac{1}{{}_c\tilde{M}} ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj}) - \frac{1}{{}_c\tilde{x}_{gj}^2} ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj}) ({}_c\tilde{M}_{(gj)} - {}_c\tilde{M}) - \right. \\ &\quad \left. \frac{1}{{}_c\tilde{M}} \frac{{}_c\bar{\epsilon}}{{}_c\tilde{M}} ({}_c\tilde{M}_{gj} - {}_c\tilde{M}_g) + \frac{1}{{}_c\tilde{x}_{gj}^2} \frac{{}_c\bar{\epsilon}}{{}_c\tilde{M}} ({}_c\tilde{M}_{gj} - {}_c\tilde{M}_g) ({}_c\tilde{M}_{(gj)} - {}_c\tilde{M}) \right] \end{aligned}$$

and the jackknife variance estimator as

$$\begin{aligned} v_J(y_{PS}) &= \sum_g \frac{n_g - 1}{n_g} \sum_j (y_{PS(gj)} - \bar{y}_{PS})^2 \\ &= \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{M} \left\{ \frac{1}{{}_c\tilde{M}} ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj}) - \frac{1}{{}_c\tilde{x}_{gj}^2} ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj}) ({}_c\tilde{M}_{(gj)} - {}_c\tilde{M}) - \right. \right. \\ &\quad \left. \left. \frac{1}{{}_c\tilde{M}} \frac{{}_c\bar{\epsilon}}{{}_c\tilde{M}} ({}_c\tilde{M}_{gj} - {}_c\tilde{M}_g) + \frac{1}{{}_c\tilde{x}_{gj}^2} \frac{{}_c\bar{\epsilon}}{{}_c\tilde{M}} ({}_c\tilde{M}_{gj} - {}_c\tilde{M}_g) ({}_c\tilde{M}_{(gj)} - {}_c\tilde{M}) \right\} \right]^2 \\ &= \hat{v}_1 + \hat{v}_2 + \hat{v}_3 + \hat{v}_4 + 2(\text{cross-product terms}) \end{aligned}$$

where

$$\begin{aligned} \hat{v}_1 &= \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{M} ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj}) \right]^2, \\ \hat{v}_2 &= \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{M} \frac{1}{{}_c\tilde{x}_{gj}^2} ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj}) ({}_c\tilde{M}_{(gj)} - {}_c\tilde{M}) \right]^2, \\ \hat{v}_3 &= \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{M} \frac{1}{{}_c\tilde{M}^2} \frac{{}_c\bar{\epsilon}}{{}_c\tilde{M}} ({}_c\tilde{M}_{gj} - {}_c\tilde{M}_g) \right]^2 \quad \text{and} \\ \hat{v}_4 &= \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{M} \frac{1}{{}_c\tilde{x}_{gj}^2} \frac{{}_c\bar{\epsilon}}{{}_c\tilde{M}} ({}_c\tilde{M}_{gj} - {}_c\tilde{M}_g) ({}_c\tilde{M}_{(gj)} - {}_c\tilde{M}) \right]^2. \end{aligned}$$

To establish the consistency of $v_J(\bar{y}_{PS})$ it suffices to show

$$n_T(\hat{v}_1 - V(\bar{y}_{PS})) \longrightarrow_p 0 \quad (2.4.7)$$

$$n_T \hat{v}_2 \longrightarrow_p 0 \quad (2.4.8)$$

$$n_T \hat{v}_3 \longrightarrow_p 0 \quad (2.4.9)$$

$$n_T \hat{v}_4 \longrightarrow_p 0 \quad (2.4.10)$$

since the Schwarz inequality will imply the cross-product terms all converge to zero in probability.

First, consider $\hat{v}_1$. We have

$$\begin{aligned} n_T \hat{v}_1 &= n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_c M}{{}_c \hat{M}} ({}_c \bar{\epsilon}_g - {}_c \epsilon_{gj}) \right]^2 \\ &= n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c ({}_c \bar{\epsilon}_g - {}_c \epsilon_{gj}) + \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) ({}_c \epsilon_{gj} - {}_c \epsilon_g) \right]^2 \\ &= n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c ({}_c \bar{\epsilon}_g - {}_c \epsilon_{gj}) \right]^2 + \\ &\quad n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) ({}_c \epsilon_{gj} - {}_c \epsilon_g) \right]^2 + 2(\text{cross-product terms}) \\ &= n_T \sum_g \frac{n_g}{n_g - 1} \sum_j ({}_c \bar{\epsilon}_g - {}_c \epsilon_{gj})^2 + n_T \hat{v}_{12} + 2(\text{cross-product terms}) \\ &= n_T v_L^*(\bar{y}_{PS}) + n_T \hat{v}_{12} + 2(\text{cross-product terms}) \end{aligned}$$

since $\sum_c ({}_c \bar{\epsilon}_g - {}_c \epsilon_{gj}) = (\bar{\epsilon}_g - \epsilon_{gj})$. We know that $n_T v_L^*(y_{PS}) \longrightarrow_p n_T V(y_{PS})$ from Theorem 2.6, so it remains to show that $n_T \hat{v}_{12} \longrightarrow_p 0$ and the cross-product terms converge to zero in probability. Now,

$$\begin{aligned} n_T \hat{v}_{12} &\leq n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right)^2 \right] \left[\sum_c ({}_c \epsilon_{gj} - {}_c \epsilon_g)^2 \right] \\ &= \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right)^2 n_T \sum_c \sum_g \frac{n_g}{n_g - 1} \sum_j ({}_c \epsilon_{gj} - {}_c \epsilon_g)^2 \\ &= \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right)^2 \right] \left[\sum_c n_T v_L({}_c \epsilon) \right] \\ &\longrightarrow_p 0 \end{aligned}$$

since $({}_c M / {}_c \hat{M}) \longrightarrow_p 1$ and $n_T v_L({}_c \bar{\epsilon})$ is stochastically bounded for all c by C5 noting that $E(n_T v_L({}_c \bar{\epsilon})) = n_T V({}_c \bar{\epsilon})$. Thus $n_T \hat{v}_{12} \longrightarrow_p 0$ and an application of the Schwarz

inequality shows that the cross-products also converge to zero in probability. Therefore, we have

$$n_T(\hat{v}_1 - V(\hat{y}_{PS})) \longrightarrow_p 0.$$

Before showing $n_T \hat{v}_2 \longrightarrow_p 0$, we note that since the function x^{-2} is continuous for $x \neq 0$, for every $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that $\left| \frac{1}{x} - \frac{1}{y} \right| < \epsilon$ whenever $|x - y| < \delta_\epsilon$. Thus, since

$$\max_{g,j} \left| {}_c\hat{M}_{(gj)} - {}_c\hat{M} \right| \longrightarrow_p 0$$

for all c , and ${}_c x_{gj}$ lies between ${}_c\hat{M}_{(gj)}$ and ${}_c\hat{M}$ for all c we have

$$P \left\{ \max_{g,j} \left| \frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right| < \epsilon \right\} \geq P \left\{ \max_{g,j} \left| {}_c\hat{M}_{(gj)} - {}_c\hat{M} \right| < \delta_\epsilon \right\} \rightarrow 1.$$

That is,

$$\max_{g,j} \left| \frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right| \longrightarrow_p 0 \quad (2.4.11)$$

for all c .

Now consider $\hat{v}_2$.

$$\begin{aligned} n_T \hat{v}_2 &\leq n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c\hat{M}}{{}_c\hat{M}} \right)^2 ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj})^2 \right] \left[\sum_c \left(\frac{1}{{}_c x_{gj}^2} \right)^2 ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \right] \\ &\leq n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj})^2 \right] \left[\sum_c \left(\frac{1}{{}_c x_{gj}^2} \right)^2 ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \right] \\ &\leq \max_{g,j,c} ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c ({}_c\bar{\epsilon}_g - {}_c\epsilon_{gj})^2 \right] \times \\ &\quad \left[\sum_c \left(\frac{1}{{}_c\hat{M}^2} + \frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
&= \max_{g,j,c} (\hat{M}_{(gj)} - \hat{M})^2 n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c (\epsilon_{gj} - \epsilon_{gj})^2 \right] \times \\
&\quad \left[\sum_c \left\{ \frac{1}{\hat{M}^4} + \left(\frac{1}{\epsilon_{gj}^2} - \frac{1}{\hat{M}^2} \right)^2 + 2 \frac{1}{\hat{M}^2} \left(\frac{1}{\epsilon_{gj}^2} - \frac{1}{\hat{M}^2} \right) \right\} \right] \\
&\leq \max_{g,j,c} (\hat{M}_{(gj)} - \hat{M})^2 \left(\sum_c \left(\frac{\hat{M}}{\epsilon_{gj}} \right)^4 \left(\frac{\hat{M}}{\epsilon_{gj}} \right)^4 \right) \left[\sum_c n_{TVL}(\epsilon) \right] + \\
&\quad \max_{g,j,c} (\hat{M}_{(gj)} - \hat{M})^2 \max_{g,j,c} \left(\frac{1}{\epsilon_{gj}^2} - \frac{1}{\hat{M}^2} \right)^2 C' \left[\sum_c n_{TVL}(\epsilon) \right] + \\
&\quad 2 \max_{g,j,c} (\hat{M}_{(gj)} - \hat{M})^2 n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c (\epsilon_{gj} - \epsilon_{gj})^2 \right] \times \\
&\quad \left[\sum_c \frac{1}{\hat{M}^2} \left(\frac{1}{\epsilon_{gj}^2} - \frac{1}{\hat{M}^2} \right) \right] \\
&\rightarrow_p 0
\end{aligned}$$

since $\max_{g,j,c} (\hat{M}_{(gj)} - \hat{M}) \rightarrow_p 0$, $\epsilon \hat{M} / \epsilon \hat{M} \rightarrow_p 1$, $\hat{M} / \epsilon \hat{M} \rightarrow 1/\epsilon \lambda < \infty$ and $n_{TVL}(\epsilon \bar{\epsilon})$ is stochastically bounded for all ϵ implies that the first term converges to zero in probability. The additional results that

$$\max_{g,j,c} \left(\frac{1}{\epsilon_{gj}^2} - \frac{1}{\hat{M}^2} \right)^2 \rightarrow_p 0,$$

and the number of poststrata, C' , is finite implies that the second term converges to zero in probability. The Schwarz inequality then implies that the third term converges to zero in probability.

Next, consider $\hat{v}_3$.

$$\begin{aligned}
n_T \hat{v}_3 &\leq n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{\hat{M}}{\epsilon_{gj}} \right)^2 \frac{1}{\hat{M}^4} \epsilon_{gj}^2 \right] \left[\sum_c (\hat{M}_{gj} - \hat{M}_g)^2 \right] \\
&= \sum_c \left(\frac{\hat{M}}{\epsilon_{gj}} \right)^6 \left(\frac{\hat{M}}{\epsilon_{gj}} \right)^4 \epsilon_{gj}^2 [n_{TVL}(\epsilon \bar{\epsilon})] \\
&\rightarrow_p 0
\end{aligned}$$

since ${}_cM/{}_c\hat{M} \rightarrow_p 1$, $M/{}_cM \rightarrow 1/{}_c\lambda < \infty$, ${}_c\bar{\epsilon} \rightarrow_p 0$ and $n_T v_L({}_c\hat{M})$ is stochastically bounded for all c . Finally, we need to show that $n_T \hat{v}_4$ converges to zero in probability.

$$\begin{aligned}
n_T \hat{v}_4 &= n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{\hat{M}} \frac{1}{{}_c x_{gj}^2} \frac{{}_c\bar{\epsilon}}{{}_c\hat{M}} ({}_c\hat{M}_{gj} - {}_c\hat{M}_g) ({}_c\hat{M}_{(gj)} - {}_c\hat{M}) \right]^2 \\
&\leq \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_cM}{\hat{M}} \right)^2 {}_c\bar{\epsilon}^2 ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \right] \left[\sum_c \left(\frac{1}{{}_c x_{gj}^2} \right)^2 ({}_c\hat{M}_{gj} - {}_c\hat{M}_g)^2 \right] \\
&\leq \max_{g,j,c} ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \left[\sum_c \left(\frac{{}_cM}{\hat{M}} \right)^2 {}_c\bar{\epsilon}^2 \right] n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \sum_c \left[\frac{1}{{}_c\hat{M}^4} + \right. \\
&\quad \left. \left(\frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right)^2 + 2 \frac{1}{{}_c\hat{M}^2} \left(\frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right) \right] ({}_c\hat{M}_{gj} - {}_c\hat{M}_g)^2 \\
&\leq \max_{g,j,c} ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \left[\sum_c \left(\frac{{}_cM}{\hat{M}} \right)^2 {}_c\bar{\epsilon}^2 \right] \sum_c \left(\frac{{}_cM}{\hat{M}} \right)^4 \left(\frac{M}{{}_cM} \right)^4 n_T v_L({}_c\hat{M}) + \\
&\quad \max_{g,j,c} ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \max_{g,j,c} \left(\frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right)^2 \left[\sum_c \left(\frac{{}_cM}{\hat{M}} \right)^2 {}_c\bar{\epsilon}^2 \right] \sum_c n_T v_L({}_c\hat{M}) + \\
&\quad 2 \max_{g,j,c} ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \left[\sum_c \left(\frac{{}_cM}{\hat{M}} \right)^2 {}_c\bar{\epsilon}^2 \right] \times \\
&\quad \left[\sum_c n_T \sum_g \frac{n_g}{n_g - 1} \sum_j ({}_c\hat{M}_{gj} - {}_c\hat{M}_g)^2 \frac{1}{{}_c\hat{M}^2} \left(\frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right) \right] \\
&\rightarrow_p 0
\end{aligned}$$

since $\max_{g,j,c} ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \rightarrow_p 0$, ${}_cM/{}_c\hat{M} \rightarrow_p 1$, ${}_c\bar{\epsilon} \rightarrow_p 0$, $M/{}_cM \rightarrow 1/{}_c\lambda < \infty$, and $n_T v_L({}_c\hat{M})$ is stochastically bounded implies the first term converges to zero in probability. With the additional fact that

$$\max_{g,j,c} \left(\frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right)^2 \rightarrow_p 0,$$

the second term also converges to zero in probability. An application of the Schwarz inequality implies that the third term also converges to zero in probability.

Summarizing the above results, we have

$$n_T(\hat{v}_1 - V(y_{PS})) \longrightarrow_p 0$$

$$n_T \hat{v}_2 \longrightarrow_p 0$$

$$n_T \hat{v}_3 \longrightarrow_p 0$$

and

$$n_T \hat{v}_4 \longrightarrow_p 0.$$

Use of the Schwarz inequality will imply the cross-product terms all converge to zero in probability. Thus, we have

$$n_T v_J(\hat{y}_{PS}) \longrightarrow_p n_T V(y_{PS})$$

or

$$n_T(v_J(\hat{y}_{PS}) - V(y_{PS})) \longrightarrow_p 0.$$

■

Chapter 3

Two-way Poststratification

3.1 Introduction

Valliant(1993) uses the superpopulation approach as an alternative to the design based theory for complex surveys. Under a superpopulation model, the observed responses, y_{hik} , are treated as realizations of random variables while the auxiliary variables, x_{hik} , are treated as fixed. The resulting inferences are conditioned on the sample, s . Using this approach, Valliant (1993) investigated the properties of several commonly used variance estimators in the case of the one-way poststratified estimator, $\hat{Y}_{PS}$.

In practice, it is common to form poststrata according to two or more auxiliary variables. If the resulting poststrata population totals are available, the one-way poststratified estimator can be used to increase the efficiency of the estimates. However, these population totals may not be known in practice. For instance, totals may be known only for age groups and race groups but not for individual age-race groups. This means that in terms of a two-way table, the marginal totals are known but not the individual cell totals. The situation has been referred to as "incomplete

poststratification" by Deville and Särndal (1992) and can be handled by using a generalized regression estimator(GREG).

In this chapter, the work of Valliant will be presented in Section 3.2. Valliant's superpopulation model is introduced as well as his asymptotic setup. The poststratified estimator is shown to be model unbiased and its model variance is obtained. Properties of the linearization, balanced repeated replication and jackknife variance estimators are examined. A robust linearization variance estimator, along with the BRR and jackknife variance estimators, are shown to be model consistent for the variance of the poststratified estimator. Second-order asymptotic equivalence of the jackknife variance estimator and a linearization variance estimator, under a design-based framework, is also established in the important special case of $n_h = 2$ clusters per stratum. This result generalizes a previous result of Rao and Wu (1985) to the poststratified estimator.

In Section 3.3, the results of Section 3.2 are extended to the two way poststratification case using the GREG estimator. Valliant's superpopulation model is generalized to the two-way situation and model unbiasedness of the GREG estimator is established as well as its model variance. Linearization, BRR and jackknife variance estimators are discussed and consistency of a robust linearization, the BRR and the jackknife variance estimators are established. Finally, the second order equivalence exhibited for the poststratified estimator is shown to hold for the GREG estimator also. Section 3.4 will present results of a simulation study investigating the finite sample properties, unconditional and conditional, of the linearization, robust linearization and jackknife variance estimators for the poststratified and GREG estimators.

3.2 One-Way Poststratification

Suppose we have the same stratified multi-stage design described in Chapter 2. We will assume the following superpopulation model (Valliant, 1993):

$$E_m(y_{hik}) = {}_c\mu \quad (hik) \text{ unit in the } c^{th} \text{ poststratum} \quad (3.2.1)$$

and

$$\text{cov}_m(y_{hik}, y_{h'i'k'}) = \begin{cases} {}_c\sigma_{hi}^2 & (hik) = (h'i'k'), (hik) \in {}_c s \\ {}_c\sigma_{hi}^2 {}_c\rho_{hi} & h = h', i = i', k \neq k', (hik) \in {}_c s, (h'i'k') \in {}_c s \\ {}_c c' \tau_{hi} & h = h', i = i', k \neq k', (hik) \in {}_c s, (h'i'k') \in {}_{c'} s \\ 0 & \text{otherwise} \end{cases}$$

where the subscript m denotes that the expectations and covariances are taken with respect to the model.

The model assumes that the y -values from different clusters are uncorrelated and independent of each other. Units in a poststratum have a common mean, ${}_c\mu$, and are correlated within a cluster. The size of the covariances ${}_c\sigma_{hi}^2 {}_c\rho_{hi}$ and ${}_c c' \tau_{hi}$ are allowed to vary among the clusters and depend upon whether or not the units are in the same poststratum.

3.2.1 Model Bias and Prediction Variance

The model bias of the basic expansion estimator of the total, $\hat{Y} = \sum_{(hik) \in s} w_{hik} y_{hik}$, is $E_m(\hat{Y} - Y) = \sum_c {}_c\mu({}_c\hat{M} - {}_cM)$ which will be zero if ${}_c\hat{M} = {}_cM$ for all c . This is generally not true for a particular sample, although true in expectation over all possible samples that a design can produce. On the other hand, the poststratified estimator, $\hat{Y}_{PS}$, given by (2.2.1) is easily shown to be unbiased under model (3.2.1):

$$E_m(\hat{Y}_{PS} - Y) = \sum_c \sum_{(hik) \in s} {}_c w_{hik} {}_c \delta_{hik} E_m(y_{hik}) - \sum_c {}_c \mu {}_c M$$

$$\begin{aligned}
&= \sum_c {}_c\mu {}_cM - \sum_c {}_c\mu {}_cM \\
&= 0
\end{aligned}$$

since $\sum_{(hik) \in s} {}_cw_{hik} {}_c\delta_{hik} = {}_cM$, which is the benchmarking condition satisfied by the poststratified weights ${}_cw_{hik}$; see equation (2.2.2).

For studying the asymptotic properties, working with the poststratified estimator of the mean y_{PS} is more convenient and allows us to place more meaningful conditions on the sample weights. Let $\hat{w}_{hik} = w_{hik}/M$ as before, then

$$\bar{y}_{PS} = \sum_c \frac{{}_cM}{M} \frac{{}_c\hat{Y}}{{}_c\hat{M}} = \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik} y_{hik} {}_c\delta_{hik}$$

where

$$\begin{aligned}
{}_c\hat{Y} &= \sum_{(hik) \in s} \hat{w}_{hik} y_{hik} {}_c\delta_{hik}, \\
{}_c\hat{M} &= \sum_{(hik) \in s} \hat{w}_{hik} {}_c\delta_{hik}
\end{aligned}$$

and

$$\hat{w}_{hik} = {}_cw_{hik}/M.$$

Valliant assumed the following regularity conditions.

(i) $n_T/M \rightarrow 0$ as $L \rightarrow \infty$

(ii)

$$\max_{h,s} M_{hs} = O(1) \quad \max_h N_h = O(1)$$

$$\max_{h,s} m_{hs} = O(1) \quad \max_h n_h = O(1)$$

(iii) $n_T \max_{h,s,k} m_{hs} \hat{w}_{hik} = O(1)$

(iv) ${}_c\hat{R} = {}_cM/{}_c\hat{M} \rightarrow {}_cR$ a constant not necessarily equal to 1 for all c

(v) $n_T V_m(\bar{y}) = O(1)$ where $\bar{y} = \sum_{(h,k) \in s} \hat{w}_{hik} y_{hik}$.

Conditions (i)-(iii) place bounding conditions on sample and population sizes and the weights, and condition (iv) requires each poststratification factor, ${}_c\hat{R}$, to converge to a constant not necessarily 1. Valliant showed that under conditions (i) to (v) the prediction variance, $V_m(\bar{y}_{PS} - \bar{Y})$, is approximately equal to the model variance of $\bar{y}_{PS}$:

$$V_m(\bar{y}_{PS} - \bar{Y}) \approx V_m(\bar{y}_{PS}) = \sum_h \sum_{\nu \in s_h} \left(\sum_c {}_c\nu_{h\nu} + \sum_c \sum_{\substack{c' \\ c' \neq c}} {}_cL_{h\nu} {}_{c'}L_{h\nu} {}_{cc'}\tau_{h\nu} \right)$$

where the subscript m indicates the variance is with respect to the model,

$${}_cL_{h\nu} = \sum_{k \in s_{h\nu}} {}_c\hat{w}_{hik} {}_c\delta_{hik}$$

and

$${}_c\nu_{h\nu} = {}_c\sigma_{h\nu}^2 \left(\sum_k {}_c\hat{u}_{hik}^2 + \sum_k \sum_{\substack{k' \\ k' \neq k}} {}_c\hat{w}_{hik} {}_c\hat{w}_{hik'} \right).$$

Valliant argues that since ${}_c\hat{R}$ is fixed under the model, $V_m(\bar{y}_{PS})$ is a variance conditional on the values of ${}_c\hat{R}$ or equivalently ${}_c\hat{M}$.

3.2.2 Variance Estimation

For variance estimation, Valliant compared three types of variance estimators: linearization, balanced repeated replication (BRR) and jackknife variance estimators. Although Valliant treated the clusters as sampled without replacement by including an ad hoc finite population correction factor we will, as in Chapter 2, treat the

clusters as sampled with replacement. This approximation generally leads to over-estimation of the variance, but the relative bias is likely to be small if the first-stage sampling fractions are small.

Linearization Variance Estimation

Taylor linearization variance estimators for poststratified estimators have been discussed for general sample designs by Rao (1985) and others. A customary linearization variance estimator for $\bar{y}_{PS}$ is given as

$$v_L(\bar{y}_{PS}) = \sum_{h=1}^L \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (\epsilon_{hi}^* - \bar{\epsilon}_h^*)^2 = v(\epsilon_{hi}^*) \quad (3.2.2)$$

see equation (2.3.10). The operator notation $v(\epsilon_{hi}^*)$ denotes that $v_L(\bar{y}_{PS})$ depends only on the ϵ_{hi}^* 's. The variance estimator (3.2.2) is design-consistent as shown in Chapter 2 but not model consistent for $V_m(\bar{y}_{PS})$. In fact, Valliant has shown that it is model consistent for $V_m(\bar{y})$, the model variance of the basic estimator $\bar{y}$.

Rao (1985) and Valliant (1993) give an alternative linearization variance estimator using the poststratified weights ${}_c\hat{w}_{hik}$:

$$v_{L2}(\bar{y}_{PS}) = v(\tilde{\epsilon}_{hi}^*) \quad (3.2.3)$$

with

$$\tilde{\epsilon}_{hi}^* = \sum_c \sum_k {}_c\hat{w}_{hik} \epsilon_{hik}^*.$$

With the additional regularity condition

- (vi) $M^{1+\delta} \sum_h E_m |\sum_k \hat{w}_{hik} {}_c\delta_{hik} (y_{hik} - {}_c\mu)|^{2+\delta} = O(1)$ for some $\delta > 0$ and for all i and c

Valliant has shown that under conditions (i)-(vi) the variance estimator (3.2.3) is model consistent for $V_m(\bar{y}_{PS})$. It is also design consistent, noting that

$${}_cM / {}_c\dot{M} \longrightarrow_p 1.$$

BRR Variance Estimation

Balanced repeated replication (BRR) or balanced half-sample variance estimators, proposed by McCarthy (1969), are often used in complex sample surveys. Suppose we have $n_h = 2$ sampled clusters in each stratum. Let $\bar{y}_{PS}^{(r)}$ be the poststratified estimator based on the r^{th} half-sample. A set of R half-samples is defined by the indicator variables

$$\delta_{hi}^{(r)} = \begin{cases} 1 & \text{if } (hi)^{th} \text{ cluster is in the } r^{th} \text{ half-sample,} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2$ and $r = 1, \dots, R$. Let

$$\gamma_h^{(r)} = 2\delta_{h1}^{(r)} - 1 = \begin{cases} +1 & \text{if the } 1^{st} \text{ cluster is in the } r^{th} \text{ half-sample,} \\ -1 & \text{if the } 2^{nd} \text{ cluster is in the } r^{th} \text{ half-sample.} \end{cases}$$

A set of R half-samples is orthogonally balanced if

$$\sum_{r=1}^R \gamma_h^{(r)} = \sum_{r=1}^R \gamma_h^{(r)} \gamma_{h'}^{(r)} = 0 \quad (h \neq h')$$

with a minimal set of half-samples having $L + 1 \leq R \leq L + 4$ where L is the number of strata. A set of R orthogonally balanced half-samples can be constructed from Hadamard matrices by choosing any L columns, excluding the columns of $+1$'s. To construct a BRR variance estimator, let

$${}_c p_{hi} = \sum_k n_h \hat{w}_{hik} y_{hik} {}_c \delta_{hik}$$

and

$${}_c t_{hi} = \sum_k n_h \hat{w}_{hik} {}_c \delta_{hik}.$$

We can then rewrite ${}_c \hat{Y}$ and ${}_c \tilde{M}$ as

$${}_c \hat{Y} = \sum_{(hi)} {}_c p_{hi} / n_h,$$

and

$${}_c\hat{M} = \sum_{(h)} {}_c t_{h1} / n_h.$$

Defining $\Delta {}_c p_h = \frac{1}{2}({}_c p_{h1} - {}_c p_{h2})$ and $\Delta {}_c t_h = \frac{1}{2}({}_c t_{h1} - {}_c t_{h2})$, the r^{th} half-sample analogues of ${}_c\hat{Y}$ and ${}_c\hat{M}$ may be written as

$$\begin{aligned} {}_c\hat{Y}^{(r)} &= {}_c\hat{Y} + \sum_h \gamma_h^{(r)} \Delta {}_c p_h, \\ {}_c\hat{M}^{(r)} &= {}_c\hat{M} + \sum_h \gamma_h^{(r)} \Delta {}_c t_h. \end{aligned}$$

The poststratified estimator based on the r^{th} half-sample is then given by

$$\bar{y}_{PS}^{(r)} = \sum_c \frac{{}_c M}{{}_c \hat{M}^{(r)}} \frac{{}_c \hat{Y}^{(r)}}{{}_c \hat{M}^{(r)}},$$

and a BRR variance estimator for $\bar{y}_{PS}$ is

$$v_{BRR}(\bar{y}_{PS}) = \frac{1}{R} \sum_r (\bar{y}_{PS}^{(r)} - \bar{y}_{PS})^2. \quad (3.2.4)$$

In applying the BRR procedure, it is questioned whether or not the estimators of the normalized poststrata totals should be recomputed for each half-sample, as above, or the full sample estimate, ${}_c\hat{M}$, can be used at each replication. First we will consider the case where ${}_c\hat{M}^{(r)}$ is computed for each half-sample. Letting $f({}_c\hat{Y}^{(r)}, {}_c\hat{M}^{(r)}) = {}_c\hat{Y}^{(r)} / {}_c\hat{M}^{(r)}$, and expanding this function around the full sample estimates $({}_c\hat{Y}, {}_c\hat{M})$, Valliant (1993) obtains the following approximation:

$$\frac{{}_c\hat{Y}^{(r)}}{{}_c\hat{M}^{(r)}} \cong \frac{{}_c\hat{Y}}{{}_c\hat{M}} + \frac{1}{{}_c\hat{M}} ({}_c\hat{Y}^{(r)} - {}_c\hat{Y}) - \frac{{}_c\hat{Y}}{{}_c\hat{M}^2} ({}_c\hat{M}^{(r)} - {}_c\hat{M}),$$

where the neglected terms are of lower order. Noting that $\bar{y}_{PS}^{(r)} = \sum_c \frac{{}_c M}{{}_c \hat{M}^{(r)}} \frac{{}_c \hat{Y}^{(r)}}{{}_c \hat{M}^{(r)}}$, and $\bar{y}_{PS} = \sum_c \frac{{}_c M}{{}_c \hat{M}} \frac{{}_c \hat{Y}}{{}_c \hat{M}}$, we have

$$\bar{y}_{PS}^{(r)} - \bar{y}_{PS} = \sum_c \frac{{}_c M}{{}_c \hat{M}} \left(\frac{{}_c \hat{Y}^{(r)}}{{}_c \hat{M}^{(r)}} - \frac{{}_c \hat{Y}}{{}_c \hat{M}} \right)$$

$$\begin{aligned}
&\cong \sum_c \frac{{}_cM}{M} \frac{1}{{}_c\hat{M}} \left({}_c\hat{Y}^{(r)} - {}_c\hat{Y} - {}_c\bar{y}({}_c\hat{M}^{(r)} - {}_c\hat{M}) \right) \\
&= \sum_c \frac{{}_cM}{M} \frac{1}{{}_c\hat{M}} \sum_h \gamma_h^{(r)} (\Delta_c p_h - {}_c\bar{y} \Delta_c t_h) \\
&= \sum_c \frac{{}_cM}{M} \frac{1}{{}_c\hat{M}} \sum_h \gamma_h^{(r)} \Delta_c \epsilon_h^*
\end{aligned}$$

where $\Delta_c \epsilon_h^* = \Delta_c p_h - {}_c\bar{y} \Delta_c t_h$ and ${}_c\bar{y} = {}_c\hat{Y} / {}_c\hat{M}$.

The BRR variance estimator (3.2.4) is then approximated as

$$v_{BRR}(\bar{y}_{PS}) \cong \frac{1}{R} \sum_r \left[\sum_c \frac{{}_cM}{M} \frac{1}{{}_c\hat{M}} \sum_h \gamma_h^{(r)} \Delta_c \epsilon_h^* \right]^2.$$

Noting that

$$\sum_r \gamma_h^{(r)} \gamma_{h'}^{(r)} = \begin{cases} R & \text{if } h = h' \\ 0 & \text{otherwise} \end{cases}$$

the BRR variance estimator becomes

$$v_{BRR}(\bar{y}_{PS}) \cong \sum_h \left(\sum_c \frac{{}_cM}{M} \frac{1}{{}_c\hat{M}} \Delta_c \epsilon_h^* \right)^2.$$

Now,

$$\begin{aligned}
\Delta_c \epsilon_h^* &= \Delta_c p_h - {}_c\bar{y} \Delta_c t_h \\
&= \frac{1}{2} ({}_c p_{h1} - {}_c\bar{y} {}_c t_{h1} - ({}_c p_{h2} - {}_c\bar{y} {}_c t_{h2})) \\
&= ({}_c \epsilon_{h1}^* - {}_c \epsilon_{h2}^*)
\end{aligned}$$

since $n_h = 2$ for all strata and ${}_c \epsilon_{h1}^* = \sum_k \hat{w}_{h1k} {}_c \delta_{h1k} (y_{h1k} - {}_c\bar{y})$. Thus,

$$\begin{aligned}
v_{BRR}(\bar{y}_{PS}) &\cong \sum_h \left(\sum_c \frac{{}_cM}{M} \frac{1}{{}_c\hat{M}} ({}_c \epsilon_{h1}^* - {}_c \epsilon_{h2}^*) \right)^2 \\
&= v_{L2}(\bar{y}_{PS}).
\end{aligned} \tag{3.2.5}$$

Since $v_{L2}(\bar{y}_{PS})$ is model consistent for $V_m(\bar{y}_{PS})$, it now follows that $v_{BRR}(\bar{y}_{PS})$ is also model consistent for $V_m(\bar{y}_{PS})$.

Suppose now we use the normalized full-sample estimates of ${}_cM$ for each half-sample and denote the resulting variance estimator $v_{BRR}^*(y_{PS})$. The poststratified estimator based on the r^{th} half-sample is then given by

$$\bar{y}_{PS}^{*(r)} = \sum \frac{{}_cM}{M} \frac{{}_c\hat{Y}^{(r)}}{{}_c\hat{M}},$$

and

$$\begin{aligned}\bar{y}_{PS}^{*(r)} - \bar{y}_{PS} &= \sum_c \frac{{}_cM}{M {}_c\hat{M}} ({}_c\hat{Y}^{(r)} - {}_c\hat{Y}) \\ &= \sum_c \frac{{}_cM}{M {}_c\hat{M}} \left(\sum_h \gamma_h^{(r)} \Delta_c p_h \right).\end{aligned}$$

The incorrect BRR variance estimator is

$$\begin{aligned}v_{BRR}^*(y_{PS}) &= \frac{1}{R} \sum_r (\bar{y}_{PS}^{*(r)} - \bar{y}_{PS})^2 \\ &= \frac{1}{R} \sum_r \left(\sum_c \frac{{}_cM}{M {}_c\hat{M}} \sum_h \gamma_h^{(r)} \Delta_c p_h \right)^2 \\ &= \sum_h \left(\sum_c \sum_k {}_c\hat{w}_{h1k} y_{h1k} - \sum_c \sum_k {}_c\hat{w}_{h2k} y_{h2k} \right)^2.\end{aligned}$$

That is, $v_{BRR}^*(\bar{y}_{PS})$ is the same as $v_{BRR}(y_{PS})$ except that the errors e_{hik} have been replaced by y_{hik} . Valliant (1993) has shown that the model expectation of $v_{BRR}^*(y_{PS})$ is

$$E_m(v_{BRR}^*(y_{PS})) = V_m(\bar{y}_{PS}) + 2 \sum_h \sum_i \left(E_m(y_{hi}^*) - \frac{1}{2} \sum_j E_m(y_{hj}^*) \right)^2$$

where $y_{hi}^* = \sum_c \sum_k {}_c\hat{w}_{hik} y_{hik} {}_c\hat{\delta}_{hik}$. Therefore, using the full-sample estimate of ${}_cM$ leads to an over estimation of the model variance of y_{PS} .

Jackknife Variance Estimation

Recall that the jackknife variance estimator is given in Chapter 2 as

$$v_J(\bar{y}_{PS}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (y_{PS(gj)} - \bar{y}_{PS})^2$$

where

$$\begin{aligned} y_{PS(g_j)} &= \sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik(g_j)} y_{hik} {}_c \delta_{hik} \\ {}_c \hat{w}_{hik(g_j)} &= \frac{{}_c M}{M {}_c \hat{M}_{(g_j)}} \hat{w}_{hik(g_j)} \\ {}_c \hat{M}_{(g_j)} &= \sum_{(hik) \in s} \hat{w}_{hik(g_j)} {}_c \delta_{hik} \end{aligned}$$

with the normalized jackknife basic weights $\hat{w}_{hik(g_j)} = w_{hik(g_j)}/M$ and $w_{hik(g_j)}$ defined by (2.4.1).

When calculating the jackknife variance estimator, as with the BRR variance estimator, one could use the correct poststratified jackknife weights, ${}_c \hat{w}_{hik(g_j)}$, or simply the jackknife adjusted weights with the full sample poststratification adjustment

$${}_c \hat{w}_{hik(g_j)}^* = \hat{w}_{hik(g_j)} \frac{{}_c M}{M {}_c \hat{M}}.$$

First we will consider the case where the poststratified adjusted jackknife weights, ${}_c \hat{w}_{hik(g_j)}$, are used.

Letting $f({}_c \hat{Y}_{(g_j)}, {}_c \hat{M}_{(g_j)}) = {}_c \hat{Y}_{(g_j)}/{}_c \hat{M}_{(g_j)}$ and expanding this function around the full sample estimates $({}_c \hat{Y}, {}_c \hat{M})$, Valliant (1993) obtained the following approximation

$$\frac{{}_c \hat{Y}_{(g_j)}}{{}_c \hat{M}_{(g_j)}} \cong \frac{{}_c \hat{Y}}{{}_c \hat{M}} + \frac{1}{{}_c \hat{M}} ({}_c \hat{Y}_{(g_j)} - {}_c \hat{Y}) - \frac{{}_c \hat{Y}}{{}_c \hat{M}^2} ({}_c \hat{M}_{(g_j)} - {}_c \hat{M}).$$

Therefore,

$$\begin{aligned} y_{PS(g_j)} - y_{PS} &= \sum_c \frac{{}_c M}{M} \left(\frac{{}_c \hat{Y}_{(g_j)}}{{}_c \hat{M}_{(g_j)}} - \frac{{}_c \hat{Y}}{{}_c \hat{M}} \right) \\ &\cong \sum_c \frac{{}_c M}{M} \frac{1}{{}_c \hat{M}} \left({}_c \hat{Y}_{(g_j)} - {}_c \hat{Y} - {}_c \bar{y} ({}_c \hat{M}_{(g_j)} - {}_c \hat{M}) \right) \\ &= \frac{n_g}{n_g - 1} \sum_c \frac{{}_c M}{M} \frac{1}{{}_c \hat{M}} \left({}_c \bar{e}_g^* - {}_c e_{g_j}^* \right) \end{aligned}$$

see equation (2.4.5). Further,

$$\begin{aligned} v_J(\bar{y}_{PS}) &= \sum_g \frac{n_g}{n_g - 1} \sum_j \left(\bar{y}_{PS(g_j)} - \bar{y}_{PS} \right)^2 \\ &\cong \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{\hat{M}} \right) ({}_c \bar{\epsilon}_g^* - {}_c \bar{\epsilon}_{g_j}^*) \right]^2 \\ &= v_{L2}(\bar{y}_{PS}). \end{aligned}$$

Since $v_{L2}(\bar{y}_{PS})$ is model consistent for $V_m(\bar{y}_{PS})$, it follows that $v_J(\bar{y}_{PS})$ is also model consistent for $V_m(\bar{y}_{PS})$.

If we use the full sample poststratification adjustment instead of the correct poststratification adjustment, we get as an estimator of Y :

$$\bar{y}_{PS(g_j)}^* = \sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik(g_j)}^* y_{hik} {}_c \delta_{hik}$$

where

$${}_c \hat{w}_{hik(g_j)}^* = \frac{\hat{w}_{hik(g_j)} {}_c M}{\hat{M}}.$$

Then $\bar{y}_{PS(g_j)}^* - \bar{y}_{PS}$ becomes

$$\bar{y}_{PS(g_j)}^* - \bar{y}_{PS} = \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_c \sum_{(gjk) \in s} {}_c \hat{w}_{gjk} y_{gjk} {}_c \delta_{gjk} - \sum_c \sum_{(gjk) \in s} {}_c \hat{w}_{gjk} y_{gjk} {}_c \delta_{gjk} \right).$$

The incorrect jackknife variance estimator is then given by

$$v_J^*(\bar{y}_{PS}) = \sum_g \frac{n_g}{n_g - 1} \sum_j \left(\bar{y}_{PS(g_j)}^* - \bar{y}_{PS} \right)^2 = v(y_{h_i}^*)$$

with $y_{h_i}^* = \sum_c \sum_k {}_c \hat{w}_{hik} y_{hik} {}_c \delta_{hik}$. The model expectation of $v_J^*(y_{PS})$ is

$$E_m(v_J^*(y_{PS})) = V_m(y_{PS}) + \sum_h \frac{n_h}{n_h - 1} \sum_i \left(E_m(y_{h_i}^*) - \frac{1}{n_h} \sum_j E_m(y_{h_j}^*) \right)^2.$$

As was the case with the incorrect BRR variance estimator, this incorrect jackknife variance estimator will overestimate the true model variance of y_{PS} .

3.2.3 Second Order Analysis

Rao and Wu (1985) have shown that for the common two sample clusters per stratum design and for general smooth functions of the basic expansion estimator, the jackknife variance estimator is equivalent to the linearization variance estimator to higher orders terms, under the designed-based setup. We will show here, that for the poststratified estimator a similar higher order equivalence holds between the jackknife variance estimator and the alternative linearization variance estimator, $v_{L2}(y_{PS})$.

A second order Taylor series expansion of $1/{}_c\hat{M}_{(gj)}$ around ${}_c\hat{M}$ gives

$$\frac{1}{{}_c\hat{M}_{(gj)}} = \frac{1}{{}_c\hat{M}} - \frac{1}{{}_c\hat{M}^2}({}_c\hat{M}_{(gj)} - {}_c\hat{M}) + \frac{1}{{}_c\hat{M}^3}({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \quad (3.2.6)$$

where ${}_c\hat{x}_{gj}$ lies between ${}_c\hat{M}_{(gj)}$ and ${}_c\hat{M}$. Letting

$$\begin{aligned} {}_cl_{gj} &= \frac{1}{{}_c\hat{M}^2}({}_c\hat{M}_{(gj)} - {}_c\hat{M}) \\ {}_cq_{gj} &= \frac{1}{{}_c\hat{M}^3}({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2 \end{aligned}$$

and

$${}_cr_{gj} = \left(\frac{1}{{}_c\hat{x}_{gj}^3} - \frac{1}{{}_c\hat{M}^3} \right) ({}_c\hat{M}_{(gj)} - {}_c\hat{M})^2$$

we can rewrite (3.2.6) as

$$\frac{1}{{}_c\hat{M}_{(gj)}} = \frac{1}{{}_c\hat{M}} - {}_cl_{gj} + {}_cq_{gj} + {}_cr_{gj}.$$

We now have

$$\begin{aligned} \frac{{}_c\hat{Y}_{(gj)}}{{}_c\hat{M}_{(gj)}} - \frac{{}_c\hat{Y}}{{}_c\hat{M}} &= \frac{1}{{}_c\hat{M}_{(gj)}}({}_c\hat{e}_g^* - {}_c\hat{e}_{gj}^*) \\ &= \frac{n_g}{n_g - 1}({}_c\hat{e}_g^* - {}_c\hat{e}_{gj}^*) \left(\frac{1}{{}_c\hat{M}} - {}_cl_{gj} + {}_cq_{gj} + {}_cr_{gj} \right) \end{aligned}$$

and the jackknife variance estimator is

$$\begin{aligned} v_J(\bar{y}_{PS}) &= \sum_g \frac{n_g - 1}{n_g} \sum_j (\bar{y}_{PS(gj)} - \bar{y}_{PS})^2 \\ &= \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_c\tilde{M}}{{}_c\tilde{M}^2} \left\{ \frac{1}{{}_c\tilde{M}} ({}_c\bar{\epsilon}_g^* - {}_c\bar{\epsilon}_{gj}^*) - {}_c\bar{l}_{gj} ({}_c\bar{\epsilon}_g^* - {}_c\bar{\epsilon}_{gj}^*) + \right. \right. \\ &\quad \left. \left. {}_c\bar{q}_{gj} ({}_c\bar{\epsilon}_g^* - {}_c\bar{\epsilon}_{gj}^*) + {}_c\bar{r}_{gj} ({}_c\bar{\epsilon}_g^* - {}_c\bar{\epsilon}_{gj}^*) \right\} \right]^2 \end{aligned}$$

Noting that $n_h \geq 2$ implies $n_h \leq 2(n_h - 1)$, we have for all c ,

$$\begin{aligned} \max_{g,j} |{}_c\bar{l}_{gj}| &= \max_{g,j} \frac{1}{{}_c\tilde{M}^2} |{}_c\hat{M}_{(gj)} - {}_c\tilde{M}| \\ &= \max_{g,j} \frac{1}{{}_c\tilde{M}^2} \frac{n_g}{n_g - 1} \left| \frac{1}{n_g} \sum_{(gik) \in s} \hat{u}_{gik} \delta_{gik} - \sum_{(gjk) \in s} \hat{u}_{gjk} \delta_{gjk} \right| \\ &\leq \frac{2}{{}_c\tilde{M}^2} \max_{g,j} \left| \frac{1}{n_g} \sum_{(gik) \in s} \hat{u}_{gik} \delta_{gik} - \sum_{(gjk) \in s} \hat{u}_{gjk} \delta_{gjk} \right| \\ &\leq \frac{2}{{}_c\tilde{M}^2} \max_{g,j} \left| \frac{1}{n_g} \sum_{(gik) \in s} \hat{u}_{gik} \delta_{gik} + \sum_{(gjk) \in s} \hat{u}_{gjk} \delta_{gjk} \right| \\ &\leq \frac{2}{{}_c\tilde{M}^2} \left(\max_{g,i} \frac{1}{n_g} \sum_{(gik) \in s} \hat{u}_{gik} \delta_{gik} + \max_{g,j} \sum_{(gjk) \in s} \hat{u}_{gjk} \delta_{gjk} \right) \\ &\leq \frac{2}{{}_c\tilde{M}^2} \left(\max_{g,i} \frac{1}{n_g} \sum_{(gik) \in s} \hat{u}_{gik} + \max_{g,j} \sum_{(gjk) \in s} \hat{u}_{gjk} \right) \\ &\leq \frac{2}{{}_c\tilde{M}^2} \left(\max_{g,i,k} m_{gi} \hat{u}_{gik} + \max_{g,j,k} m_{gj} \hat{u}_{gjk} \right) \\ &= O\left(\frac{1}{n_T}\right) \end{aligned} \tag{3.2.7}$$

by C4 and the fact that $1/{}_c\tilde{M} \rightarrow 1/{}_c\lambda < \infty$ under C4; see Lemma (2.7). Noting that ${}_c\bar{q}_{gj}$ is of order $O(1/n_T)^2$ and ${}_c\bar{r}_{gj}$ is of order $o(1/n_T)^2$ by continuity of the function $1/x^3$ for $x \neq 0$, the jackknife variance estimator may be written as

$$v_J(\bar{y}_{PS}) = \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_c\tilde{M}}{{}_c\tilde{M}^2} ({}_c\bar{\epsilon}_g^* - {}_c\bar{\epsilon}_{gj}^*) \right]^2 -$$

$$\begin{aligned}
& 2 \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{cM}{cM} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right] \left[\sum_c \frac{cM}{M} c l_{gj} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right] + \\
& \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{cM}{M} c l_{gj} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right]^2 + \\
& 2 \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{cM}{cM} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right] \left[\sum_c \frac{cM}{M} c q_{gj} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right] + \\
& \text{lower order terms.}
\end{aligned}$$

We will now show that

$$\sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{cM}{M} c l_{gj} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right]^2 = O_p \left(\frac{1}{n_T} \right)^3 \quad (3.2.8)$$

and

$$\sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{cM}{cM} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right] \left[\sum_c \frac{cM}{M} c q_{gj} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right] = O_p \left(\frac{1}{n_T} \right)^3. \quad (3.2.9)$$

To show (3.2.8), we have

$$\begin{aligned}
& \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{cM}{M} c l_{gj} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right]^2 \\
& \leq \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*)^2 \right] \left[\sum_c \left(\frac{cM}{M} \right)^2 c l_{gj}^2 \right] \\
& \leq \max_{g,j,c} c l_{gj}^2 \sum_c \left(\frac{cM}{M} \right)^2 \sum_c v_L(c\bar{\epsilon}) \\
& = O_p \left(\frac{1}{n_T} \right)^3
\end{aligned}$$

since $v_L(c, \epsilon) = O_p(1/n_T)$ for all c , $cM/M < 1$ for all c and $\max_{g,j,c} |c l_{gj}| = O(1/n_T)$.

Now to show (3.2.9), we have

$$\begin{aligned}
& \sum_g \sum_j \left[\left(\frac{n_g}{n_g - 1} \right)^{1/2} \sum_c \frac{cM}{cM} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right] \left[\left(\frac{n_g}{n_g - 1} \right)^{1/2} \sum_c \frac{cM}{M} c q_{gj} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right] \\
& \leq \left[\sum_g \frac{n_g}{n_g - 1} \sum_j \left(\sum_c \frac{cM}{cM} (c\bar{\epsilon}_g^* - c\epsilon_{gj}^*) \right)^2 \right]^{1/2} \times
\end{aligned}$$

$$\begin{aligned}
& \left[\sum_g \frac{n_g}{n_g - 1} \sum_j \left(\sum_c \frac{cM}{M} {}^c q_{gj} ({}^c \epsilon_g^* - {}^c \epsilon_{gj}^*) \right)^2 \right]^{1/2} \\
& \leq [v_{L2}(\tilde{y}_{PS})]^{1/2} \left[\sum_g \frac{n_g}{n_g - 1} \sum_j \left\{ \sum_c \left(\frac{cM}{M} {}^c q_{gj} \right)^2 \right\} \left\{ \sum_c ({}^c \epsilon_g^* - {}^c \epsilon_{gj}^*)^2 \right\} \right]^{1/2} \\
& \leq [v_{L2}(\tilde{y}_{PS})]^{1/2} \left[\max_{g,j,c} {}^c q_{gj}^2 \sum_g \frac{n_g}{n_g - 1} \sum_j \left\{ \sum_c \left(\frac{cM}{M} \right)^2 \right\} \left\{ \sum_c ({}^c \epsilon_g^* - {}^c \epsilon_{gj}^*)^2 \right\} \right]^{1/2} \\
& = O_p \left(\frac{1}{n_T} \right)^{1/2} \left[O_p \left(\frac{1}{n_T} \right)^4 O_p \left(\frac{1}{n_T} \right) \right]^{1/2} \\
& = O_p \left(\frac{1}{n_T} \right)^3
\end{aligned}$$

since $v_L({}^c \epsilon) = O_p(1/n_T)$, $v_{L2}(\tilde{y}_{PS}) = O_p(1/n_T)$, $cM/M < 1$ and $\max_{g,j,c} |{}^c q_{gj}| = O(1/n_T)^2$.

Thus we have

$$\begin{aligned}
& v_J(\tilde{y}_{PS}) \\
& = \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{cM}{M} ({}^c \epsilon_g^* - {}^c \epsilon_{gj}^*) \right]^2 - \\
& \quad 2 \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{cM}{M} ({}^c \epsilon_g^* - {}^c \epsilon_{gj}^*) \right] \left[\sum_c \frac{cM}{M} {}^c l_{gj} ({}^c \epsilon_g^* - {}^c \epsilon_{gj}^*) \right] + O_p \left(\frac{1}{n_T} \right)^3 \\
& = v_{L2}^*(\tilde{y}_{PS}) - 2B + O_p \left(\frac{1}{n_T} \right)^3
\end{aligned}$$

where

$$B = \sum_g \frac{n_g}{n_g - 1} \sum_j \sum_c \sum_{c'} \frac{cM}{M} \frac{c'M}{M} \frac{1}{cM} ({}^c \epsilon_g^* - {}^c \epsilon_{gj}^*) {}^{c'} l_{gj} ({}^{c'} \epsilon_g^* - {}^{c'} \epsilon_{gj}^*).$$

Now if $n_h = 2$ for all h , $B = 0$ since

$${}^c \epsilon_g^* - {}^c \epsilon_{g1}^* = -({}^c \epsilon_g^* - {}^c \epsilon_{g2}^*)$$

and

$${}^c l_{g1} = -{}^c l_{g2}.$$

Thus we have

$$\begin{aligned} v_J(\bar{y}_{PS}) &= v_{L2}(\bar{y}_{PS}) + O_p\left(\frac{1}{n_T}\right)^3 \\ &= v_{L2}(\bar{y}_{PS}) \left[1 + O_p\left(\frac{1}{n_T}\right)^2\right] \end{aligned}$$

which is a generalization of the result of Rao and Wu (1985). This second order equivalence does not hold for the customary linearization variance estimator, $v_L(y_{PS})$, because $v_L(y_{PS}) = v_{L2}(\bar{y}_{PS}) + O_p(1/n_T)^{\frac{3}{2}}$, noting that ${}_c\hat{M}/{}_c\hat{M} = 1 + O_p(1/n_T)^{-\frac{1}{2}}$. Thus from a statistical viewpoint, the choice of $v_J(\bar{y}_{PS})$ and $v_{L2}(\bar{y}_{PS})$ should depend more on issues such as computing costs. In this case, $v_{L2}(\bar{y}_{PS})$ is more favourable as the jackknife variance estimator requires repeated computation of ${}_c\hat{M}$ and ${}_c\hat{Y}$.

3.3 Two-Way Poststratification

Suppose now that the sampled units are classified according to two poststratification variables and are placed into the cells of a two-way table. If the cell population counts are known, methods in Section 3.2 can be used by treating the cells as poststrata. In this section, we will be considering the case where the marginal population counts are known but not the cell population counts.

We will assume that we have two poststratification variables with A and B levels respectively, and that the following superpopulation model holds:

$$E(y_{hik}) = \mu_a + \mu_b \quad (hik) \in s_{ab} \quad (3.3.1)$$

where s_{ab} is the sample falling in cell (a, b) , and

$$\text{cov}(y_{hik}, y_{h'i'k'}) = \begin{cases} \sigma_{h'iab}^2 & (hik) = (h'i'k'), (hik) \in s_{ab}, \\ \sigma_{h'iab}^2 \rho_{h'iab} & h = h', i = i', k \neq k', (hik) \in s_{ab}, (h'i'k') \in s_{ab}, \\ \tau_{h'ia}^{bb'} & h = h', i = i', k \neq k', (hik) \in s_{ab}, (h'i'k') \in s_{ab'}, \\ \tau_{h'ib}^{aa'} & h = h', i = i', k \neq k', (hik) \in s_{ab}, (h'i'k') \in s_{a'b}, \\ \tau_{hi} & h = h', i = i', k \neq k', (hik) \in s_{ab}, (h'i'k') \in s_{a'b'}, \\ 0 & \text{otherwise.} \end{cases}$$

Under this set up, we have an additive model and units in different clusters are uncorrelated and independent of each other. Units within a cluster are correlated, with the correlation depending on whether or not the units are in the same row and/or column of the two-way table.

The generalized regression estimator, $\hat{Y}_{GREG}$, can be used in the case where there is more than one poststratification variable (Huang and Fuller, 1978; Deville and Särndal, 1992). Let $\mathbf{x}_{hik}$ be a vector of indicator auxiliary variables corresponding to the categories of the poststratification variables and with known population counts $\mathbf{X}$. The generalized regression estimator of the population total, Y , is given by

$$\hat{Y}_{GREG} = \hat{Y} + \hat{\beta}^T (\mathbf{X} - \hat{\mathbf{X}}) \quad (3.3.2)$$

where $\hat{Y}$ is the basic expansion estimator,

$$\hat{\mathbf{X}} = \sum_{(hik) \in s} w_{hik} \mathbf{x}_{hik},$$

and $\hat{\beta}$ is the vector of estimated regression coefficients

$$\begin{aligned} \hat{\beta} &= \left[\sum_{(hik) \in s} w_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T \right]^{-1} \left[\sum_{(hik) \in s} w_{hik} \mathbf{x}_{hik} y_{hik} \right] \\ &= \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}} \quad \text{say.} \end{aligned}$$

The poststratified estimator, $\hat{Y}_{PS}$, is a special case of (3.3.2) where the $\mathbf{x}_{hik}$ denote the vector of indicator variables for the poststrata. In this case, $\hat{\mathbf{X}} = (\hat{M}_1, \dots, \hat{M}_c)^T$, $\mathbf{X} = (M_1, \dots, M_c)^T$, and $\hat{\beta} = (\hat{y}_1, \dots, \hat{y}_c)^T$, with $\hat{y}_c = \hat{Y}/\hat{M}_c$. Then

$$\hat{Y}_{GREG} = \hat{Y} + \sum_c \hat{y}_c (\hat{M}_c - M_c) = \hat{Y}_{PS}.$$

In the special case of $\mathbf{x}_{hik}$ being a vector of indicator variables, the generalized regression estimator reduces to the projection estimator, see Särndal (1989), given as

$$\hat{Y}_{GREG} = \hat{\beta}^T \mathbf{X}. \quad (3.3.3)$$

Thus we will work with this simpler form of the regression estimator and not (3.3.2). The generalized regression estimator, (3.3.3), can be rewritten as

$$\hat{Y}_{GREG} = \sum_{(hik) \in s} w_{hik}^* y_{hik}$$

where $w_{hik}^* = w_{hik} \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \mathbf{X}$. In the special case of $\hat{Y}_{PS}$,

$$w_{hik}^* = w_{hik} \frac{\hat{M}_c}{\hat{M}_c} = w_{hik}.$$

The GREG estimator also possesses the benchmarking property of the poststratified estimator in the sense that if the regression weights, w_{hik}^* , are used to estimate the total of the auxiliary variables, $\hat{\mathbf{X}}_{GREG}$ will be exactly $\mathbf{X}$.

3.3.1 Model Bias and Prediction Variance

For studying the properties of the GREG estimator, we will work with the GREG estimator of the population mean, $\bar{Y}$. Let $\hat{w}_{hik}^* = w_{hik}^*/M$, then

$$\begin{aligned} y_{GREG} &= \sum_{(hik) \in s} \hat{w}_{hik}^* y_{hik} \\ &= \hat{\beta}^T \hat{\mathbf{X}}. \end{aligned}$$

Model unbiasedness of $\bar{y}_{GREG}$ is shown in the following Theorem.

Theorem 3.1 *Under the model (3.3.1), $\bar{y}_{GREG}$ is an unbiased estimator of the population mean $\bar{Y}$.*

Proof

Let

$$\delta_{hik}^{ab} = \begin{cases} 1 & \text{if } (hik) \in s_{ab} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\bar{y}_{GREG} = \sum_a \sum_b \sum_{(hik) \in s} \hat{w}_{hik}^* y_{hik} \delta_{hik}^{ab},$$

and

$$\begin{aligned} \bar{y}_{GREG} - Y &= \sum_a \sum_b \sum_{(hik) \in s} (\hat{w}_{hik}^* - M^{-1}) y_{hik} \delta_{hik}^{ab} \\ &= \sum_a \sum_b \sum_h \sum_{i \in s_h} \sum_{k \notin s_h} y_{hik} \delta_{hik}^{ab} / M \\ &= \sum_a \sum_b \sum_h \sum_{i \notin s_h} \sum_{k=1}^{M_{hi}} y_{hik} \delta_{hik}^{ab} / M. \end{aligned}$$

Further,

$$\begin{aligned} E_m(\bar{y}_{GREG} - Y) &= \sum_a \sum_b \mu_a \left(\sum_{(hik) \in s} \hat{w}_{hik}^* \delta_{hik}^{ab} - \sum_h \sum_i \sum_k \delta_{hik}^{ab} / M \right) + \\ &\quad \sum_a \sum_b \mu_b \left(\sum_{(hik) \in s} \hat{w}_{hik}^* \delta_{hik}^{ab} - \sum_h \sum_i \sum_k \delta_{hik}^{ab} / M \right) \\ &= S_a + S_b, \quad \text{say.} \end{aligned}$$

We note that,

$$\begin{aligned} S_a &= \sum_a \mu_a \left(\sum_b \sum_{(hik) \in s} \hat{w}_{hik}^* \delta_{hik}^{ab} - \sum_b \sum_h \sum_i \sum_k \delta_{hik}^{ab} / M \right) \\ &= \sum_a \frac{\mu_a}{M} (X_a - X_a) \\ &= 0 \end{aligned}$$

by the benchmarking property of the regression weights $\hat{w}_{hik}^*$. Similarly, $S_b = 0$ by symmetry. Thus, y_{GREG} is model unbiased for $\bar{Y}$. ■

Rewriting y_{GREG} as

$$\begin{aligned}\bar{y}_{GREG} &= \sum_a \sum_b \sum_{(hik) \in s} \hat{w}_{hik}^* y_{hik} \delta_{hik}^{ab} \\ &= \sum_a \sum_b \sum_{(hi) \in s} t_{hi}^{ab}\end{aligned}$$

where $t_{hi}^{ab} = \sum_k \hat{w}_{hik}^* y_{hik} \delta_{hik}^{ab}$, we have under model (3.3.1)

$$\text{cov}(t_{hi}^{ab}, t_{h'i'}^{a'b'}) = \begin{cases} \nu_{h1ab} & h = h', i = i', a = a', b = b' \\ \nu_{h1abb'} & h = h', i = i', a = a', b \neq b' \\ \nu_{h1baa'} & h = h', i = i', a \neq a', b = b' \\ \nu_{h1aa'bb'} & h = h', i = i', a \neq a', b \neq b' \\ c & \text{otherwise} \end{cases} \quad (3.3.4)$$

where

$$\begin{aligned}\nu_{h1ab} &= \sigma_{h1ab}^2 \left(\sum_k \hat{w}_{hik}^{*2} \delta_{hik}^{ab} + \sum_k \sum_{k', k' \neq k} \hat{w}_{hik}^* \hat{w}_{hik'}^* \rho_{h1ab} \delta_{hik}^{ab} \delta_{hik'}^{ab} \right), \\ \nu_{h1abb'} &= \tau_{h1a}^{bb'} \sum_k \sum_{k', k' \neq k} \hat{w}_{hik}^* \hat{w}_{hik'}^* \delta_{hik}^{ab} \delta_{hik'}^{ab'}, \\ \nu_{h1baa'} &= \tau_{h1b}^{a'a'} \sum_k \sum_{k', k' \neq k} \hat{w}_{hik}^* \hat{w}_{hik'}^* \delta_{hik}^{ab} \delta_{hik'}^{a'b'}, \text{ and} \\ \nu_{h1aa'bb'} &= \tau_{hi} \sum_k \sum_{k', k' \neq k} \hat{w}_{hik}^* \hat{w}_{hik'}^* \delta_{hik}^{ab} \delta_{hik'}^{a'b'}.\end{aligned}$$

To account for the two-way poststratification, we need the additional condition

- (vii) $g_{hik} = \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \mathbf{X} \rightarrow \gamma_{hik} = \mathbf{x}_{hik}^T \mathbf{A}^{-1} \bar{\mathbf{X}}$ a constant, where $\hat{\mathbf{A}}$ is obtained from $\hat{\mathbf{A}}$ by replacing w_{hik} with $\hat{w}_{hik}$. That is, $\hat{\mathbf{A}} \rightarrow \mathbf{A}$, a matrix of constants.

This condition is similar to condition (iv) in the case of the one-way poststratified estimator.

Theorem 3.2 Under conditions (i), (ii), (iii), (v) and (vi) the prediction variance of $\bar{y}_{GREG}$ is given by

$$\begin{aligned} V_m(\bar{y}_{GREG} - \bar{Y}) &\approx V_m(\bar{y}_{GREG}) \\ &= \sum_h \sum_{u \in s_h} \sum_a \sum_b \nu_{h1ab} + \sum_h \sum_{u \in s_h} \sum_a \sum_b \sum_{b' \neq b} \nu_{h1abb'} + \\ &\quad \sum_h \sum_{u \in s_h} \sum_a \sum_{a' \neq a} \sum_b \nu_{h1baa'} + \sum_h \sum_{u \in s_h} \sum_a \sum_{a' \neq a} \sum_b \sum_{b' \neq b} \nu_{h1aa'bb'}. \end{aligned}$$

Proof

Recall

$$\begin{aligned} \bar{y}_{GREG} - \bar{Y} &= \sum_a \sum_b \sum_{(h1k) \in s} (\tilde{w}_{h1k}^* - 1/M) y_{h1k} \delta_{h1k}^{ab} - \\ &\quad \sum_a \sum_b \sum_h \sum_{u \in s_h} \sum_{k \notin s_{h1}} y_{h1k} \delta_{h1k}^{ab} / M - \\ &\quad \sum_a \sum_b \sum_h \sum_{u \notin s_h} \sum_{k=1}^{M_{h1}} y_{h1k} \delta_{h1k}^{ab} / M \\ &= T_1 - T_2 - T_3. \end{aligned}$$

Thus,

$$V_m(\bar{y}_{GREG} - \bar{Y}) = V_m(T_1) + V_m(T_2) + V_m(T_3) - 2Cov_m(T_1 T_2)$$

since T_2 and T_3 and T_1 and T_3 are summed over different clusters and are independent. First, consider $V_m(T_1)$. Using (3.3.1), we have

$$\begin{aligned} V_m(T_1) &= \sum_h \sum_{u \in s_h} V_m \left(\sum_a \sum_b \sum_{k \in s_{h1}} (\tilde{w}_{h1k}^* - 1/M) y_{h1k} \delta_{h1k}^{ab} \right) \\ &= \sum_h \sum_{u \in s_h} \sum_a T_{11h1a} + \sum_h \sum_{u \in s_h} \sum_a \sum_{a', a' \neq a} T_{12h1a'} \end{aligned}$$

where

$$T_{11h1a} = V_m \left(\sum_b \sum_{k \in s_{h1}} (\tilde{w}_{h1k}^* - 1/M) y_{h1k} \delta_{h1k}^{ab} \right)$$

$$\begin{aligned}
&= \sum_b V_m(t_{h1}^{ab}) + \sum_b \sum_{b', b' \neq b} Cov_m(t_{h1}^{ab}, t_{h1}^{ab'}) \\
&= \sum_b \nu_{h1ab} + \sum_b \sum_{b', b' \neq b} \nu_{h1abb'}
\end{aligned} \tag{3.3.5}$$

and

$$\begin{aligned}
T_{12h1a'} &= \sum_b V_m \left(\sum_{k \in s_{h1}} (\hat{u}_{h1k}^* - 1/M) y_{h1k} \delta_{h1k}^{ab} \right) + \\
&\quad \sum_b \sum_{b', b' \neq b} Cov_m \left(\sum_{k \in s_{h1}} (\hat{u}_{h1k}^* - 1/M) y_{h1k} \delta_{h1k}^{ab}, \sum_{k \in s_{h1}} (\hat{u}_{h1k}^* - 1/M) y_{h1k} \delta_{h1k}^{ab'} \right) \\
&= \sum_b \nu_{h1a'a'b} + \sum_b \sum_{b', b' \neq b} \nu_{h1a'a'bb'}.
\end{aligned} \tag{3.3.6}$$

Combining equations (3.3.5) and (3.3.6), we get

$$\begin{aligned}
V_m(T_1) &= \sum_h \sum_{i \in s_h} \sum_a \sum_b \nu_{h1ab} + \sum_h \sum_{i \in s_h} \sum_a \sum_b \sum_{b' \neq b} \nu_{h1abb'} + \\
&\quad \sum_h \sum_{i \in s_h} \sum_a \sum_{a' \neq a} \sum_b \nu_{h1baa'} + \sum_h \sum_{i \in s_h} \sum_a \sum_{a' \neq a} \sum_b \sum_{b' \neq b} \nu_{h1a'a'bb'} \\
&= V_m(\dot{y}_{GREG}).
\end{aligned}$$

From conditions (v) and (vii), we have $V_m(T_1) = O(1/n_T)$.

In a similar fashion, one can show that

$$\begin{aligned}
V_m(T_2) &= \frac{1}{M^2} \sum_h \sum_{i \in s_h} \sum_a \sum_b (M_{h1ab} - m_{h1ab}) \left[\sigma_{h1ab}^2 (1 + (M_{h1ab} - m_{h1ab} - 1) \rho_{h1ab}) + \right. \\
&\quad \left. 2 \sum_{b', b' \neq b} (M_{h1ab'} - m_{h1ab'} - 1) \tau_{h1a}^{bb'} + \sum_{a', a' \neq a} (M_{h1a'b} - m_{h1a'b} - 1) \tau_{h1b}^{aa'} + \right. \\
&\quad \left. \sum_{a', a' \neq a} \sum_{b', b' \neq b} (M_{h1a'b'} - m_{h1a'b'} - 1) \tau_{h1} \right] \\
&= O(n_T/M^2) \text{ from (ii),} \\
V_m(T_3) &= \frac{1}{M^2} \sum_h \sum_{i \in s_h} \sum_a \sum_b M_{h1ab} \left[\sigma_{h1ab}^2 (1 + (M_{h1ab} - 1) \rho_{h1ab}) + \right. \\
&\quad \left. 2 \sum_{b', b' \neq b} (M_{h1ab'} - 1) \tau_{h1a}^{bb'} + \sum_{a', a' \neq a} (M_{h1a'b} - 1) \tau_{h1b}^{aa'} + \right.
\end{aligned}$$

$$\begin{aligned}
& \left. \sum_{a', a' \neq a} \sum_{b', b' \neq b} (M_{h_{1a'b'}} - 1) \tau_{h_{1a'b'}} \right] \\
& = O((N - n_T)/M^2) \text{ from (ii) and} \\
\text{Cov}_m(T_1, T_2) &= \frac{1}{M} \sum_a \sum_b \sum_{(h_{1k}) \in s} (\hat{w}_{h_{1k}}^* - 1/M) \left[(M_{h_{1ab}} - m_{h_{1ab}} - 1) \sigma_{h_{1ab}}^2 \rho_{h_{1ab}} + \right. \\
& \quad \sum_{a', a' \neq a} (M_{h_{1a'b}} - m_{h_{1a'b}} - 1) \tau_{h_{1a'b}}^{aa'} + \\
& \quad \sum_{b', b' \neq b} (M_{h_{1ab'}} - m_{h_{1ab'}} - 1) \tau_{h_{1ab'}}^{bb'} + \\
& \quad \left. \sum_{a', a' \neq a} \sum_{b', b' \neq b} (M_{h_{1a'b'}} - m_{h_{1a'b'}} - 1) \tau_{h_{1a'b'}} \right] \\
& = O(1/M) \text{ from (ii), (iii) and (vii).}
\end{aligned}$$

The dominant term of the prediction variance is $V_m(T_1)$, establishing the result. ■

3.3.2 Variance Estimation

In this subsection, we will consider the properties of the linearization, the BRR and the jackknife variance estimators for the generalized regression estimator of the mean $\bar{Y}$.

Linearization Variance Estimator

A customary linearization variance estimator is given by

$$v_L(y_{GREG}) = v(\epsilon_{ht})$$

where $\epsilon_{ht} = \sum_k \hat{w}_{h_{1k}} \epsilon_{h_{1k}}^*$ and $\epsilon_{h_{1k}}^* = y_{h_{1k}} - \hat{\beta}^T \mathbf{x}_{h_{1k}}$. Särndal et al (1989) proposed an alternative variance estimator under a unistage sampling design. Following that method, we have for stratified multi-stage designs

$$v_{L2}(y_{GREG}) = v(\epsilon_{ht}^*) \quad (3.3.7)$$

where $e_{hi}^* = \sum_k \hat{w}_{hik} e_{hik}^*$. In the special case of $\hat{y}_{PS}$, the alternative variance estimator (3.3.7) reduces to $v_{L2}(y_{PS})$.

BRR Variance Estimator

Again, we assume that $n_h = 2$ for all strata. Let $\hat{y}_{GREG}^{(r)}$ be the generalized regression estimator based on the r^{th} half-sample.

$$\hat{y}_{GREG}^{(r)} = \hat{\beta}^{(r)T} \mathbf{X}.$$

To define $\hat{\beta}^{(r)}$ first let

$$\begin{aligned} s_{hi} &= \sum_k n_h \hat{w}_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T, \\ t_{hi} &= \sum_k n_h \hat{w}_{hik} \mathbf{x}_{hik} y_{hik}, \\ \Delta s_h &= \frac{1}{2}(s_{h1} - s_{h2}) \end{aligned}$$

and

$$\Delta t_h = \frac{1}{2}(t_{h1} - t_{h2}).$$

Also, let $\delta_{hi}^{(r)}$ and $\gamma_h^{(r)}$ be as defined previously. Then

$$\begin{aligned} \hat{\mathbf{A}}^{(r)} &= \sum_h \sum_i \delta_{hi}^{(r)} s_{hi} \\ &= \hat{\mathbf{A}} + \sum_h \gamma_h^{(r)} \Delta s_h, \\ \hat{\mathbf{b}}^{(r)} &= \sum_h \sum_i \delta_{hi}^{(r)} t_{hi} \\ &= \hat{\mathbf{b}} + \sum_h \gamma_h^{(r)} \Delta t_h \end{aligned}$$

and

$$\hat{\beta}^{(r)} = \hat{\mathbf{A}}^{(r)-1} \hat{\mathbf{b}}^{(r)}$$

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{b}}$ are obtained from $\hat{\mathbf{A}}$ and $\hat{\mathbf{b}}$ by replacing w_{hik} with $\hat{w}_{hik}$. A BRR variance estimator for y_{GREG} is given by

$$v_{BRR}(y_{GREG}) = \frac{1}{R} \sum_r (\hat{y}_{GREG}^{(r)} - y_{GREG})^2 \quad (3.3.8)$$

where R is the number of half-samples.

We now show that the BRR variance estimator is approximately equal to the alternative linearization variance estimator (3.3.7). Consider

$$\begin{aligned} \hat{\mathbf{A}}^{(r)} &= \hat{\mathbf{A}} + \sum_h \gamma_h^{(r)} \Delta s_h \\ &= \hat{\mathbf{A}} (\mathbf{I} + \mathbf{A}^{-1} \mathbf{D}^{(r)}) \end{aligned}$$

where $\mathbf{D}^{(r)} = \sum_h \gamma_h^{(r)} \Delta s_h$. Using the matrix equality

$$(\mathbf{I} + \mathbf{PQ})^{-1} = \mathbf{I} - \mathbf{P}(\mathbf{I} + \mathbf{QP})^{-1} \mathbf{Q} \quad (3.3.9)$$

we can show

$$\hat{\beta}^{(r)} \cong \hat{\beta} + \hat{\mathbf{A}}^{-1} (\mathbf{c}^{(r)} - \mathbf{D}^{(r)} \hat{\beta}) \quad (3.3.10)$$

where $\mathbf{c}^{(r)} = \sum_h \gamma_h^{(r)} \Delta t_h$. Note that $\mathbf{c}^{(r)} - \mathbf{D}^{(r)} \hat{\beta}$ can be rewritten as

$$\mathbf{c}^{(r)} - \mathbf{D}^{(r)} \hat{\beta} = \sum_h \gamma_h^{(r)} (\Delta t_h - \Delta s_h \hat{\beta}).$$

We have,

$$\begin{aligned} \hat{y}_{GREG}^{(r)} - y_{GREG} &= (\hat{\beta}^{(r)} - \hat{\beta})^T \mathbf{X} \\ &\cong \sum_h \gamma_h^{(r)} (\Delta t_h - \Delta s_h \hat{\beta})^T \hat{\mathbf{A}}^{-1} \mathbf{X} \end{aligned} \quad (3.3.11)$$

using (3.3.10). Now for $i = 1$, we have

$$\begin{aligned} (t_{h1} - s_{h1} \hat{\beta})^T \hat{\mathbf{A}}^{-1} \mathbf{X} &= (\hat{w}_{h1k} \mathbf{x}_{h1k} y_{h1k} - \hat{w}_{h1k} \mathbf{x}_{h1k} \mathbf{x}_{h1k}^T \hat{\beta})^T \hat{\mathbf{A}}^{-1} \mathbf{X} \\ &= n_h \hat{w}_{h1k} \epsilon_{h1k}^* \mathbf{x}_{h1k}^T \hat{\mathbf{A}}^{-1} \mathbf{X} \\ &= n_h \hat{\epsilon}_{h1}^* \end{aligned}$$

where $\tilde{\epsilon}_{h1}^* = \sum_k \hat{w}_{hik}^* \epsilon_{hik}^*$. Therefore,

$$\begin{aligned} (\Delta t_h - \Delta s_h \hat{\beta})^T \hat{\mathbf{A}}^{-1} \mathbf{X} &= \frac{n_h}{2} (\tilde{\epsilon}_{h1}^* - \tilde{\epsilon}_{h2}^*) \\ &= \Delta \tilde{\epsilon}_h^* \end{aligned}$$

since $n_h = 2$ for all strata. This gives

$$y_{GREG}^{(\tau)} - \bar{y}_{GREG} \cong \sum_h \gamma_h^{(\tau)} \Delta \tilde{\epsilon}_h^*.$$

Thus, equation (3.3.8) can be approximated by

$$\begin{aligned} v_{BRR}(y_{GREG}) &\cong \frac{1}{R} \sum_{\tau} \left(\sum_h \gamma_h^{(\tau)} \Delta \tilde{\epsilon}_h^* \right)^2 \\ &= \sum_h \sum_{h'} \Delta \tilde{\epsilon}_h^* \Delta \tilde{\epsilon}_{h'}^* \frac{1}{R} \sum_{\tau} \gamma_h^{(\tau)} \gamma_{h'}^{(\tau)} \\ &= \sum_h \Delta \tilde{\epsilon}_h^{*2} \end{aligned}$$

since

$$\sum_{\tau} \gamma_h^{(\tau)} \gamma_{h'}^{(\tau)} = \begin{cases} R & \text{if } h = h' \\ 0 & \text{otherwise.} \end{cases}$$

Note that with $n_h = 2$ for all h ,

$$\begin{aligned} v_{L2}(y_{GREG}) &= \sum_h \frac{n_h}{(n_h - 1)} \sum_i (\tilde{\epsilon}_{hi}^* - \bar{\tilde{\epsilon}}_h^*)^2 \\ &= \sum_h \Delta \tilde{\epsilon}_h^{*2} \end{aligned}$$

where $\bar{\tilde{\epsilon}}_h^* = (1/n_h) \sum_i \tilde{\epsilon}_{hi}^*$. That is,

$$v_{BRR}(y_{GREG}) \cong v_{L2}(y_{GREG}).$$

This result generalizes Valliant's result (3.2.5) to the GREG estimator.

Suppose we recalculate only those totals which involve y_{hik} for each half-sample.

The generalized regression estimator based on the r^{th} half-sample is then

$$y_{GREG}^{*(\tau)} = \hat{\beta}^{(\tau)T} \bar{\mathbf{X}}$$

where $\hat{\beta}^{(r)} = \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}}^{(r)}$, i.e.

$$\hat{\beta}^{(r)} = \hat{\beta} + \hat{\mathbf{A}}^{-1} \sum_h \gamma_h^{(r)} \Delta t_h.$$

We have

$$\begin{aligned} y_{GREG}^{*(r)} - y_{GREG} &= (\hat{\beta}^{(r)} - \hat{\beta})^T \mathbf{X} \\ &= \sum_h \gamma_h^{(r)} \Delta t_h^T \hat{\mathbf{A}}^{-1} \mathbf{X} \\ &= \sum_h \gamma_h^{(r)} \Delta r_h \end{aligned}$$

where $\Delta r_h = \frac{1}{2}(r_{h1} - r_{h2})$, and $r_{hi} = \sum_k u_h \hat{w}_{hik} y_{hik}$.

We will denote this incorrect *BRR* variance estimator as

$$\begin{aligned} v_{BRR}^*(y_{GREG}) &= \frac{1}{R} \sum_r (y_{GREG}^{*(r)} - y_{GREG})^2 \\ &= \sum_h \gamma_h^{(r)} \Delta r_h^* \end{aligned}$$

i.e.

$$v_{BRR}^*(y_{GREG}) = v_{L2}^*(y_{GREG}),$$

where $v_{L2}^*(y_{GREG})$ is equation (3.3.7) with $\hat{c}_h^*$ replaced by

$$y_{hi}^* = \sum_k u_h \hat{w}_{hik} y_{hik}.$$

The model expectation of this incorrect variance estimator is

$$E_m(v_{BRR}^*(y_{GREG})) = V_m(\bar{y}_{GREG}) + 2 \sum_h \sum_i \left(E_m(y_{hi}^*) - \frac{1}{2} \sum_j E_m(y_{hj}^*) \right)^2$$

where $y_{hi}^* = \sum_k \hat{w}_{hik} y_{hik}$. Therefore, this incorrect BRR variance estimator will overestimate the model variance of y_{GREG} .

Jackknife Variance Estimation

As in the case of the BRR variance estimator, we will write the generalized regression estimator as

$$\bar{y}_{GREG} = \hat{\beta}^T \bar{\mathbf{X}}.$$

To construct a jackknife variance estimator, first define the following quantities when the $(gj)^{th}$ cluster has been deleted

$$\begin{aligned}\hat{\beta}_{(gj)} &= \hat{\mathbf{A}}_{(gj)}^{-1} \hat{\mathbf{b}}_{(gj)}, \\ \hat{\mathbf{A}}_{(gj)} &= \sum_{(hik) \in s} \hat{w}_{hik(gj)} \mathbf{x}_{hik} \mathbf{x}_{hik}^T, \text{ and} \\ \hat{\mathbf{b}}_{(gj)} &= \sum_{(hik) \in s} \hat{w}_{hik(gj)} \mathbf{x}_{hik} y_{hik}\end{aligned}$$

with $\hat{w}_{hik(gj)} = w_{hik(gj)}/M$ and $w_{hik(gj)}$ defined as in equation (2.4.1). The generalized regression estimator obtained when the $(gj)^{th}$ cluster has been deleted is then given by

$$y_{GREG(gj)} = \hat{\beta}_{(gj)}^T \bar{\mathbf{X}}.$$

and the jackknife variance estimator by

$$v_J(\bar{y}_{GREG}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (y_{GREG(gj)} - \bar{y}_{GREG})^2.$$

We can write $\hat{\beta}_{(gj)}$ as following

$$\hat{\beta}_{(gj)} = (\hat{\mathbf{A}} + \mathbf{D}_{gj})^{-1} (\hat{\mathbf{b}} + \mathbf{c}_{gj})$$

where

$$\begin{aligned}\mathbf{D}_{gj} &= \frac{n_g}{n_j - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} \mathbf{x}_{gik} \mathbf{x}_{gik}^T - \sum_{(gjk) \in s} \hat{w}_{gjk} \mathbf{x}_{gjk} \mathbf{x}_{gjk}^T \right), \text{ and} \\ \mathbf{c}_{gj} &= \frac{n_g}{n_j - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} \mathbf{x}_{gik} y_{gik} - \sum_{(gjk) \in s} \hat{w}_{gjk} \mathbf{x}_{gjk} y_{gjk} \right).\end{aligned}$$

Using matrix equality (3.3.9) we can show that

$$\hat{\beta}_{(gJ)} \cong \hat{\beta} + \hat{\mathbf{A}}^{-1}(\mathbf{c}_{gJ} - \mathbf{D}_{gJ}\hat{\beta}). \quad (3.3.12)$$

Note that

$$\mathbf{c}_{gJ} - \mathbf{D}_{gJ}\hat{\beta} = \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} \mathbf{x}_{gik} \epsilon_{gik}^* - \sum_{(gjk) \in s} \hat{w}_{gjk} \mathbf{x}_{gjk} \epsilon_{gjk}^* \right) \quad (3.3.13)$$

where $\epsilon_{hik}^* = y_{hik} - \hat{\beta}^T \mathbf{x}_{hik}$.

Now the difference $\bar{y}_{GREG(gJ)} - y_{GREG}$ is simply

$$\begin{aligned} & \bar{y}_{GREG(gJ)} - y_{GREG} \\ &= (\hat{\beta}_{(gJ)} - \hat{\beta})^T \mathbf{X} \\ &\cong (\mathbf{c}_{gJ} - \mathbf{D}_{gJ}\hat{\beta})^T \hat{\mathbf{A}}^{-1} \mathbf{X} \\ &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} \epsilon_{gik}^* \mathbf{x}_{gik}^T \hat{\mathbf{A}}^{-1} \mathbf{X} - \sum_{(gjk) \in s} \hat{w}_{gjk} \epsilon_{gjk}^* \mathbf{x}_{gjk}^T \hat{\mathbf{A}}^{-1} \mathbf{X} \right) \end{aligned}$$

using equations (3.3.12) and (3.3.13).

Thus, the jackknife variance estimator is

$$v_J(\bar{y}_{GREG}) \cong \sum_g \frac{n_g}{n_g - 1} \sum_j \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik}^* \epsilon_{gik}^* - \sum_{(gjk) \in s} \hat{w}_{gjk}^* \epsilon_{gjk}^* \right)^2$$

i.e. $v_J(y_{GREG}) \cong v_{L2}(y_{GREG})$.

Suppose we jackknife only those totals which involve y_{hik} . The regression estimator then becomes

$$\bar{y}_{GREG(gJ)}^* = \hat{\beta}_{(gJ)}^T \mathbf{X}$$

where

$$\begin{aligned} \hat{\beta}_{(gJ)} &= \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}}_{(gJ)} \\ &= \hat{\beta} + \hat{\mathbf{A}}^{-1} \mathbf{c}_{gJ}. \end{aligned}$$

We have

$$\begin{aligned}
 \hat{y}_{GREG(g_j)}^* - y_{GREG} &= (\hat{\beta}_{(g_j)} - \hat{\beta})^T \mathbf{X} \\
 &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(g_1k) \in s} \hat{w}_{g_1k} y_{g_1k} \mathbf{x}_{g_1k}^T - \sum_{(g_jk) \in s} \hat{w}_{g_jk} y_{g_jk} \mathbf{x}_{g_jk}^T \right) \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}} \\
 &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(g_1k) \in s} \hat{w}_{g_1k}^* y_{g_1k} - \sum_{(g_jk) \in s} \hat{w}_{g_jk}^* y_{g_jk} \right).
 \end{aligned}$$

The jackknife variance estimator is then

$$v_J^*(y_{GREG}) = \sum_g \frac{n_g}{n_g - 1} \sum_j \left(\frac{1}{n_g} \sum_{(g_1k) \in s} \hat{w}_{g_1k}^* y_{g_1k} - \sum_{(g_jk) \in s} \hat{w}_{g_jk}^* y_{g_jk} \right)^2.$$

That is, $v_J^*(y_{GREG})$ is the same as $v_{L2}(y_{GREG})$ except the errors, e_{hik}^* have been replaced by y_{hik} . The model expectation of this incorrect variance estimator is

$$E_m(v_J^*(y_{GREG})) = V_m(y_{GREG}) + \sum_h \frac{n_h}{n_h - 1} \sum_i \left(E_m(y_{hi}^*) - \frac{1}{2} \sum_j E_m(y_{hj}^*) \right)^2$$

where $y_{hi}^* = \sum_k \hat{w}_{hik}^* y_{hik}$. This incorrect jackknife variance estimator will therefore overestimate the model variance of $\bar{y}_{GREG}$.

On a practical note, the equivalence of $v_J(y_{GREG})$ and $v_{L2}(y_{GREG})$ has been exploited to produce savings in computing time and computing costs at Statistics Canada. Many of Statistics Canada's social surveys use the GREG estimator with multiple poststratification variables. For instance, the Family Expenditure (FAMEX) survey uses the GREG estimator with poststrata based on two age groups and two groups based on the number of persons residing in a household. For more information on the FAMEX survey, see Statistics Canada Publication #62-555. FAMEX currently uses the jackknife variance method to produce variance estimates but due to the considerable computational costs, Statistics Canada is seeking an alternative variance estimation method. In order to keep continuity with

Table 3.1: FAMEX Average Expenditures for Married Couple Households

Province	$SE_J(\hat{\theta})$	$SE_{I_2}(\hat{\theta})$	RD
Newfoundland	2.73	2.66	2.6 %
Prince Edward Island	2.30	2.28	0.9 %
Nova Scotia	0	0	n/a
New Brunswick	19.55	18.97	3.0 %
Quebec	8.15	8.11	0.1 %
Ontario	0.61	0.61	0.0 %
Manitoba	5.10	5.08	0.1 %
Saskatchewan	0	0	n/a
Alberta	12.53	12.51	0.1 %
British Columbia	0.21	0.21	0.0 %

published variance estimates, the new variance estimates must be close to the jackknife variance estimates. In an internal working paper, Bechakhee (1995) made an empirical study to compare the jackknife variance estimates and estimates obtained from *GLSREG* (1).

Using data from the 1992 FAMEX survey, she calculated jackknife and linearized jackknife variance estimators for the average expenditures for married couple households at a provincial level. Table 3.1 presents standard error estimates along with relative differences defined as

$$RD = \frac{SE_{I_2}(\hat{\theta}) - SE_J(\hat{\theta})}{SE_J(\hat{\theta})} \times 100\%.$$

As we can see, the two variance estimators are extremely close with relative differences less than 3.0%. Comparing CPU times and costs, we see that the linearization variance estimator, *v*_{I₂}, used significantly less CPU time — 22 versus 45.12 for the

2

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

jackknife variance estimator and the cost significantly less: \$8.18 versus \$71.54 for the jackknife variance estimator.

3.3.3 Large Sample Properties

To investigate the large sample properties of $\bar{y}_{GREG}$ and its variance estimators, we will first consider the basic estimator of the population mean:

$$\bar{y} = \sum_{(hik)cs} \hat{w}_{hik} y_{hik}$$

obtained by replacing $g_{hik} = \mathbf{x}_{hik}^T \hat{A}^{-1} \bar{\mathbf{X}}$ in $\hat{w}_{hik}^*$ with 1. Note, that under the superpopulation model, g_{hik} is a constant. Since $\bar{y} = \bar{y}_{GREG}$ with $g_{hik} = 1$, it is clear that

$$\begin{aligned} V_m(\bar{y}) = & \sum_h \sum_{i \in s_h} \sum_a \sum_b \nu'_{hiab} + \sum_h \sum_{i \in s_h} \sum_a \sum_b \sum_{b' \neq b} \nu'_{hiabb'} + \\ & \sum_h \sum_{i \in s_h} \sum_a \sum_{a' \neq a} \sum_b \nu'_{hiaba'} + \sum_h \sum_{i \in s_h} \sum_a \sum_{a' \neq a} \sum_b \sum_{b' \neq b} \nu'_{hiaba'b'b'} \end{aligned}$$

where the prime on the ν -values indicate that $g_{hik} = 1$ for all (hik) in the expression for $V_m(\bar{y}_{GREG})$.

Now consider the linearization variance estimator

$$v_L(\bar{y}_{GREG}) = \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi}^* - \bar{e}_h^*)^2$$

with

$$\begin{aligned} \epsilon_{hi}^* &= \sum_k \hat{w}_{hik} \epsilon_{hik}^*, \\ \epsilon_{hik}^* &= y_{hik} - \hat{\beta}^T \mathbf{x}_{hik}, \quad \text{and} \\ \bar{e}_h^* &= \frac{1}{n_h} \sum_i \epsilon_{hi}^*. \end{aligned}$$

We will replace the estimated residual, e_{hik}^* , with the true residual, $e_{hik} = y_{hik} - \beta^T \mathbf{x}_{hik}$, and work with

$$v_L^*(\bar{y}_{GREG}) = \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2. \quad (3.3.14)$$

Using techniques in Chapter 2 it can be shown that

$$n_T(v_L(\bar{y}_{GREG}) - v_L^*(\bar{y}_{GREG})) \longrightarrow_p 0,$$

so that it is sufficient to establish the model consistency of $v_L^*(\bar{y}_{GREG})$. Note that here the convergence in probability refers to the probability distribution generated by the model. To prove model consistency of $v_L(\bar{y}_{GREG})$, we need the additional condition

(viii) $n_T^{1+\delta} \sum_h \sum_i E_m |\epsilon_{hi} - E_m(\epsilon_{hi})|^{2+\delta} = O(1)$ for some $\delta > 0$ where $\epsilon_{hi} = \sum_k \hat{w}_{hik} e_{hik}$ and $e_{hik} = y_{hik} - \beta^T \mathbf{x}_{hik}$ are the residuals.

Note that this condition is analogous to condition (vi) in the one-way poststratification case. The following theorem will show that $v_L^*(\bar{y}_{GREG})$ is model consistent for $V_m(\bar{y})$ but not for $V_m(\bar{y}_{GREG})$.

Theorem 3.3 *Under conditions (v) and (viii), we have*

$$n_T(v_L^*(\bar{y}_{GREG}) - V_m(\bar{y})) \longrightarrow_p 0.$$

Proof

We have

$$\begin{aligned} v_L(\bar{y}_{GREG}) &= \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 \\ &= \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - E(e_{hi}))^2 - \sum_h \frac{n_h^2}{n_h - 1} (\bar{e}_h - E(e_{hi}))^2 \\ &= v_1 - v_2. \end{aligned}$$

Now,

$$\begin{aligned} E_m(v_1) &= \sum_h \frac{n_h}{n_h - 1} \sum_i E_m(\epsilon_{hi} - E(\epsilon_{hi}))^2 \\ &= \sum_h \frac{n_h}{n_h - 1} \sum_i V_m(\epsilon_{hi}) \end{aligned}$$

with

$$\begin{aligned} V_m(\epsilon_{hi}) &= V_m \left(\sum_k \hat{w}_{hik} (y_{hik} - \beta^T \mathbf{x}_{hik}) \right) \\ &= V_m \left(\sum_a \sum_b \sum_k \hat{w}_{hik} (y_{hik} - \beta^T \mathbf{x}_{hik}) \delta_{hik}^{ab} \right) \\ &= V_m \left(\sum_a \sum_b \sum_k \hat{w}_{hik} y_{hik} \delta_{hik}^{ab} \right) \\ &= V_m(t'_{hi}) \end{aligned}$$

where

$$\begin{aligned} t'_{hi} &= \sum_a \sum_b t_{hi}^{ab}, \\ t_{hi}^{ab} &= \sum_k \hat{w}_{hik} y_{hik} \delta_{hik}^{ab}, \\ V_m(t'_{hi}) &= \sum_a \sum_b \nu'_{hiab} + \sum_a \sum_b \sum_{b' \neq b} \nu'_{hiabb'} + \\ &\quad \sum_a \sum_{a' \neq a} \sum_b \nu'_{hitbaa'} + \sum_a \sum_{a' \neq a} \sum_b \sum_{b' \neq b} \nu'_{hiaa'bb'} \end{aligned}$$

and the prime indicates that $g_{hik} = 1$ for all (hik) .

To show that

$$\begin{aligned} n_T v_1 &\longrightarrow_p n_T E_m(v_1) \\ &= n_T \sum_h \frac{n_h}{n_h - 1} \sum_i V_m(t'_{hi}), \end{aligned}$$

let

$$X_{hi} = n_T \left(\frac{n_h}{n_h - 1} \right)^{1/2} (\epsilon_{hi} - E_m(\epsilon_{hi}))$$

and note that $n_h \geq 2$ implies that $n_h/(n_h - 1) \leq 2$. Then

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E_m |X_{hi}^2|^{1+\delta/2} &= n_T^{1+\delta} \sum_h \left(\frac{n_h}{n_h - 1} \right)^{1+\delta/2} \sum_i E_m |e_{hi} - E_m(e_{hi})|^{2+\delta} \\ &\leq 2^{1+\delta/2} n_T^{1+\delta} \sum_h \sum_i E_m |e_{hi} - E_m(e_{hi})|^{2+\delta} \\ &= O(1) \end{aligned}$$

by condition (viii). Applying the Law of Large Numbers (see Lemma 2.2) gives

$$\frac{1}{n_T} \sum_h \sum_i X_{hi}^2 = n_T v_1 \rightarrow_p n_T \sum_h \frac{n_h}{n_h - 1} \sum_i V_m(t'_{hi}). \quad (3.3.15)$$

Now, consider v_2 .

$$\begin{aligned} E_m(v_2) &= \sum_h \frac{n_h^2}{n_h - 1} E_m(\bar{e}_h - E_m(e_{hi}))^2 \\ &= \sum_h \frac{n_h^2}{n_h - 1} V_m(\bar{e}_h) \\ &= \sum_h \frac{1}{n_h - 1} \sum_i V_m(t'_{hi}). \end{aligned}$$

Let

$$X_h = (Ln_T)^{1/2} \frac{n_h}{(n_h - 1)^{1/2}} (\bar{e}_h - E_m(e_{hi}))$$

and $0 < \delta'/2 < \min(\delta/2, 1/2)$. Then

$$\begin{aligned} \frac{1}{L} \sum_h E_m |X_h^2|^{1+\delta'/2} &= \left(\frac{L}{n_T} \right)^{\delta'/2} n_T^{1+\delta'} \sum_h \frac{n_h^{2+\delta'}}{(n_h - 1)^{1+\delta'/2}} E_m |\bar{e}_h - E_m(e_{hi})|^{2+\delta'} \\ &\leq \left(\frac{L}{n_T} \right)^{\delta'/2} 2^{1+\delta'/2} n_T^{1+\delta'} \sum_h n_h^{1+\delta'/2} E_m |\bar{e}_h - E_m(e_{hi})|^{2+\delta'} \\ &= O(1) \end{aligned}$$

if we can show

$$n_T^{1+\delta'} \sum_h n_h^{1+\delta'/2} E_m |\bar{e}_h - E_m(e_{hi})|^{2+\delta'} = O(1).$$

Now, by the Cauchy-Schwarz inequality and Lemma (2.4),

$$\begin{aligned} E_m |\bar{e}_h - E_m(e_{hi})|^{2+\delta'} &\leq \left(E_m(\bar{e}_h - E_m(e_{hi}))^2 \right)^{1/2} \left(E_m |\bar{e}_h - E_m(e_{hi})|^{2+2\delta'} \right)^{1/2} \\ &\leq \left(\frac{1}{n_h^2} \sum_i V_m(t'_{hi}) \right)^{1/2} \left(\frac{1}{n_h} \sum_i E_m |e_{hi} - E_m(e_{hi})|^{2+2\delta'} \right)^{1/2} \end{aligned}$$

so we have

$$\begin{aligned} n_T^{1+\delta'} \sum_h n_h^{1+\delta'/2} E_m |\bar{e}_h - E_m(e_{hi})|^{2+\delta'} &\leq n_T^{1+\delta'} \sum_h n_h^{1+\delta'/2} \left(\frac{1}{n_h^2} \sum_i V_m(t'_{hi}) \right)^{1/2} \left(\frac{1}{n_h} \sum_i E_m |e_{hi} - E_m(e_{hi})|^{2+2\delta'} \right)^{1/2} \\ &\leq n_T^{1+\delta'} \sum_h n_h \left(\frac{1}{n_h} \sum_i V_m(t'_{hi}) \right)^{1/2} \left(\frac{1}{n_h} \sum_i E_m |e_{hi} - E_m(e_{hi})|^{2+2\delta'} \right)^{1/2} \\ &\leq \left(n_T \sum_h \sum_i V_m(t'_{hi}) \right)^{1/2} \left(n_T^{1+2\delta'} \sum_h \sum_i E_m |e_{hi} - E_m(e_{hi})|^{2+2\delta'} \right)^{1/2} \\ &= O(1) \end{aligned}$$

by conditions (v) and (viii). Lemma 2.2 then implies

$$n_T v_2 \longrightarrow_p n_T \sum_h \frac{1}{n_h - 1} \sum_i V_m(t'_{hi}). \quad (3.3.16)$$

Combining equations (3.3.15) and (3.3.16), and noting that

$$\begin{aligned} E_m(v_L^*(\bar{y}_{GREG})) &= E_m(v_1) - E_m(v_2) \\ &= \sum_h \sum_i V_m(t'_{hi}) \left(\frac{n_h}{n_h - 1} - \frac{1}{n_h - 1} \right) \\ &= V_m(\bar{y}). \end{aligned}$$

we have

$$\begin{aligned} n_T v_L^*(\bar{y}_{GREG}) &= n_T v_1 - n_T v_2 \\ &\longrightarrow_p n_T V_m(\bar{y}) \end{aligned}$$

or

$$n_T(v_L(\bar{y}_{GREG}) - V_m(\bar{y})) \longrightarrow_p 0.$$

■

Now consider the variance estimator

$$v_{L2}(\bar{y}_{GREG}) = \sum_h \frac{n_h}{n_h - 1} \sum_i (\tilde{e}_{hi}^* - \tilde{e}_h^*)^2$$

where $\tilde{e}_{hi}^* = \sum_k \hat{w}_{hik}^* e_{hik}^*$ and $\tilde{e}_h^* = (1/n_h) \sum_i \tilde{e}_{hi}^*$. As was previously done, we will replace the estimated residual, $e_{hik}^* = y_{hik} - \hat{\beta}^T \mathbf{x}_{hik}$, with the true residual, $e_{hik} = y_{hik} - \beta^T \mathbf{x}_{hik}$ and consider the variance estimator

$$v_{L2}^*(\bar{y}_{GREG}) = \sum_h \frac{n_h}{n_h - 1} \sum_i (\tilde{e}_{hi} - \tilde{e}_h)^2$$

with $\tilde{e}_{hi} = \sum_k \hat{w}_{hik}^* e_{hik}$ and $\tilde{e}_h = (1/n_h) \sum_i \tilde{e}_{hi}$. Again, it is sufficient to establish the model consistency of $v_{L2}^*(\bar{y}_{GREG})$ since it can be shown that $n_T(v_{L2}(\bar{y}_{GREG}) - v_{L2}^*(\bar{y}_{GREG})) \longrightarrow_p 0$.

By adding and subtracting γ_{hik} to $\hat{w}_{hik}^*$ we get

$$\begin{aligned} \tilde{e}_{hi} &= \sum_k \hat{w}_{hik} e_{hik} (\gamma_{hik} + g_{hik} - \gamma_{hik}) \\ &= \sum_k \hat{w}_{hik} e_{hik} \gamma_{hik} + \sum_k \hat{w}_{hik} e_{hik} (g_{hik} - \gamma_{hik}) \\ &= e_{hi}^r + e_{hi}^\Delta \end{aligned}$$

where e_{hi}^r is the weighted cluster total of the residuals using the regression weight and e_{hi}^Δ is the weighted cluster total of the residuals using the design weight multiplied by the difference $g_{hik} - \gamma_{hik}$. The variance estimator $v_{L2}^*(\bar{y}_{GREG})$ then becomes

$$\begin{aligned} v_{L2}^*(\bar{y}_{GREG}) &= \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^r - \bar{e}_h^r)^2 + \\ &\quad \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^\Delta - \bar{e}_h^\Delta)^2 + \end{aligned}$$

$$\begin{aligned}
& 2 \text{ (cross-product terms)} \\
& = V_1 + V_2 + 2 \text{ (cross-product terms)} \quad (3.3.17)
\end{aligned}$$

where $\bar{e}_h^r = 1/n_h \sum_i e_{hi}^r$, $\bar{e}_h^\Delta = 1/n_h \sum_i e_{hi}^\Delta$.

$$\begin{aligned}
V_1 &= \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^r - \bar{e}_h^r)^2 \quad \text{and} \\
V_2 &= \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^\Delta - \bar{e}_h^\Delta)^2.
\end{aligned}$$

Consistency of $v_{L2}^*(\tilde{y}_{GREG})$ is established in the following theorem by showing

$$n_T V_1 \longrightarrow_p n_T V_m(\tilde{y}_{GREG})$$

$$n_T V_2 \longrightarrow_p 0 \quad \text{and}$$

$$n_T(\text{cross-product terms}) \longrightarrow_p 0.$$

Theorem 3.4 *If conditions (v) and (vii) hold along with*

$$n_T^{1+\delta} \sum_h \sum_i E_m |e_{hi}^r - E_m(e_{hi}^r)|^{2+\delta} = O(1) \quad (3.3.18)$$

and

$$n_T \sum_h \frac{n_h}{n_h - 1} \sum_i \|\hat{\mathbf{Z}}_{hi} - \hat{\mathbf{Z}}_h\|^2 = O_p(1) \quad (3.3.19)$$

where $\hat{\mathbf{Z}}_{hi} = \sum_k \hat{w}_{hik} e_{hik} \mathbf{x}_{hik}$ and $\hat{\mathbf{Z}}_h = (1/n_h) \sum_i \hat{\mathbf{Z}}_{hi}$. then

$$n_T(v_{L2}^*(\tilde{y}_{GREG}) - V_m(\tilde{y}_{GREG})) \longrightarrow_p 0.$$

Condition (3.3.19) is simply a bounding condition on the variance of the $\hat{\mathbf{Z}}_{hi}$.

Proof

By adding and subtracting $E_m(e_{hi}^r)$ to V_1 from equation (3.3.17), we have

$$\begin{aligned}
V_1 &= \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^r - E_m(e_{hi}^r))^2 - \sum_h \frac{n_h^2}{n_h - 1} (\bar{e}_h^r - E_m(\bar{e}_h^r))^2 \\
&= V_{11} - V_{12}.
\end{aligned}$$

Considering V_{11} first, let

$$X_{hi} = n_T \left(\frac{n_h}{n_h - 1} \right)^{1/2} (e_{hi}^r - E_m(e_{hi}^r))$$

and then noting that $n_h \geq 2$ implies $n_h/(n_h - 1) \leq 2$, we have

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E_m |X_{hi}^2|^{1+\delta/2} &\leq n_T^{1+\delta} 2^{1+\delta/2} \sum_h \sum_i E_m |e_{hi}^r - E_m(e_{hi}^r)|^{2+\delta} \\ &= O(1) \end{aligned}$$

by equation (3.3.18). Lemma 2.2 then gives

$$\begin{aligned} n_T V_{11} &\longrightarrow_p n_T \sum_h \frac{n_h}{n_h - 1} \sum_i E_m (e_{hi}^r - E_m(\bar{e}_h^r))^2 \\ &= \sum_h \frac{n_h}{n_h - 1} \sum_i V_m(t_{hi}) \end{aligned}$$

where

$$\begin{aligned} V_m(t_{hi}) &= V_m \left(\sum_a \sum_b t_{hi}^{ab} \right) \\ &= \sum_a \sum_b \nu_{hia b} + \sum_a \sum_b \sum_{b' \neq b} \nu_{hia b b'} + \\ &\quad \sum_a \sum_{a' \neq a} \sum_b \nu_{hia a' b} + \sum_a \sum_{a' \neq a} \sum_b \sum_{b' \neq b} \nu_{hia a' b b'}. \end{aligned}$$

For V_{12} , let

$$X_h = (Ln_T)^{1/2} \frac{n_h}{(n_h - 1)^{1/2}} (\bar{e}_h^r - E_m(e_{hi}^r))$$

and $0 < \delta'/2 < \min(\delta/2, 1/2)$. Using the same arguments as in the previous theorem, we have

$$\begin{aligned} \frac{1}{L} \sum_h E_m |X_h^2|^{1+\delta'/2} &\leq \left(n_T \sum_h \sum_i V_m(t_{hi}) \right)^{1/2} \left(n_T^{1+2\delta'} \sum_h \sum_i E_m |e_{hi}^r - E_m(e_{hi}^r)|^{2+2\delta'} \right)^{1/2} \\ &= O(1) \end{aligned}$$

under condition (v) and equation (3.3.18). Applying Lemma (2.2), we have

$$n_T V_{12} \longrightarrow_p n_T \sum_h \frac{1}{n_h - 1} \sum_i V_m(t_{hi}).$$

Combining these two results gives

$$n_T V_1 = n_T V_{11} - n_T V_{12} \longrightarrow_p n_T \sum_h \sum_i V_m(t_{hi}) = V_m(y_{GREG}).$$

It remains to show that

$$n_T V_2 \longrightarrow_p 0$$

and

$$n_T \text{ cross-product terms } \longrightarrow_p 0$$

where

$$V_2 = \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi}^\Delta - \bar{\epsilon}_h^\Delta)^2.$$

Consider the difference $g_{hik} - \gamma_{hik}$,

$$\begin{aligned} g_{hik} - \gamma_{hik} &= \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}} - \mathbf{x}_{hik}^T \mathbf{A}^{-1} \bar{\mathbf{X}} \\ &= \mathbf{x}_{hik}^T (\hat{\mathbf{A}}^{-1} \bar{\mathbf{X}} - \mathbf{A}^{-1} \bar{\mathbf{X}}) \\ &= \mathbf{x}_{hik}^T \mathbf{\Gamma} \end{aligned}$$

where $\mathbf{\Gamma} = \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}} - \mathbf{A}^{-1} \bar{\mathbf{X}}$. Note that since $g_{hik} \rightarrow \gamma_{hik}$ for all (hik) , each element of the matrix $\mathbf{\Gamma}$ converges to zero. We can now write $\epsilon_{hi}^\Delta - \bar{\epsilon}_h^\Delta$ as

$$\begin{aligned} \epsilon_{hi}^\Delta - \bar{\epsilon}_h^\Delta &= \left(\sum_k \hat{w}_{hik} e_{hik} \mathbf{x}_{hik} - \frac{1}{n_h} \sum_j \sum_k \hat{w}_{hjk} e_{hjk} \mathbf{x}_{hjk} \right)^T \mathbf{\Gamma} \\ &= (\tilde{\mathbf{Z}}_{hi} - \tilde{\mathbf{Z}}_h)^T \mathbf{\Gamma} \end{aligned}$$

and we have

$$\begin{aligned} (\epsilon_{hi}^\Delta - \bar{\epsilon}_h^\Delta)^2 &= [(\tilde{\mathbf{Z}}_{hi} - \tilde{\mathbf{Z}}_h)^T \mathbf{\Gamma}]^2 \\ &\leq \|\tilde{\mathbf{Z}}_{hi} - \tilde{\mathbf{Z}}_h\|^2 \|\mathbf{\Gamma}\|^2. \end{aligned}$$

Therefore,

$$\begin{aligned}
 n_T V_2 &= n_T \sum_h \frac{n_h}{n_h - 1} \sum_i (\epsilon_{hi}^\Delta - \bar{\epsilon}_h^\Delta)^2 \\
 &\leq n_T \sum_h \frac{n_h}{n_h - 1} \sum_i \|\hat{\mathbf{Z}}_{hi} - \hat{\mathbf{Z}}_h\|^2 \|\Gamma\|^2 \\
 &= \|\Gamma\|^2 n_T \sum_h \frac{n_h}{n_h - 1} \sum_i \|\hat{\mathbf{Z}}_{hi} - \hat{\mathbf{Z}}_h\|^2 \\
 &\longrightarrow_p 0
 \end{aligned}$$

since $\|\Gamma\|^2 \rightarrow 0$ and equation (3.3.19).

Appealing to the Schwarz inequality, the cross-product terms also converge to zero in probability. Thus we have

$$\begin{aligned}
 n_T v_{L2}^*(\bar{y}_{GREG}) &= n_T V_1 + n_T V_2 + 2n_T (\text{cross-product terms}) \\
 &\longrightarrow_p n_T V_m(\bar{y}_{GREG}).
 \end{aligned}$$

That is,

$$n_T (v_{L2}^*(\bar{y}_{GREG}) - V_m(\bar{y}_{GREG})) \longrightarrow_p 0.$$

■

Since $v_{L2}(\bar{y}_{GREG})$ is model consistent for $V_m(\bar{y}_{GREG})$, as a consequence of earlier results, $v_{BRR}(\bar{y}_{GREG})$ and $v_J(\bar{y}_{GREG})$ are also model consistent for $V_m(\bar{y}_{GREG})$.

3.3.4 Second Order Analysis

Rao and Wu (1985) have shown that for the common two sample clusters per stratum design and for general smooth functions of the basic expansion estimator, the jackknife variance estimator is equivalent to the linearization variance estimator to higher orders terms, under the designed-based setup. We showed that the equivalence also holds for the poststratified estimator in the Section 3.2.3. In this subsection, we will show that for the generalized regression estimator a similar higher

order equivalence holds between the jackknife variance estimator and the alternative linearization variance estimator, $v_{L2}(\bar{y}_{GREG})$.

The GREG estimator when the $(gj)^{th}$ cluster has been deleted is

$$\bar{y}_{GREG(gj)} = \hat{\beta}_{(gj)}^T \bar{X}. \quad (3.3.26)$$

Using the matrix equality $(I + PQ)^{-1} = I - P(I + QP)^{-1}Q$, we can approximate $\hat{\beta}_{(gj)}$. Writing $\hat{A}_{(gj)}$ as $\hat{A} + D_{gj}$ we have

$$\begin{aligned} \hat{A}_{(gj)} &= \hat{A} + D_{gj} \\ &= \hat{A}(\hat{A}^{-1}\hat{A} + \hat{A}^{-1}D_{gj}) \\ &= \hat{A}(I + \hat{A}^{-1}D_{gj}) \end{aligned}$$

and

$$\hat{A}_{(gj)}^{-1} = (I + \hat{A}^{-1}D_{gj})^{-1}\hat{A}^{-1}.$$

Using the matrix equality given above,

$$\begin{aligned} (I + \hat{A}^{-1}D_{gj})^{-1} &= I - \hat{A}^{-1}(I + D_{gj}\hat{A}^{-1})^{-1}D_{gj} \\ &= I - \hat{A}^{-1}(I + (\hat{A}_{(gj)} - \hat{A})\hat{A}^{-1})^{-1}D_{gj} \\ &= I - \hat{A}_{(gj)}^{-1}D_{gj}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \hat{A}_{(gj)}^{-1} &= (I + \hat{A}^{-1}D_{gj})^{-1}\hat{A}^{-1} \\ &= (I - \hat{A}_{(gj)}^{-1}D_{gj})\hat{A}^{-1} \\ &= \hat{A}^{-1} - \hat{A}_{(gj)}^{-1}D_{gj}\hat{A}^{-1} \\ &= \hat{A}^{-1} - (\hat{A}^{-1} - \hat{A}_{(gj)}^{-1}D_{gj}\hat{A}^{-1})D_{gj}\hat{A}^{-1} \\ &= \hat{A}^{-1} - \hat{A}^{-1}D_{gj}\hat{A}^{-1} + \hat{A}_{(gj)}^{-1}D_{gj}\hat{A}^{-1}D_{gj}\hat{A}^{-1} \end{aligned}$$

$$\begin{aligned}
&= \hat{\mathbf{A}}^{-1} - \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} + \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} - \\
&\quad \hat{\mathbf{A}}_{(gj)}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \\
&= \hat{\mathbf{A}}^{-1} - \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} + \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} + O\left(\frac{1}{n_T}\right)^3
\end{aligned}$$

since

$$\hat{\mathbf{A}} = O(1)$$

and

$$\begin{aligned}
\mathbf{D}_{gj} &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} \mathbf{x}_{gik} \mathbf{x}_{gik}^T - \sum_{(gjk) \in s} \hat{w}_{gjk} \mathbf{x}_{gjk} \mathbf{x}_{gjk}^T \right) \\
&= O\left(\frac{1}{n_T}\right)
\end{aligned}$$

under condition C4. Noting that $\hat{\beta}_{(gj)} = \hat{\mathbf{A}}_{(gj)}^{-1} \hat{\mathbf{b}}_{(gj)}$, we have

$$\begin{aligned}
\hat{\beta}_{(gj)} &= \hat{\mathbf{A}}_{(gj)}^{-1} \hat{\mathbf{b}}_{(gj)} \\
&= \left(\hat{\mathbf{A}}^{-1} - \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} + \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} + O\left(\frac{1}{n_T}\right)^3 \right) (\hat{\mathbf{b}} + \mathbf{c}_{gj}) \\
&= \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}} + \hat{\mathbf{A}}^{-1} \mathbf{c}_{gj} - \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}} - \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \mathbf{c}_{gj} + \\
&\quad \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} (\hat{\mathbf{b}} + \mathbf{c}_{gj}) + O\left(\frac{1}{n_T}\right)^3 \\
&= \hat{\beta} + \hat{\mathbf{A}}^{-1} (\mathbf{c}_{gj} - \mathbf{D}_{gj} \hat{\beta}) - \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} (\mathbf{c}_{gj} - \mathbf{D}_{gj} \hat{\beta}) + O\left(\frac{1}{n_T}\right)^3
\end{aligned}$$

since $\hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} = O(1/n_T)^2$, $\mathbf{c}_{gj} = O(1/n_T)$, and $\mathbf{D}_{gj} = O(1/n_T)$. That is,

$$\hat{\beta}_{(gj)} = \hat{\beta} + \hat{\mathbf{A}}^{-1} (\mathbf{c}_{gj} - \mathbf{D}_{gj} \hat{\beta}) - \hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1} (\mathbf{c}_{gj} - \mathbf{D}_{gj} \hat{\beta}) + O\left(\frac{1}{n_T}\right)^3.$$

This gives

$$\begin{aligned}
\tilde{\mathbf{y}}_{GREG(gj)} - \tilde{\mathbf{y}}_{GREG} \\
&= (\hat{\beta}_{(gj)} - \hat{\beta})^T \tilde{\mathbf{X}}
\end{aligned}$$

$$\begin{aligned}
&= \left[\hat{\mathbf{A}}^{-1}(\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta}) - \hat{\mathbf{A}}^{-1}\mathbf{D}_{g_j}\hat{\mathbf{A}}^{-1}(\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta}) + O\left(\frac{1}{n_T}\right)^3 \right]^T \mathbf{X} \\
&= (\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta})^T \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}} - \\
&\quad (\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta})^T (\hat{\mathbf{A}}^{-1}\mathbf{D}_{g_j}\hat{\mathbf{A}}^{-1})^T \bar{\mathbf{X}} + O\left(\frac{1}{n_T}\right)^3
\end{aligned}$$

We have shown previously, (3.3.13), that

$$\begin{aligned}
(\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta})^T \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}} &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gjk) \in s} \hat{u}_{gjk}^* \hat{\epsilon}_{gjk}^* - \sum_{(gjk) \in s} \hat{u}_{gjk}^* \hat{\epsilon}_{gjk}^* \right) \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}} \\
&= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gjk) \in s} \hat{u}_{gjk}^* \hat{\epsilon}_{gjk}^* - \sum_{(gjk) \in s} \hat{u}_{gjk}^* \hat{\epsilon}_{gjk}^* \right) \\
&= \frac{n_g}{n_g - 1} (\hat{\epsilon}_g^* - \hat{\epsilon}_{g_j}^*).
\end{aligned}$$

Thus we have

$$\begin{aligned}
&\bar{y}_{GREG(g_j)} - \bar{y}_{GREG} \\
&= \frac{n_g}{n_g - 1} (\hat{\epsilon}_g^* - \hat{\epsilon}_{g_j}^*) - (\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta})^T (\hat{\mathbf{A}}^{-1}\mathbf{D}_{g_j}\hat{\mathbf{A}}^{-1})^T \bar{\mathbf{X}} + O\left(\frac{1}{n_T}\right)^3
\end{aligned}$$

and the jackknife variance estimator is

$$\begin{aligned}
&v_J(\bar{y}_{GREG}) \\
&= \sum_g \frac{n_g - 1}{n_g} \sum_j (\bar{y}_{GREG(g_j)} - \bar{y}_{GREG})^2 \\
&= \sum_g \frac{n_g - 1}{n_g} \sum_j \left[(\hat{\epsilon}_g^* - \hat{\epsilon}_{g_j}^*) - (\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta})^T (\hat{\mathbf{A}}^{-1}\mathbf{D}_{g_j}\hat{\mathbf{A}}^{-1})^T \bar{\mathbf{X}} + O\left(\frac{1}{n_T}\right)^3 \right]^2 \\
&= \sum_g \frac{n_g}{n_g - 1} \sum_j (\hat{\epsilon}_g^* - \hat{\epsilon}_{g_j}^*)^2 + \\
&\quad \sum_g \frac{n_g - 1}{n_g} \sum_j \left[(\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta})^T (\hat{\mathbf{A}}^{-1}\mathbf{D}_{g_j}\hat{\mathbf{A}}^{-1})^T \bar{\mathbf{X}} \right]^2 - \\
&\quad 2 \sum_g \sum_j (\hat{\epsilon}_g^* - \hat{\epsilon}_{g_j}^*) (\mathbf{c}_{g_j} - \mathbf{D}_{g_j}\hat{\beta})^T (\hat{\mathbf{A}}^{-1}\mathbf{D}_{g_j}\hat{\mathbf{A}}^{-1})^T \bar{\mathbf{X}} + O_p\left(\frac{1}{n_T}\right)^3
\end{aligned}$$

$$\begin{aligned}
&= \sum_g \frac{n_g}{n_g - 1} \sum_j (\tilde{e}_g^* - \tilde{e}_{gj}^*)^2 - 2 \sum_g \sum_j (\tilde{e}_g^* - \tilde{e}_{gj}^*)(\mathbf{c}_{gj} - \mathbf{D}_{gj}\hat{\beta})^T \times \\
&\quad (\hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1})^T \bar{\mathbf{X}} + O_p\left(\frac{1}{n_T}\right)^3 \\
&= v_{L2}(y_{GREG}) + 2R + O_p\left(\frac{1}{n_T}\right)^3
\end{aligned}$$

where

$$R = \sum_g \sum_j (\tilde{e}_g^* - \tilde{e}_{gj}^*)(\mathbf{c}_{gj} - \mathbf{D}_{gj}\hat{\beta})^T (\hat{\mathbf{A}}^{-1} \mathbf{D}_{gj} \hat{\mathbf{A}}^{-1})^T \bar{\mathbf{X}}.$$

In the special case of $n_h = 2$ for all h , we have $R = 0$ since

$$\begin{aligned}
\tilde{e}_g^* - \tilde{e}_{g1}^* &= -(\tilde{e}_g^* - \tilde{e}_{g2}^*) \\
\mathbf{c}_{g1} &= -\mathbf{c}_{g2}
\end{aligned}$$

and

$$\mathbf{D}_{g1} = -\mathbf{D}_{g2}.$$

Thus we have

$$\begin{aligned}
v_J(\bar{y}_{GREG}) &= v_{L2}(\bar{y}_{GREG}) + O_p\left(\frac{1}{n_T}\right)^3 \\
&= v_{L2}(\bar{y}_{GREG}) \left(1 + O_p\left(\frac{1}{n_T}\right)^2\right).
\end{aligned}$$

This result establishes the equivalence of $v_J(\bar{y}_{GREG})$ and $v_{L2}(\bar{y}_{GREG})$ to higher order terms.

3.4 Simulation Study

We performed simulation studies to investigate the conditional and unconditional finite sample properties of the jackknife and the linearized jackknife variance estimators for the poststratified estimator and the GREG estimator. For both studies

we used a fixed finite population, considered by Valliant (1993), consisting of 10,841 persons included in the September 1988 Current Population Survey (CPS) of the United States. The variable used in the study was weekly wages for each person. For the one-way poststratified estimator we used eight poststrata defined on the basis of age, race and sex (see Valliant, 1993). For the GREG estimator, we used the variables age, with five levels, and race, with two levels.

The study population contained 2,826 geographic segments, each composed of about four neighbouring households. One hundred design strata were created with each stratum having about the same total number of households. A two-stage stratified sample design was used in which the segments were selected as clusters and persons as second-stage units. A sample of $n_h = 2$ segments was selected with replacement from each stratum with probability proportional to the number of persons in each segment. Within each sampled segment, a simple random sample of four persons was selected without replacement if the segment contained more than four persons. In segments with four or fewer persons, all persons in the segment were selected.

We selected two sets of 10,000 samples, one set for the one-way poststratification case and another set for the two-way poststratification case. In each sample we computed the basic estimator, $\hat{Y}$, the relevant poststratified estimator, $\hat{Y}_{PS}$ or $\hat{Y}_{GREG}$, and five variance estimators: the customary linearization variance estimator, v_L , the robust linearization or linearized jackknife, v_{JL} , the jackknife, v_J , and the incorrect jackknife, v_J^* .

To compare the variance estimators we computed the relative bias of the variance estimator v as

$$\text{rel. bias} = \frac{\sum_i v_i / 10,000}{\text{MSE}} - 1$$

Table 3.2: Unconditional Results for The Poststratified Estimator

Performance Measure	$v_L(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_J(\hat{Y}_{PS})$	$v_J^*(\hat{Y}_{PS})$
Relative Bias	-0.44 %	0.12 %	0.26 %	37.16 %
Error Rate	5.20 %	5.09 %	5.06 %	2.41 %
Lower Error Rate	2.41 %	2.35 %	2.33%	0.99 %
Upper Error Rate	2.79 %	2.74 %	2.73 %	1.42 %
Average Length	3.81	3.82	3.83	4.48

where v_i is the variance estimate of interest in the i^{th} sample. The mean squared error, MSE, was calculated as

$$MSE = \frac{1}{10,000} \sum_{i=1}^{10,000} (\hat{Y}_i - Y)^2$$

where $\hat{Y}_i$ is the estimator, poststratified or GREG, from the i^{th} sample and Y is the population total. Normal theory confidence intervals were calculated using each variance estimator and error rates corresponding to a nominal error rate of 5% were calculated as

$$\text{error rate} = 1 - \frac{\text{number of samples with } L_i \leq Y \leq U_i}{10,000}$$

where (L_i, U_i) is a confidence interval on Y for the i^{th} sample. Lower and upper error rates were calculated as

$$\begin{aligned} \text{lower error rate} &= \frac{\text{number of samples with } Y < L_i}{10,000} \\ \text{upper error rate} &= \frac{\text{number of samples with } U_i < Y}{10,000}. \end{aligned}$$

Table 3.2 presents unconditional results averaged over all 10,000 samples for the poststratified estimator. As indicated by the theory in Section 3.2, the jackknife and

the linearized jackknife variance estimators estimate the MSE of the poststratified estimator very well while the incorrect jackknife severely overestimates the MSE. Furthermore, v_J and v_{JL} are close, supporting the asymptotic equivalence of v_J and v_{JL} . We note that the linearization variance estimator, v_L , is also estimating the MSE of the poststratified estimator well, contrary to Valliant's (1993) claim. It is well known that in the design-based framework, see Theorem 2.6, that v_L is consistent for the variance of the poststratified estimator. Thus, it is not surprising to see v_L performing well in repeated sampling. Valliant (1993) reported an overestimation of 35% for v_L using the same dataset. In view of the asymptotic consistency of v_L , we conjecture that Valliant's calculation might be incorrect. Turning to confidence intervals, as one would expect from the relative bias values the error rates associated with v_L , v_J and v_{JL} are close to the nominal 5% while for v_J^* the error rate is considerably less than 5%. Also the average length of the confidence interval associated with v_J^* is larger than those associated with v_L , v_J and v_{JL} due to the overestimation by v_J^* .

Unconditional results for the GREG estimator appear in Table 3.3. As was the case for the poststratified estimator, v_L , v_J , and v_{JL} perform well both in terms of relative bias and error rates of the confidence intervals. The incorrect jackknife severely overestimates the variance of $\hat{Y}_{GREG}$ which is reflected in the lower than nominal error rates and the larger average length of the confidence intervals.

To investigate the conditional properties of the variance estimators for the poststratified estimator, we divided the 10,000 samples into 10 groups of 1,000 based on the bias measure

$$D_{PS} = \sum_c \left(\frac{{}_c\hat{M}}{{}_cM} - 1 \right).$$

Valliant (1993) used this bias measure arguing that if the poststrata totals, ${}_cY$, are approximately equal, then D_{PS} is proportional to the model bias of the basic estimator, $\hat{Y}$. More generally, D_{PS} is a measure of how "balanced" the sample is with respect to the distribution of poststrata counts. To form the 10 groups of 1,000

Table 3.3: Unconditional Results for The GREG Estimator

Performance Measure	$v_L(\hat{Y}_{GREG})$	$v_{JL}(\hat{Y}_{GREG})$	$v_J(\hat{Y}_{GREG})$	$v_J^*(\hat{Y}_{GREG})$
Relative Bias	-0.96 %	0.76 %	0.57 %	25.87 %
Error Rate	5.30 %	5.27 %	5.23 %	3.07 %
Lower Error Rate	2.24 %	2.21 %	2.19 %	1.08 %
Upper Error Rate	3.06 %	3.06 %	3.04 %	1.99 %
Average Length	3.94	3.95	3.95	4.44

D_{PS} was calculated for each sample and the 10,000 samples were sorted in ascending order by the value of D_{PS} . The 10,000 samples were then divided into 10 groups of 1,000 samples. For the GREG estimator the 10,000 samples were divided based on the bias measure

$$D_G = \sum_a \left(\frac{{}_a\hat{M}}{{}_aM} - 1 \right) + \sum_b \left(\frac{{}_b\hat{M}}{{}_bM} - 1 \right)$$

where a and b index the levels of the two poststratification variables.

The conditional results for both the poststratified estimator and the GREG estimator are similar to the unconditional results. The results for the poststratified estimator appear in Tables 3.4 and 3.5 and those for the GREG estimator in Tables 3.6 and 3.7. The three estimators v_L , v_J , and v_{JL} all performed well, although v_L seems to be somewhat worse in the extreme groups 1 and 10, while the incorrect jackknife variance estimator performed poorly. It is surprising to see v_L performing so well conditionally. A possible explanation is that with our particular sample design, we have $\hat{M} = \sum_{(h_1k)cs} w_{h_1k} = M$ so that

$$\sum_c {}_c\hat{M} = M.$$

Because of this, we do not see samples which are poorly “balanced” since if some poststrata estimates, ${}_c\hat{M}$, are gross overestimates, say, the remaining estimates correct for the overestimation. Thus, we see mostly well balanced samples where v_L is

Table 3.4: Conditional Relative Biases for The Poststratified Estimator

Group	$v_L(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_J(\hat{Y}_{PS})$	$v_J^*(\hat{Y}_{PS})$
1	-5.00 %	-8.05 %	-7.88 %	17.83 %
2	0.55 %	-1.18 %	-1.01 %	28.06 %
3	8.33 %	7.03 %	7.19 %	41.29 %
4	-1.10 %	-1.56 %	-1.42 %	31.82 %
5	-0.76 %	-0.69 %	-0.55 %	34.77 %
6	2.50 %	3.39 %	3.53 %	41.69 %
7	6.10 %	7.51 %	7.66 %	48.86 %
8	6.60 %	8.82 %	8.96 %	53.54 %
9	-4.46 %	-1.43 %	-1.31 %	41.11 %
10	-13.56 %	-9.17 %	-9.07 %	36.63 %

expected to perform well.

In conclusion, the three estimators v_L , v_J and v_{JL} all performed similarly both unconditionally and conditionally. The incorrect jackknife performed poorly indicating that reweighting must be done correctly after a cluster is deleted.

Table 3.5: Conditional Error Rates for The Poststratified Estimator

Group	$v_L(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_J(\hat{Y}_{PS})$	$v_J^*(\hat{Y}_{PS})$
1	5.5 %	5.9 %	5.9 %	3.4 %
2	4.6 %	4.8 %	4.8 %	2.9 %
3	3.7 %	3.8 %	3.8 %	1.9 %
5	5.7 %	5.8 %	5.8 %	2.9 %
5	4.9 %	4.8 %	4.7 %	2.6 %
6	5.1 %	5.0 %	4.8 %	2.2 %
7	5.2 %	4.8 %	4.8 %	2.1 %
8	4.5 %	4.3 %	4.3 %	1.3 %
9	5.8 %	5.4 %	5.4 %	2.4 %
10	7.0 %	6.3 %	6.3 %	2.4 %

Table 3.6: Conditional Relative Biases for The GREG Estimator

Group	$v_L(\hat{Y}_{GREG})$	$v_{JL}(\hat{Y}_{GREG})$	$v_J(\hat{Y}_{GREG})$	$v_J^*(\hat{Y}_{GREG})$
1	9.25 %	4.95 %	5.13 %	26.51 %
2	3.99 %	1.50 %	1.67 %	24.96 %
3	-3.24 %	-4.76 %	-4.59 %	17.53 %
4	-2.66 %	-3.43 %	-3.26 %	20.53 %
5	7.90 %	7.61 %	7.80 %	35.46 %
6	-3.60 %	-3.12 %	-2.94 %	23.38 %
7	-9.24 %	-8.27 %	-8.08 %	17.41 %
8	3.34 %	5.30 %	5.50 %	35.84 %
9	-3.75 %	-0.85 %	-0.62 %	30.84 %
10	-8.68 %	-4.15 %	-3.92 %	28.50 %

Table 3.7: Conditional Error Rates for The GREG Estimator

Group	$v_L(\hat{Y}_{GREG})$	$v_{JL}(\hat{Y}_{GREG})$	$v_J(\hat{Y}_{GREG})$	$v_J^*(\hat{Y}_{GREG})$
1	4.3 %	4.5 %	4.4 %	3.0 %
2	4.9 %	5.0 %	5.0 %	3.3 %
3	5.0 %	5.1 %	5.1 %	3.8 %
4	5.7 %	5.9 %	5.9 %	3.3 %
5	3.9 %	4.0 %	4.0 %	2.3 %
6	5.7 %	5.8 %	5.7 %	3.0 %
7	5.9 %	5.8 %	5.8 %	2.9 %
8	5.8 %	5.7 %	5.7 %	2.8 %
9	5.5 %	5.1 %	4.9 %	3.0 %
10	6.3 %	5.8 %	5.8 %	3.3 %

Chapter 4

Nonresponse Adjustments for Unit Nonresponse

4.1 Introduction

Multi-purpose surveys collect responses on a large number of items from each sampled unit. Unfortunately, it is rare to receive responses to all questions from every sampled unit. The results presented in previous chapters assumed that responses were available to all questions from every sampled unit, commonly referred to as the full-response case. In practice, it is more common that a survey has incomplete data, or suffers from nonresponse. Nonresponse may occur due to reasons such as non-contact, respondent refusal or failed consistency edits. Nonresponse affects survey estimates by reducing the effective sample size or by possibly biasing the estimates if the nonrespondents are systematically different than the respondents.

There are two main types of nonresponse, unit nonresponse and item nonresponse. Unit nonresponse occurs when no information is obtained from the sampled unit, while item nonresponse occurs when partial information is obtained. When handling nonresponse, it is common to divide the sample into classes called weighting classes in the case of unit nonresponse and imputation classes in the case of item nonresponse. These classes are usually formed based on characteristics known from

all selected sampling units, regardless of whether or not the unit responded.

Unit nonresponse is usually compensated through nonresponse adjustments to the weights used for estimation. When weighting classes exist, the nonresponse adjustments are calculated within each class. This chapter will address the problem of unit nonresponse when using a poststratified or regression estimator. Item nonresponse will be addressed in later chapters.

In Section 4.2, we will consider the poststratified estimator under a single weighting class and under multiple weighting classes. In the case of a single weighting class, asymptotic consistency and asymptotic normality will be established for the poststratified estimator and asymptotic consistency of the jackknife variance estimator will also be established. In practice, the use of multiple weighting classes is more common than a single weighting class. Usually the weighting classes cut across the poststrata but it is possible to use the poststrata as weighting classes themselves (Jones and Chromy, 1982). When the poststrata are used as weighting classes, the poststratified estimator reduces to the single weighting class case. In the case where the weighting classes cut across the poststrata, we will present the poststratified estimator along with a jackknife variance estimator. Finally, linearization variance estimators will be presented for both the single weighting class situation and the multiple weighting class situation.

In Section 4.3, we investigate the generalized regression estimator under a single weighting class and multiple weighting classes. We establish the asymptotic normality of the generalized regression estimator and the asymptotic consistency of the jackknife variance estimator in the single weighting class case. As a consequence of showing consistency of the jackknife variance estimator, we obtain a linearization variance estimator which retains properties of the jackknife variance estimator. In the multiple weighting class case we simply present the generalized regression estimator and a jackknife variance estimator along with the corresponding linearization variance estimator.

4.2 Poststratified Estimator

4.2.1 One Weighting Class

Suppose we have a stratified multi-stage design with the sample divided into respondents, s_r , and nonrespondents, s_m . Assuming that we have only one weighting class the nonresponse adjustment factor is

$$d = \frac{\sum_{(hik) \in s_r} w_{hik}}{\sum_{(hik) \in s_r} w_{hik}} = \frac{\hat{U}}{\hat{R}}$$

where w_{hik} is the design weight. Using this nonresponse adjustment factor an estimator of the total, Y , is

$$\hat{Y}_N = \sum_{(hik) \in s_r} d w_{hik} y_{hik}.$$

The nonresponse adjustment factor simply increases the basic weights, w_{hik} , to compensate for the nonresponding units.

Now, suppose we poststratify the respondents into C poststrata with known population counts, ${}_1M, \dots, {}_CM$. The poststratification adjustment, for the c^{th} poststratum, in the presence of nonresponse is

$${}_cb = \frac{{}_cM}{\sum_{(hik) \in s_r} d w_{hik} {}_c\delta_{hik}}$$

where ${}_c\delta_{hik}$ is the poststratum indicator variable as before. The poststratification adjustment uses the nonresponse adjusted weights, $d w_{hik}$, in estimating the poststratum counts. The poststratified estimator of the total when nonresponse is present is

$$\begin{aligned} \hat{Y}_{N,PS} &= \sum_c \sum_{(hik) \in s_r} d {}_cb w_{hik} y_{hik} {}_c\delta_{hik} \\ &= \sum_c \sum_{(hik) \in s_r} \frac{{}_cM}{\sum_{(hik) \in s_r} d w_{hik} {}_c\delta_{hik}} w_{hik} y_{hik} {}_c\delta_{hik} \\ &= \sum_c \sum_{(hik) \in s_r} {}_c w_{hik}^N y_{hik} {}_c\delta_{hik} \end{aligned}$$

where

$$\begin{aligned} {}_c w_{hik}^N &= \frac{{}_c M}{\sum_{(hik) \in s_r} w_{hik} {}_c \delta_{hik}} w_{hik} \\ &= \frac{{}_c M}{{}_c \hat{M}^r} w_{hik}, \quad \text{and} \\ {}_c \hat{M}^r &= \sum_{(hik) \in s_r} w_{hik} {}_c \delta_{hik}. \end{aligned}$$

This estimator is exactly the poststratified estimator based on the respondents only.

If we let

$$a_{hik} = \begin{cases} 1 & \text{if } (hik) \in s_r \\ 0 & \text{otherwise} \end{cases}$$

be a response indicator variable, we can write $\hat{Y}_{N,PS}$ as

$$\hat{Y}_{N,PS} = \sum_c \sum_{(hik) \in s} {}_c w_{hik}^N a_{hik} y_{hik} {}_c \delta_{hik}.$$

It can easily be shown that this estimator still benchmarks to the known population totals if it is used to estimate the poststratum counts. We note that if the poststratification adjustment is performed first and the poststratified weights are used to calculate the nonresponse adjustment, the resulting poststratified estimator will not benchmark to the known population totals. Thus, we consider only the case where the nonresponse adjustment is calculated before the poststratification adjustment.

In order to apply the results of Chapter 2, we will need the following additional assumptions. We will assume a uniform response mechanism or that $E(a_{hik}) = p > 0$, for all $(hik) \in s$, and that the indicator variables $\{a_{hik}\}$ are independent of each other and of the response variable y_{hik} . Then we can show that $\hat{Y}_{N,PS}$ is an approximately unbiased estimator of the population total,

$$\begin{aligned} E(\hat{Y}_{N,PS}) &= E \left(\sum_c \sum_{(hik) \in s} \frac{{}_c M}{\sum_{(hik) \in s} a_{hik} w_{hik} {}_c \delta_{hik}} w_{hik} a_{hik} y_{hik} {}_c \delta_{hik} \right) \\ &\approx \sum_c \frac{{}_c M p_c Y}{p_c M} \\ &= Y. \end{aligned}$$

In order to work with the estimator of the mean, let $\hat{w}_{hik} = w_{hik}/M$ and we have

$$\bar{y}_{N,PS} = \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik}^N a_{hik} y_{hik} {}_c\delta_{hik},$$

with ${}_c\hat{w}_{hik}^N = {}_cM/(M {}_c\hat{M}^r) \hat{w}_{hik}$ and ${}_c\hat{M}^r = \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_c\delta_{hik}$. Writing $\bar{Y}$ as

$$\bar{Y} = \sum_c \frac{{}_cM}{M} {}_c\bar{Y}$$

we have

$$\begin{aligned} \bar{y}_{N,PS} - \bar{Y} &= \sum_c \frac{{}_cM}{M {}_c\hat{M}^r} \left(\sum_{(hik) \in s} \hat{w}_{hik} a_{hik} (y_{hik} - {}_c\bar{Y}) {}_c\delta_{hik} \right) \\ &= \sum_c \frac{{}_cM}{M {}_c\hat{M}^r} \left(\sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_ce_{hik} {}_c\delta_{hik} \right) \\ &= \frac{1}{p} \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} e_{hik} + \sum_c \left(\frac{{}_cM}{M {}_c\hat{M}^r} - \frac{1}{p} \right) \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_ce_{hik} {}_c\delta_{hik} \\ &= \frac{1}{p} \bar{e}^r + \sum_c \left(\frac{{}_cM}{M {}_c\hat{M}^r} - \frac{1}{p} \right) {}_c\bar{e}^r \end{aligned} \quad (4.2.1)$$

where

$$\begin{aligned} {}_ce_{hik} &= y_{hik} - {}_c\bar{Y}, \\ e_{hik} &= y_{hik} - \sum_c {}_c\bar{Y} {}_c\delta_{hik}, \\ \bar{e}^r &= \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} e_{hik} \end{aligned}$$

and

$${}_c\bar{e}^r = \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_ce_{hik} {}_c\delta_{hik}.$$

Assuming the following regularity conditions,

C7 $n_T^{1+\delta} \sum_h \sum_i E|e_{hi} - E(e_{hi})|^{2+\delta} = O(1)$ where $e_{hi} = \sum_k \hat{w}_{hik} a_{hik} e_{hik}$ and $e_{hik} = y_{hik} - \sum_c {}_c\bar{Y} {}_c\delta_{hik}$ are the residuals

C8 $n_TV({}_c\bar{e}^r) = O(1)$ where ${}_c\bar{e}^r = \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_ce_{hik} {}_c\delta_{hik}$ and ${}_ce_{hik} = y_{hik} - {}_c\bar{Y}$

C9 $n_T V(\bar{\epsilon}^r) \rightarrow \sigma_{\epsilon^r}^2$ where $\bar{\epsilon}^r = \sum_{(h,k) \in s} \hat{w}_{hik} a_{hik} \epsilon_{hik}$

we have the following theorems.

Theorem 4.1 *Under C4, C7, C8 and the uniform response mechanism,*

$$\bar{y}_{N,PS} \longrightarrow_p \bar{Y}.$$

Note that it is enough to have bounded $1 + \delta$ moments in C7 for Theorem 4.1.

Proof

We will show that both terms in (4.2.1) converge to zero in probability. For the first term, let

$$X_{hi} = n_T(\epsilon_{hi} - E(\epsilon_{hi})).$$

Condition C7 and Lemma 2.5 gives

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{1+\delta} &= n_T^\delta \sum_h \sum_i E|\epsilon_{hi} - E(\epsilon_{hi})|^{1+\delta} \\ &\leq \left(n_T^{1+\delta} \sum_h \sum_i E|\epsilon_{hi} - E(\epsilon_{hi})|^{2+\delta} \right)^{\frac{1+\delta}{2+\delta}} \\ &= O(1). \end{aligned}$$

Therefore, by Lemma 2.2, we have

$$\frac{1}{n_T} \sum_h \sum_i X_{hi} = \sum_h \sum_i (\epsilon_{hi} - E(\epsilon_{hi})) \longrightarrow_p 0$$

or

$$\bar{\epsilon}^r \longrightarrow_p 0$$

since $\sum_h \sum_i E(\epsilon_{hi}) = 0$.

Now, if we let E_2 and V_2 denote expectation and variance conditional on s , we have

$$V({}_c \hat{M}^r) = V E_2({}_c \hat{M}^r) + E V_2({}_c \hat{M}^r). \quad (4.2.2)$$

Under C4,

$$\begin{aligned} n_TV E_2({}_c\hat{M}^r) &= n_TV(p{}_c\hat{M}) \\ &= p^2 n_TV({}_c\hat{M}) \\ &= O(1) \end{aligned}$$

from Lemma 2.7. Also, under C4,

$$\begin{aligned} n_TE V_2({}_c\hat{M}^r) &= p(1-p)n_TE \left(\sum_{(hik) \in s} \tilde{w}_{hik}^2 {}_c\delta_{hik} \right) \\ &= O(1) \end{aligned}$$

since

$$\begin{aligned} n_TE \left(\sum_{(hik) \in s} \tilde{w}_{hik}^2 {}_c\delta_{hik} \right) &\leq n_T \sum_h \sum_i E \left(\sum_k \tilde{w}_{hik}^2 \right) \\ &\leq n_T \max_{h,i,k} \tilde{w}_{hik} E \left(\sum_{(hik) \in s} \tilde{w}_{hik} \right) \\ &\leq n_T \max_{h,i,k} m_{hi} \tilde{w}_{hik} \\ &= O(1) \end{aligned}$$

noting that $E \left(\sum_{(hik) \in s} \tilde{w}_{hik} \right) = 1$ and $m_{hi} > 0$. Therefore, $n_TV({}_c\hat{M}^r) = O(1)$ which implies ${}_c\hat{M}^r \rightarrow_p p{}_cM/M$ and by Slutsky's Theorem, Lemma (2.3),

$$\frac{{}_cM}{M{}_c\hat{M}^r} = \frac{{}_cM}{{}_c\hat{M}^r} \rightarrow_p \frac{1}{p} \quad (4.2.3)$$

for $p > 0$. Finally, noting that $E({}_c\bar{e}^r) = 0$, we have ${}_c\bar{e}^r \rightarrow_p 0$ since

$$P \{ |{}_c\bar{e}^r| > \epsilon \} \leq \frac{1}{\epsilon^2} V({}_c\bar{e}^r) \rightarrow 0$$

under C8. Thus, for all c we have

$$\left(\frac{{}_cM}{M{}_c\hat{M}^r} - \frac{1}{p} \right) {}_c\bar{e}^r \rightarrow_p 0$$

and

$$\begin{aligned}\bar{y}_{N,PS} - \bar{Y} &= \frac{1}{p} \bar{e}^r + \sum_c \left(\frac{{}_c M}{{}_c \hat{M}^r} - \frac{1}{p} \right) {}_c \bar{e}^r \\ &\xrightarrow{p} 0.\end{aligned}$$

Theorem 4.2 *Under C4, C7, C8, C9 and the uniform response mechanism.*

$$\sqrt{n_T}(\bar{y}_{N,PS} - \bar{Y}) \xrightarrow{d} N(0, \sigma_{er}^2/p^2)$$

where p is the probability of a response under the uniform response mechanism.

Proof

Working with the first term in (4.2.1), let

$$X_{hi} = n_T(e_{hi} - E(e_{hi})).$$

Under C9, we have $(1/n_T) \sum_h \sum_i V(X_{hi}) \xrightarrow{p} \sigma_{er}^2$ satisfying condition 1 of Lemma 2.1 and C7 gives

$$\begin{aligned}\frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{2+\delta} &= n_T^{1+\delta} \sum_h \sum_i E|e_{hi} - E(e_{hi})|^{2+\delta} \\ &= O(1).\end{aligned}$$

By Lemma 2.1, we have

$$\sqrt{n_T} \bar{e}^r \xrightarrow{d} N(0, \sigma_{er}^2).$$

Considering the second term in (4.2.1), C4 implies ${}_c M / {}_c \hat{M}^r \xrightarrow{p} 1/p$ and C8 implies $\sqrt{n_T}({}_c \bar{e}^r - E({}_c \bar{e}^r)) = \sqrt{n_T} {}_c \bar{e}^r = O_p(1)$. Lemma 2.8 then gives

$$\left(\frac{{}_c M}{{}_c \hat{M}^r} - \frac{1}{p} \right) \sqrt{n_T} {}_c \bar{e}^r \xrightarrow{p} 0$$

for all c . Therefore, by Lemma 2.3 we have

$$\begin{aligned}\sqrt{n_T}(\bar{y}_{N,PS} - \bar{Y}) &= \sqrt{n_T} \frac{1}{p} \bar{e}^r + \sqrt{n_T} \sum_c \left(\frac{{}_c M}{{}_c \hat{M}^r} - \frac{1}{p} \right) {}_c \bar{e}^r \\ &\xrightarrow{d} N(0, \sigma_{er}^2/p^2).\end{aligned}$$

A customary linearization variance estimator is given by

$$v_L(\bar{y}_{N,PS}) = \left(\frac{\tilde{M}}{\tilde{M}^r} \right)^2 \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^* - \bar{e}_h^*)^2$$

where $\tilde{M} = \sum_{(hik) \in s} \tilde{w}_{hik}$, $\tilde{M}^r = \sum_{(hik) \in s} \tilde{w}_{hik} a_{hik}$, $e_{hi}^* = \sum_k \tilde{w}_{hik} a_{hik} e_{hik}^*$, $\bar{e}_h^* = (1/n_h) \sum_i e_{hi}^*$, $e_{hik}^* = y_{hik} - \sum_c \bar{y}_c \delta_{hik}$ and $\bar{y}_c = \tilde{Y}^r / \tilde{M}^r$. Recall, that in the proof of Theorem 2.6, we replaced the estimated residual, $e_{hik}^* = y_{hik} - \sum_c \bar{y}_c \delta_{hik}$, with the true residual, $e_{hik} = y_{hik} - \sum_c \bar{Y}_c \delta_{hik}$, and showed that the remainder converges to zero in probability. Using the same techniques, we now show consistency of $v_L(\bar{y}_{N,PS})$ by considering

$$v_L^*(\bar{y}_{N,PS}) = \left(\frac{\hat{M}}{\hat{M}^r} \right)^2 \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2$$

where $e_{hi} = \sum_k \hat{w}_{hik} a_{hik} e_{hik}$, $\bar{e}_h = (1/n_h) \sum_i e_{hi}$ and $e_{hik} = y_{hik} - \sum_c \bar{Y}_c \delta_{hik}$. Note that $v_L^*(\bar{y}_{N,PS})$ is not a true variance estimator as it involves the unknown population poststratum means, $\bar{Y}_c$.

Theorem 4.3 *Under C4, C7, C9 and the uniform response mechanism.*

$$n_T (v_L^*(\bar{y}_{N,PS}) - V(\bar{y}_{N,PS})) \longrightarrow_p 0.$$

Proof

Rewriting $v_L^*(\bar{y}_{N,PS})$ as

$$\begin{aligned} v_L^*(\bar{y}_{N,PS}) &= \left(\frac{1}{p^2} \right) \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 + \\ &\quad \left(\frac{\hat{M}}{\hat{M}^r} - \frac{1}{p} \right)^2 \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 + \\ &\quad 2 \frac{1}{p} \left(\frac{\hat{M}}{\hat{M}^r} - \frac{1}{p} \right) \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 \\ &= \frac{1}{p^2} v_1 + \left(\frac{\hat{M}}{\hat{M}^r} - \frac{1}{p} \right)^2 v_1 + 2 \frac{1}{p} \left(\frac{\hat{M}}{\hat{M}^r} - \frac{1}{p} \right) v_1. \end{aligned}$$

Now by adding and subtracting $E(e_{hi})$ to v_1 we have

$$\begin{aligned} v_1 &= \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - E(e_{hi}))^2 - \sum_h \frac{n_h^2}{n_h - 1} (\bar{e}_h - E(e_{hi}))^2 \\ &= v_{11} - v_{12}. \end{aligned}$$

Working with v_{11} , let

$$X_{hi} = n_T \left(\frac{n_h}{n_h - 1} \right)^{1/2} (e_{hi} - E(e_{hi})).$$

Then by C7,

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}^2|^{1+\delta/2} &= n_T^{1+\delta} \sum_h \sum_i E|e_{hi} - E(e_{hi})|^{2+\delta} \\ &= O(1) \end{aligned}$$

and Lemma 2.2 gives

$$\frac{1}{n_T} \sum_h \sum_i X_{hi}^2 = n_T v_{11} \longrightarrow_p n_T \sum_h \frac{n_h^2}{n_h - 1} \sigma_{eh}^2$$

where $\sigma_{eh}^2 = V(e_{hi})$.

For v_{12} , let

$$X_h = (Ln_T)^{1/2} \frac{n_h}{(n_h - 1)^{1/2}} (\bar{e}_h - E(e_{hi}))$$

and following the proof of Theorem 2.3 we have,

$$\frac{1}{L} \sum_h E|X_h^2|^{1+\delta'/2} = O(1)$$

under C7 and C9. Lemma 2.2 gives

$$\frac{1}{L} \sum_h X_h^2 = n_T v_{12} \longrightarrow_p n_T \sum_h \frac{n_h}{n_h - 1} \sigma_{eh}^2.$$

Combining these two results gives

$$\begin{aligned} n_T v_1 &= n_T v_{11} - n_T v_{12} \\ &\longrightarrow_p n_T \sigma_{er}^2 \approx n_T p^2 V(\tilde{y}_{N,PS}). \end{aligned} \tag{4.2.4}$$

Under C4, we have

$$\begin{aligned}
 n_TV(\tilde{M}) &= n_T \sum_h \sum_i V \left(\sum_k \tilde{w}_{hik} \right) \\
 &\leq n_T \sum_h \sum_i E \left(\sum_k \tilde{w}_{hik} \right)^2 \\
 &= n_T \sum_h \sum_i E \left(\sum_k \tilde{w}_{hik} \right) \left(\sum_k \tilde{w}_{hik} \right) \\
 &\leq n_T \max_{h,i,k} m_{hi} \tilde{w}_{hik} E \left(\sum_{(hik) \in s} \tilde{w}_{hik} \right) \\
 &= n_T \max_{h,i,k} m_{hi} \tilde{w}_{hik} = O(1)
 \end{aligned}$$

noting that $E \left(\sum_{(hik) \in s} \tilde{w}_{hik} \right) = 1$. Further,

$$\begin{aligned}
 n_TV(\tilde{M}^r) &= n_T \sum_h \sum_i V \left(\sum_k \tilde{w}_{hik} a_{hik} \right) \tag{4.2.5} \\
 &\leq n_T \sum_h \sum_i E \left(\sum_k \tilde{w}_{hik} a_{hik} \right)^2 \\
 &\leq n_T \sum_h \sum_i E \left(\sum_k \tilde{w}_{hik}^2 \right) \left(\sum_k a_{hik}^2 \right) \\
 &= n_T \sum_h \sum_i m_{hi} E \left(\sum_k \tilde{w}_{hik} \tilde{w}_{hik} \right) \\
 &\leq n_T \max_{h,i,k} m_{hi} \tilde{w}_{hik} E \left(\sum_{(hik) \in s} \tilde{w}_{hik} \right) \\
 &= n_T \max_{h,i,k} m_{hi} \tilde{w}_{hik} = O(1).
 \end{aligned}$$

That is, $n_TV(\tilde{M}) = O(1)$ and $n_TV(\tilde{M}^r) = O(1)$ which in turn implies $\tilde{M} \rightarrow_p 1$, $\tilde{M}^r \rightarrow_p p$ and by the continuity of $1/x$ for $x \neq 0$,

$$\frac{\tilde{M}}{\tilde{M}^r} - \frac{1}{p} \rightarrow_p 0. \tag{4.2.6}$$

Therefore we have

$$\begin{aligned}
 n_T v_L^*(\tilde{y}_{N,PS}) &= \frac{1}{p^2} n_T v_1 + \left(\frac{\tilde{M}}{\tilde{M}^r} - \frac{1}{p} \right)^2 n_T v_1 + 2 \frac{1}{p} \left(\frac{\tilde{M}}{\tilde{M}^r} - \frac{1}{p} \right) n_T v_1 \\
 &\rightarrow_p n_TV(\tilde{y}_{N,PS})
 \end{aligned}$$

by equations (4.2.4) and (4.2.6). ■

To construct a jackknife variance estimator we first define the nonresponse adjustment when the $(gj)^{th}$ cluster has been deleted as

$$d_{(gj)} = \frac{\hat{U}_{(gj)}}{\hat{R}_{(gj)}}$$

where $\hat{U}_{(gj)} = \sum_{(hik) \in s} \hat{w}_{hik(gj)}$, $\hat{R}_{(gj)} = \sum_{(hik) \in s} \hat{w}_{hik(gj)} a_{hik}$ and $\hat{w}_{hik(gj)}$ are the normalized jackknife weights. The poststratification adjustment now becomes

$${}_c b_{(gj)} = \frac{{}_c M}{\sum_{(hik) \in s} d_{(gj)} \hat{w}_{hik(gj)} a_{hik} {}_c \delta_{hik}}$$

and the poststratified estimator, with the $(gj)^{th}$ cluster deleted is

$$\begin{aligned} \bar{y}_{N,PS(gj)} &= \sum_c \sum_{(hik) \in s} d_{(gj)} {}_c b_{(gj)} \hat{w}_{hik(gj)} y_{hik} a_{hik} {}_c \delta_{hik} \\ &= \sum_c \frac{{}_c M}{M} \frac{{}_c \hat{Y}_{(gj)}^r}{{}_c \hat{M}_{(gj)}^r} \end{aligned} \quad (4.2.7)$$

where

$${}_c \hat{Y}_{(gj)}^r = \sum_{(hik) \in s} \hat{w}_{hik(gj)} a_{hik} y_{hik} {}_c \delta_{hik}$$

and

$${}_c \hat{M}_{(gj)}^r = \sum_{(hik) \in s} \hat{w}_{hik(gj)} a_{hik} {}_c \delta_{hik}.$$

Note that $\bar{y}_{N,PS(gj)}$ does not involve the nonresponse adjustment, $d_{(gj)}$. Therefore, when using the jackknife variance method with a poststratified estimator involving a nonresponse adjustment, the nonresponse adjustment need not be recalculated each time a cluster is deleted. This is contrary to the general jackknife methodology but it is clear that the nonresponse adjustment cancels out due to the form of the poststratified estimator.

The jackknife variance estimator is then defined as

$$v_J(\bar{y}_{N,PS}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (\bar{y}_{N,PS(gj)} - \bar{y}_{N,PS})^2.$$

Asymptotic consistency of the jackknife variance estimator is established in the following theorem.

Theorem 4.4 *Under C4, C7, C8, C9 and the uniform response mechanism,*

$$n_T (v_J(\bar{y}_{N,PS}) - V(\bar{y}_{N,PS})) \longrightarrow_p 0.$$

Proof

We have

$$\begin{aligned} \frac{{}_c\hat{Y}_{(gj)}^r}{{}_c\hat{M}_{(gj)}^r} - \frac{{}_c\hat{Y}^r}{{}_c\hat{M}^r} &= \frac{1}{{}_c\hat{M}_{(gj)}^r} ({}_c\hat{Y}_{(gj)}^r - {}_c\hat{Y}^r - {}_c\bar{y}({}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r)) \\ &= \frac{1}{{}_c\hat{M}_{(gj)}^r} \frac{n_g}{n_g - 1} ({}_c\bar{e}_g^* - {}_ce_{gj}^*). \end{aligned} \quad (4.2.8)$$

Using a Taylor series expansion, we obtain

$$\frac{1}{{}_c\hat{M}_{(gj)}^r} \cong \frac{1}{{}_c\hat{M}^r} - \frac{1}{{}_c\hat{M}^{r2}} ({}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r).$$

Putting this expression in (4.2.8) gives

$$\begin{aligned} \frac{{}_c\hat{Y}_{(gj)}^r}{{}_c\hat{M}_{(gj)}^r} - \frac{{}_c\hat{Y}^r}{{}_c\hat{M}^r} &\cong \frac{n_g}{n_g - 1} \frac{1}{{}_c\hat{M}^r} ({}_c\bar{e}_g^* - {}_ce_{gj}^*) - \\ &\quad \frac{n_g}{n_g - 1} \frac{1}{{}_c\hat{M}^{r2}} ({}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r) ({}_c\bar{e}_g^* - {}_ce_{gj}^*). \end{aligned}$$

Using this expression, the jackknife variance estimator is

$$\begin{aligned} v_J(\bar{y}_{N,PS}) &\cong \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{{}_c\hat{M}^r} ({}_c\bar{e}_g^* - {}_ce_{gj}^*) \right]^2 + \\ &\quad \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{{}_c\hat{M}^r} \frac{1}{{}_c\hat{M}^{r2}} ({}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r) ({}_c\bar{e}_g^* - {}_ce_{gj}^*) \right]^2 - \\ &\quad 2(\text{ cross-product terms }). \end{aligned}$$

Under C4 and C8, following the lines of Section 3.2.3 we have

$$\sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{{}_c\hat{M}^r} \frac{1}{{}_c\hat{M}^{r2}} ({}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r) ({}_c\bar{e}_g^* - {}_ce_{gj}^*) \right]^2 = O_p \left(\frac{1}{n_T} \right)^3$$

and the cross-product term is $O_p(1/n_T)^2$.

Therefore,

$$v_J(\bar{y}_{N,PS}) = v_{JL}(\bar{y}_{N,PS}) + O_p\left(\frac{1}{n_T}\right)^2$$

where

$$\begin{aligned} v_{JL}(\bar{y}_{N,PS}) &= \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{{}_c\hat{M}^r} ({}_c\bar{e}_g^* - {}_ce_{gj}^*) \right]^2 \\ &= v(r_{hi}) \end{aligned} \quad (4.2.9)$$

with $r_{hi} = \sum_k {}_c\hat{w}_{hik} a_{hik} e_{hik}$. Note that $v_{JL}(\bar{y}_{N,PS})$ is analogous to the robust variance estimator suggested by Rao (1985) in the full response situation. This variance estimator, $v_{JL}(\bar{y}_{N,PS})$, is a computationally simpler alternative to the jackknife variance estimator while still retaining properties of the jackknife variance estimator. It can be used with existing software for linearization variance estimation.

As was done in establishing consistency of $v_L(\bar{y}_{N,PS})$, we will replace the estimated residual, e_{hik}^* , with the true residual, e_{hik} , and work with

$$v_{JL}^*(\bar{y}_{N,PS}) = \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_cM}{{}_c\hat{M}^r} ({}_c\bar{e}_g - {}_ce_{gj}) \right]^2$$

where ${}_ce_{hi} = \sum_k {}_c\hat{w}_{hik} a_{hik} {}_ce_{hik}$ and ${}_ce_{hik} = y_{hik} - {}_c\bar{Y}$. The difference between $v_{JL}(\bar{y}_{N,PS})$ and $v_{JL}^*(\bar{y}_{N,PS})$ can be shown to converge to zero in probability using techniques from Chapter 2. These calculations are lengthy but straightforward and will not be shown here. Now, we can rewrite $v_{JL}^*(\bar{y}_{N,PS})$ as

$$\begin{aligned} v_{JL}^*(\bar{y}_{N,PS}) &= \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\frac{1}{p} \sum_c ({}_c\bar{e}_h - {}_ce_{hi}) \right]^2 + \\ &\quad \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \left(\frac{{}_cM}{{}_c\hat{M}^r} - \frac{1}{p} \right) ({}_c\bar{e}_h - {}_ce_{hi}) \right]^2 + \\ &\quad 2(\text{cross-product terms}) \\ &= \frac{1}{p^2} v_1 + v_2 + 2(\text{cross-product terms}). \end{aligned}$$

In Theorem 4.3 it was shown that

$$n_T v_1 \longrightarrow_p n_T \sigma_{\epsilon r}^2$$

so

$$\frac{1}{p^2} n_T v_1 \longrightarrow_p n_T \frac{\sigma_{\epsilon r}^2}{p^2}.$$

Turning to v_2 under C4 we have, from (4.2.3),

$$\frac{{}_c M}{{}_c \hat{M}^r} \longrightarrow_p \frac{1}{p}$$

and under C8 we have $n_T v_L({}_c \bar{\epsilon}^r) = O(1)$. Combining these two results gives

$$\begin{aligned} n_T v_2 &\leq \sum_c \left(\frac{{}_c M}{{}_c \hat{M}^r} - \frac{1}{p} \right)^2 n_T \sum_h \frac{n_h}{n_h - 1} \sum_i \sum_c ({}_c \bar{e}_h - {}_c e_{hi})^2 \\ &= \sum_c \left(\frac{{}_c M}{{}_c \hat{M}^r} - \frac{1}{p} \right)^2 n_T v_L({}_c \bar{\epsilon}^r) \\ &\longrightarrow_p 0 \end{aligned}$$

Finally, the Schwarz inequality implies the cross-product terms also converge to zero in probability. Thus we have,

$$n_T v_{JL}^*(\bar{y}_{N,PS}) \longrightarrow_p n_T V(\bar{y}_{N,PS}).$$

Now, since

$$v_J(\bar{y}_{N,PS}) = v_{JL}^*(\bar{y}_{N,PS}) + O_p \left(\frac{1}{n_T} \right)^2$$

we have

$$n_T (v_J(\bar{y}_{N,PS}) - V(\bar{y}_{N,PS})) \longrightarrow_p 0.$$

■

4.2.2 Multiple Weighting Classes

The assumption of equal response probabilities, p , is often unrealistic in practice. It can be relaxed by forming $J \geq 2$ weighting classes based on characteristics known for

all sample units regardless of whether the unit responds or not. We then assume a uniform response mechanism within each weighting class, but allow the probability of response, $p_j > 0$, to be different between weighting classes. The nonresponse adjustment factor in the j^{th} weighting class is

$$d_j = \frac{\sum_{(hik) \in s_j} \hat{w}_{hik}}{\sum_{(hik) \in s_j} \hat{w}_{hik} a_{hik}} = \frac{\hat{U}_j}{\hat{R}_j}$$

where s_j represents the sample in the j^{th} weighting class. An estimate of the mean is given as

$$\bar{y}_N = \sum_{j=1}^J \sum_{(hik) \in s_j} d_j \hat{w}_{hik} a_{hik} y_{hik}.$$

When using the poststratified estimator with weighting classes cutting across poststrata, the nonresponse adjustment is usually calculated first (see Flyer, Rust and Morganstein, 1989). The poststratification adjustment is then calculated using the nonresponse adjusted weights,

$$\begin{aligned} {}_c b &= \frac{{}_c M}{M \sum_j \sum_{(hik) \in s_j} d_j \hat{w}_{hik} a_{hik} {}_c \delta_{hik}} \\ &= \frac{{}_c M}{M \sum_j d_j {}_c \hat{M}_j^r} \end{aligned}$$

where ${}_c b$ is the poststratification adjustment factor for the c^{th} poststratum and

$${}_c \hat{M}_j^r = \sum_{(hik) \in s_j} \hat{w}_{hik} a_{hik} {}_c \delta_{hik}.$$

The poststratified estimator is then

$$\begin{aligned} \bar{y}_{N,PS} &= \sum_c \sum_j \sum_{(hik) \in s_j} d_j {}_c b \hat{w}_{hik} a_{hik} y_{hik} {}_c \delta_{hik} \\ &= \sum_c \frac{{}_c M}{M} \left(\frac{\sum_j d_j {}_c \hat{Y}_j^r}{\sum_j d_j {}_c \hat{M}_j^r} \right) \\ &= \sum_c \frac{{}_c M}{M} \frac{{}_c \hat{S}}{{}_c \hat{T}} \end{aligned}$$

where

$$\begin{aligned} {}_c\hat{S} &= \sum_j d_j {}_c\hat{Y}_j^r, \\ {}_c\hat{T} &= \sum_j d_j {}_c\hat{M}_j^r, \text{ and} \\ {}_c\hat{Y}_j^r &= \sum_{(hik) \in s_j} \tilde{w}_{hik} a_{hik} y_{hik} {}_c\delta_{hik}. \end{aligned}$$

Noting that

$$\begin{aligned} E(\hat{R}_j) &= p_j E(\hat{U}_j), \\ E({}_c\hat{Y}_j^r) &= p_j E({}_c\hat{Y}_j), \text{ and} \\ E({}_c\hat{M}_j^r) &= p_j E({}_c\hat{M}_j) \end{aligned}$$

we have

$$\begin{aligned} E({}_c\hat{S}) &\approx \sum_j \frac{1}{p_j} p_j E({}_c\hat{Y}_j) \\ &= E({}_c\hat{Y}) = \frac{{}_cY}{M}, \\ E({}_c\hat{T}) &\approx \sum_j \frac{1}{p_j} p_j E({}_c\hat{M}_j) \\ &= E({}_c\hat{M}) = \frac{{}_cM}{M} \text{ and} \\ E(\bar{y}_{N,PS}) &\approx \sum_c \frac{{}_cM}{{}_cM} \frac{{}_cY}{M} = \bar{Y}. \end{aligned}$$

Thus, $\bar{y}_{N,PS}$ is approximately unbiased for $\bar{Y}$. Note that the poststratified estimator depends on the nonresponse adjustment factors, d_j , unlike in the case of a single weighting class ($J = 1$).

In order to construct a jackknife variance estimator of $\bar{y}_{N,PS}$, we first need to define the following when the $(gl)^{th}$ cluster has been deleted,

$${}_c\hat{Y}_{j(gl)}^r = \sum_{(hik) \in s_j} \tilde{w}_{hik(gl)} a_{hik} y_{hik} {}_c\delta_{hik}, \quad (4.2.10)$$

$${}_c\hat{M}_{j(gl)}^r = \sum_{(hik) \in s_j} \hat{w}_{hik(gl)} a_{hik} {}_c\delta_{hik}, \quad (4.2.11)$$

$$\hat{R}_{j(gl)} = \sum_{(hik) \in s_j} \hat{w}_{hik(gl)} a_{hik}, \quad \text{and} \quad (4.2.12)$$

$$\hat{U}_{j(gl)} = \sum_{(hik) \in s_j} \hat{w}_{hik(gl)}, \quad (4.2.13)$$

where $\hat{w}_{hik(gl)}$ are the normalized jackknife weights. The poststratified estimator with the $(gl)^{th}$ cluster deleted is

$$\begin{aligned} \bar{y}_{N,PS(gl)} &= \sum_c \sum_j \sum_{(hik) \in s_j} d_{j(gl)} {}_c b_{(gl)} \hat{w}_{hik(gl)} a_{hik} y_{hik} {}_c \delta_{hik} \\ &= \sum_c \frac{{}_c M}{{}_c \hat{M}} \frac{{}_c \hat{S}_{(gl)}}{{}_c \hat{T}_{(gl)}} \end{aligned}$$

where

$${}_c \hat{S}_{(gl)} = \sum_j d_{j(gl)} {}_c \hat{Y}_{j(gl)}^r, \quad (4.2.14)$$

$${}_c \hat{T}_{(gl)} = \sum_j d_{j(gl)} {}_c \hat{M}_{j(gl)}^r. \quad (4.2.15)$$

$$d_{j(gl)} = \frac{\hat{U}_{j(gl)}}{\hat{R}_{j(gl)}}, \quad \text{and} \quad (4.2.16)$$

$${}_c b_{(gl)} = \frac{{}_c M}{{}_c \hat{M}} \frac{1}{{}_c \hat{T}_{(gl)}}. \quad (4.2.17)$$

The jackknife variance estimator of $\bar{y}_{N,PS}$ is

$$v_J(\bar{y}_{N,PS}) = \sum_g \frac{n_g - 1}{n_g} \sum_l (\bar{y}_{N,PS(gl)} - \bar{y}_{N,PS})^2.$$

Consistency of the jackknife variance estimator can be established along the same lines as Theorem 4.4 for the single weighting class. The proof is therefore omitted, for simplicity.

To obtain a linearization type variance estimator, we can linearize the jackknife variance estimator and use expressions (4.2.10) - (4.2.13) to simplify the approximation. Using a Taylor series expansion, we have

$$d_{j(gl)} \cong d_j + \frac{1}{\hat{R}_j} [\hat{U}_{j(gl)} - \bar{U}_j - d_j(\hat{R}_{j(gl)} - \bar{R}_j)]$$

$$= d_j + \Delta_{jgl}. \quad (4.2.18)$$

Using this expression for $d_{j(gl)}$ in equations (4.2.14) and (4.2.15), we obtain

$${}_c\hat{S}_{(gl)} = \sum_j d_j {}_c\hat{Y}_{j(gl)}^r + \sum_j \Delta_{jgl} {}_c\hat{Y}_j^r$$

and

$${}_c\hat{T}_{(gl)} = \sum_j d_j {}_c\hat{M}_{j(gl)}^r + \sum_j \Delta_{jgl} {}_c\hat{M}_j^r.$$

Noting that equations (4.2.10) and (4.2.11) can be rewritten as

$$\begin{aligned} {}_c\hat{Y}_{j(gl)}^r &= {}_c\hat{Y}_j^r + \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s_j} \hat{w}_{gik} a_{gik} y_{gik} {}_c\delta_{gik} - \sum_{(gik) \in s_j} \hat{w}_{gik} a_{gik} y_{gik} {}_c\delta_{gik} \right) \\ &= {}_c\hat{Y}_j^r + {}_cA_{j(gl)}^r \end{aligned}$$

and

$$\begin{aligned} {}_c\hat{M}_{j(gl)}^r &= {}_c\hat{M}_j^r + \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s_j} \hat{w}_{gik} a_{gik} {}_c\delta_{gik} - \sum_{(gik) \in s_j} \hat{w}_{gik} a_{gik} {}_c\delta_{gik} \right) \\ &= {}_c\hat{M}_j^r + {}_cB_{j(gl)}^r \end{aligned}$$

we have, ignoring lower order terms,

$$\begin{aligned} {}_c\hat{S}_{(gl)} &= \sum_j d_j {}_c\hat{Y}_j^r + \sum_j d_j {}_cA_{j(gl)}^r + \sum_j \Delta_{jgl} {}_c\hat{Y}_j^r + \sum_j \Delta_{jgl} {}_cA_{j(gl)}^r \\ &\cong {}_c\hat{S} + \sum_j d_j {}_cA_{j(gl)}^r + \sum_j \Delta_{jgl} {}_c\hat{Y}_j^r \end{aligned} \quad (4.2.19)$$

and

$$\begin{aligned} {}_c\hat{T}_{(gl)} &= \sum_j d_j {}_c\hat{M}_j^r + \sum_j d_j {}_cB_{j(gl)}^r + \sum_j \Delta_{jgl} {}_c\hat{M}_j^r + \sum_j \Delta_{jgl} {}_cB_{j(gl)}^r \\ &\cong {}_c\hat{T} + \sum_j d_j {}_cB_{j(gl)}^r + \sum_j \Delta_{jgl} {}_c\hat{M}_j^r \end{aligned} \quad (4.2.20)$$

A Taylor series expansion of

$$f({}_c\hat{S}_{(gl)}, {}_c\hat{T}_{(gl)}) = \frac{{}_c\hat{S}_{(gl)}}{{}_c\hat{T}_{(gl)}}$$

gives

$$\frac{{}_c\hat{S}_{(gl)}}{{}_c\hat{T}_{(gl)}} \cong \frac{{}_c\hat{S}}{{}_c\hat{T}} + \frac{1}{{}_c\hat{T}} \left[{}_c\hat{S}_{(gl)} - {}_c\hat{S} - \frac{{}_c\hat{S}}{{}_c\hat{T}}({}_c\hat{T}_{(gl)} - {}_c\hat{T}) \right]$$

or

$$\frac{{}_c\hat{S}_{(gl)}}{{}_c\hat{T}_{(gl)}} - \frac{{}_c\hat{S}}{{}_c\hat{T}} \cong \frac{1}{{}_c\hat{T}} \left[{}_c\hat{S}_{(gl)} - {}_c\hat{S} - {}_c\bar{y}_N({}_c\hat{T}_{(gl)} - {}_c\hat{T}) \right] \quad (4.2.21)$$

where ${}_c\bar{y}_N = {}_c\hat{S}/{}_c\hat{T}$ is a nonresponse adjusted estimate of the poststratum mean, ${}_c\bar{Y}$. Using expressions (4.2.19) and (4.2.20) in equation (4.2.21) gives

$$\begin{aligned} & \frac{{}_c\hat{S}_{(gl)}}{{}_c\hat{T}_{(gl)}} - \frac{{}_c\hat{S}}{{}_c\hat{T}} \\ & \cong \frac{1}{{}_c\hat{T}} \left[\sum_j d_j {}_cA_{j(gl)}^r + \sum_j \Delta_{jgl} {}_c\hat{Y}_j^r - {}_c\bar{y}_N \left(\sum_j d_j {}_cB_{j(gl)}^r + \sum_j \Delta_{jgl} {}_c\hat{M}_j^r \right) \right] \\ & = \frac{1}{{}_c\hat{T}} \left[\sum_j d_j ({}_cA_{j(gl)}^r - {}_c\bar{y}_N {}_cB_{j(gl)}^r) + \sum_j \Delta_{jgl} ({}_c\hat{Y}_j^r - {}_c\bar{y}_N {}_c\hat{M}_j^r) \right]. \end{aligned} \quad (4.2.22)$$

Now,

$${}_cA_{j(gl)}^r - {}_c\bar{y}_N {}_cB_{j(gl)}^r = \frac{n_g}{n_g - 1} ({}_c\bar{\epsilon}_{jg}^r - {}_c\bar{\epsilon}_{jgl}^r)$$

and

$${}_c\hat{Y}_j^r - {}_c\bar{y}_N {}_c\hat{M}_j^r = {}_c\bar{\epsilon}_j^r$$

where

$$\begin{aligned} {}_c\bar{\epsilon}_{jgl}^r &= \sum_{(g|k) \in s_j} \hat{w}_{g|k} a_{g|k} {}_c\delta_{g|k} (y_{g|k} - {}_c\bar{y}_N), \\ {}_c\bar{\epsilon}_{jg}^r &= \frac{1}{n_g} \sum_j {}_c\bar{\epsilon}_{jgl}^r \quad \text{and} \\ {}_c\bar{\epsilon}_j^r &= \sum_{(h|k) \in s_j} \hat{w}_{h|k} a_{h|k} {}_c\delta_{h|k} (y_{h|k} - {}_c\bar{y}_N). \end{aligned}$$

Also,

$$\Delta_{jgl} = \frac{1}{\hat{R}_j} [\hat{U}_{j(gl)} - \hat{U}_j - d_j (\hat{R}_{j(gl)} - \hat{R}_j)]$$

$$\begin{aligned}
&= \frac{1}{\hat{R}_j} \frac{n_g}{n_g - 1} \left[\frac{1}{n_g} \sum_{(gik) \in s_j} \hat{w}_{gik} (1 - d_j a_{gik}) - \sum_{(gik) \in s_j} \hat{w}_{gik} (1 - d_j a_{gik}) \right] \\
&= \frac{n_g}{n_g - 1} \Delta'_{jgl}.
\end{aligned}$$

Therefore, equation (4.2.22) can be rewritten as

$$\frac{{}_c\hat{S}_{(gl)}}{{}_c\hat{T}_{(gl)}} - \frac{{}_c\hat{S}}{{}_c\hat{T}} \cong \frac{n_g}{n_g - 1} \frac{1}{{}_c\hat{T}} \left[\sum_j d_j ({}_c\bar{e}_{jg}^r - {}_c\bar{e}_{jgl}^r) + \sum_j \Delta'_{jgl} {}_c\bar{e}_j^r \right]$$

and the jackknife variance estimator can be approximated by

$$\begin{aligned}
v_{JL}(\bar{y}_{N,PS}) &= \sum_g \frac{n_g}{n_g - 1} \sum_l \left[\sum_c \frac{{}_cM}{M} \frac{1}{{}_c\hat{T}} \left(\sum_j d_j ({}_c\bar{e}_{jg}^r - {}_c\bar{e}_{jgl}^r) + \sum_j \Delta'_{jgl} {}_c\bar{e}_j^r \right) \right]^2 \\
&= \sum_g \frac{n_g}{n_g - 1} \sum_l \left[\sum_c \sum_j d_j {}_c\bar{b} ({}_c\bar{e}_{jg}^r - {}_c\bar{e}_{jgl}^r) + \sum_c \sum_j {}_c\bar{b} \Delta'_{jgl} {}_c\bar{e}_j^r \right]^2 \\
&= v(\epsilon_{hi}) \tag{4.2.23}
\end{aligned}$$

where the operator notation $v(\epsilon_{hi})$ signifies $v_{JL}(\bar{y}_{N,PS})$ depends only on the cluster totals, ϵ_{hi} defined as

$$\begin{aligned}
\epsilon_{hi} &= \sum_c \sum_j \sum_k d_j {}_c\bar{b} \hat{w}_{hik} a_{hik} {}_c\bar{e}_{hik}^* {}_c\delta_{hik} \eta_{hik}^j + \\
&\quad \sum_c \sum_j \sum_k \frac{1}{\hat{R}_j} {}_c\bar{b} {}_c\bar{e}_j^r \hat{w}_{hik} \eta_{hik}^j (1 - d_j a_{hik})
\end{aligned}$$

with η_{hik}^j being the weighting class indicator variable defined as

$$\eta_{hik}^j = \begin{cases} 1 & \text{if the } (hik)^{th} \text{ unit is in the } j^{th} \text{ weighting class} \\ 0 & \text{otherwise.} \end{cases}$$

Note that the cluster totals are made up of the correctly weighted residual plus an additional term contributed by the nonresponse adjustment. This linearization variance estimator, $v_{JL}(\bar{y}_{N,PS})$, can be used with existing software for linearization variance estimators. It is an alternative to the computationally costly jackknife

variance estimator, while retaining certain important properties of the jackknife variance estimator.

When the poststrata are used as weighting classes it is easy to show that $\bar{y}_{N,PS}$ is identical to the one obtained in the case of one weighting class. The linearized jackknife variance estimator obtained for this special case coincides with the linearization variance estimator given in Jones and Chromy (1982).

4.3 The Regression Estimator

4.3.1 One Weighting Class

Suppose that instead of using a poststratified estimator, we use a generalized regression estimator with the auxiliary vector, $\mathbf{x}_{hik}$, consisting of indicator variables. In the single weighting class situation, the nonresponse adjustment remains the same as before,

$$d = \frac{\sum_{(hik)cs} \hat{w}_{hik}}{\sum_{(hik)cs} \hat{w}_{hik} a_{hik}} = \frac{\hat{U}}{\hat{R}}$$

and the regression adjustment is

$$g_{hik} = \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}}$$

where $\hat{\mathbf{A}}^{-1} = \sum_{(hik)cs} d \hat{w}_{hik} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T$, and $\bar{\mathbf{X}}$ is the vector of known population means. The regression estimator, with the nonresponse adjustment is

$$\begin{aligned} \bar{y}_{N,GREG} &= \sum_{(hik)cs} d \hat{w}_{hik}^* a_{hik} y_{hik} \\ &= \hat{\beta}^T \bar{\mathbf{X}} \end{aligned} \quad (4.3.1)$$

where

$$\begin{aligned} \hat{\beta} &= \left(\sum_{(hik)cs} d \hat{w}_{hik} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T \right)^{-1} \sum_{(hik)cs} d \hat{w}_{hik} a_{hik} y_{hik} \mathbf{x}_{hik} \\ &= \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}} \end{aligned}$$

and $\tilde{w}_{hik}^* = \tilde{w}_{hik} g_{hik}$. This estimator is precisely the regression estimator based on the respondents only. It is easily shown that this estimator will benchmark to the known population means (i.e. $\bar{\mathbf{x}}_{N,GREG} = \bar{\mathbf{X}}$). Under a uniform response mechanism, $\bar{y}_{N,GREG}$, is approximately unbiased for the population mean,

$$\begin{aligned} E(\bar{y}_{N,GREG}) &\cong \sum_{(hik)_{cs}} \frac{1}{p} p E(\tilde{w}_{hik}^* y_{hik}) \\ &= E(\bar{y}_{GREG}) = \bar{Y}. \end{aligned}$$

To investigate the large sample properties of $\bar{y}_{N,GREG}$, it is more convenient to work with the equivalent form

$$\bar{y}_{N,GREG} = \hat{Y} + \hat{\beta}^T (\bar{\mathbf{X}} - \hat{\mathbf{X}}) \quad (4.3.2)$$

where $\hat{Y} = \sum_{(hik)_{cs}} d\tilde{w}_{hik} a_{hik} y_{hik}$ and $\hat{\mathbf{X}} = \sum_{(hik)_{cs}} d\tilde{w}_{hik} a_{hik} \mathbf{x}_{hik}$. The two forms of $\bar{y}_{N,GREG}$, equations (4.3.1) and (4.3.2), are equivalent in the special case where the vector $\mathbf{1}$ is in the column space of the $m \times T$ matrix $\mathbf{X}_d$ where m is the sample size and T is the number of poststrata categories (Särndal, 1989). In our special case where $\mathbf{x}_{hik}$ are vectors of indicator variables, this condition is satisfied. Note that this equivalence also holds for the full response GREG.

Denoting β as the vector of unknown population regression coefficients, we can write $\bar{Y} = \beta^T \bar{\mathbf{X}}$ and using equation (4.3.2) we have

$$\begin{aligned} \bar{y}_{N,GREG} - \bar{Y} &= \hat{Y} - \beta^T \bar{\mathbf{X}} + \hat{\beta}^T (\bar{\mathbf{X}} - \hat{\mathbf{X}}) \\ &= \hat{Y} - \beta^T \hat{\mathbf{X}} + (\hat{\beta} - \beta)^T (\bar{\mathbf{X}} - \hat{\mathbf{X}}) \\ &= \sum_{(hik)_{cs}} d\tilde{w}_{hik} a_{hik} e_{hik} + (\hat{\beta} - \beta)^T (\bar{\mathbf{X}} - \hat{\mathbf{X}}) \\ &= d\bar{e}^r + (\hat{\beta} - \beta)^T (\bar{\mathbf{X}} - \hat{\mathbf{X}}) \end{aligned} \quad (4.3.3)$$

where $\bar{e}^r = \sum_{(hik)_{cs}} \tilde{w}_{hik} a_{hik} e_{hik}$ and $e_{hik} = y_{hik} - \beta^T \mathbf{x}_{hik}$.

We will assume the following regularity conditions

$$C10 \quad n_T^{1+\delta} \sum_h \sum_i E|e_{hi} - E(e_{hi})|^{2+\delta} = O(1) \text{ where } e_{hi} = \sum_k \tilde{w}_{hik} a_{hik} e_{hik}$$

C11 $n_TV(\bar{e}^r) \rightarrow \sigma_{er}^2$ where $\bar{e}^r = \sum_{(hik)es} \hat{w}_{hik} a_{hik} e_{hik}$

C12 $n_TV(\bar{\mathbf{Z}}) = O(1)$ where $\bar{\mathbf{Z}} = \sum_{(hik)es} \hat{w}_{hik} e_{hik} \mathbf{x}_{hik}$.

Conditions C10 and C11 are analogous to condition C12 and C13 for the one-way poststratified estimator, while condition C12 is sufficient to show $\sqrt{n_T}(\hat{\beta} - \beta) = O_p(1)$:

$$\begin{aligned} \hat{\beta} - \beta &= \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}} - \beta \\ &= \hat{\mathbf{A}}^{-1} \left(\sum_{(hik)es} d\hat{w}_{hik} a_{hik} y_{hik} \mathbf{x}_{hik} - \sum_{(hik)es} d\hat{w}_{hik} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \beta \right) \\ &= \hat{\mathbf{A}}^{-1} (d\hat{w}_{hik} a_{hik} e_{hik} \mathbf{x}_{hik}) \\ &= \hat{\mathbf{A}}^{-1} \bar{\mathbf{Z}}. \end{aligned}$$

Now, note that $E(\bar{\mathbf{Z}}) = 0$ and condition C12 imply $\bar{\mathbf{Z}} = O_p(n_T^{-1/2})$. Noting that $\hat{\mathbf{A}} = O_p(1)$, we have

$$\hat{\beta} - \beta = O_p\left(\frac{1}{\sqrt{n_T}}\right).$$

Asymptotic normality is now established in the following theorem.

Theorem 4.5 *Under conditions C4, C10, C11, C12 and the uniform response mechanism*

$$\sqrt{n_T}(\bar{y}_{N,GREG} - \bar{Y}) \longrightarrow_d N(0, \sigma_{er}^2/p^2)$$

where p is the probability of a response under the uniform response mechanism.

Proof

Working with equation (4.3.3), to establish the asymptotic normality of $\bar{y}_{N,GREG}$, we will show

$$\sqrt{n_T} \bar{e}^r \longrightarrow_d N(0, \sigma_{er}^2), \quad (4.3.4)$$

$$d - p^{-1} \longrightarrow_p 0, \quad (4.3.5)$$

and

$$\sqrt{n_T}(\hat{\beta} - \beta)^T(\bar{\mathbf{X}} - \hat{\mathbf{X}}) \longrightarrow_p 0. \quad (4.3.6)$$

To show (4.3.4), let

$$X_{hi} = n_T(e_{hi} - E(e_{hi}))$$

where $e_{hi} = \sum_k \hat{w}_{hik} a_{hik} e_{hik}$. Then under condition C11, we have

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i V(X_{hi}) &= n_T V(\bar{e}^r) \\ &\rightarrow \sigma_{er}^2. \end{aligned}$$

Now, by condition C10 we have

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{2+\delta} &= n_T^{1+\delta} \sum_h \sum_i E|e_{hi} - E(e_{hi})|^{2+\delta} \\ &= O(1). \end{aligned}$$

Therefore the conditions of Lemma (2.1) are satisfied, giving

$$\sqrt{n_T} \bar{e}^r \longrightarrow_d N(0, \sigma_{er}^2). \quad (4.3.7)$$

To show (4.3.5), we note that under C4, see (4.2.6),

$$d = \frac{\hat{M}}{\hat{M}^r} \longrightarrow_p \frac{1}{p}.$$

Therefore, by Slutsky's Theorem, Lemma (2.3),

$$d\sqrt{n_T} \bar{e}^r \longrightarrow_d N(0, \sigma_{er}^2/p^2).$$

Finally, to show (4.3.6) we write $\sqrt{n_T}(\hat{\beta} - \beta)^T(\bar{\mathbf{X}} - \hat{\mathbf{X}})$ as

$$\sqrt{n_T}(\hat{\beta} - \beta)^T(\bar{\mathbf{X}} - \hat{\mathbf{X}}) = \sum_i \sqrt{n_T}(\hat{\beta}_i - \beta_i)^T(\bar{X}_i - \hat{X}_i)$$

where $\hat{\beta}_i, \beta_i, \bar{X}_i$, and $\hat{X}_i$ are the i^{th} elements of the vectors $\hat{\beta}, \beta, \bar{\mathbf{X}}$, and $\hat{\mathbf{X}}$ respectively. Note that $\bar{X}_i = {}_iM/M$ and $\hat{X}_i = d_i \hat{M}^r$ where ${}_iM$ is the population count for

the t^{th} category, and ${}_t\hat{M}^r = \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_t\delta_{hik}$ with ${}_t\delta_{hik}$ the indicator for the t^{th} category. Under C4, we have $n_T V({}_t\hat{M}^r) = O(1)$, see equation (4.2.3), which implies that

$${}_t\hat{M}^r \longrightarrow_p p \frac{{}_tM}{M}$$

by Chebyshev's inequality. Combining this result with $d \longrightarrow_p p^{-1}$, see equation (4.2.6), gives

$$\bar{X}_t - \hat{X}_t \longrightarrow_p 0$$

for each t . Therefore, under condition C12 we have by Lemma 2.8

$$\sqrt{n_T}(\hat{\beta}_t - \beta_t)^T(\bar{X}_t - \hat{X}_t) \longrightarrow_p 0$$

for all t . That is,

$$\sqrt{n_T}(\hat{\beta} - \beta)^T(\bar{\mathbf{X}} - \hat{\mathbf{X}}) \longrightarrow_p 0.$$

Finally, by Slutsky's Theorem, Lemma (2.3), we have

$$\begin{aligned} & \sqrt{n_T}(\bar{y}_{N,GREG} - \bar{Y}) \\ &= d\sqrt{n_T}\bar{e}^r + \sqrt{n_T}(\hat{\beta} - \beta)^T(\bar{\mathbf{X}} - \hat{\mathbf{X}}) \\ &\longrightarrow_d N(0, \sigma_{e^r}^2/p^2) \end{aligned}$$

as required. ■

A customary linearization variance estimator is given by

$$v_L(\bar{y}_{N,GREG}) = d^2 \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^* - \bar{e}_h^*)^2$$

where $e_{hi}^* = \sum_k \hat{w}_{hik} a_{hik} e_{hik}^*$, $e_{hik}^* = y_{hik} - \hat{\beta}^T \mathbf{x}_{hik}$ and $\bar{e}_h^* = (1/n_h) \sum_i e_{hi}^*$. As was done in Section 4.2, we will replace the estimated residual, e_{hik}^* , with the true residual, $e_{hik} = y_{hik} - \beta^T \mathbf{x}_{hik}$, and consider

$$v_L^*(\bar{y}_{N,GREG}) = d^2 \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2$$

where $e_{hi} = \sum_k \bar{w}_{hik} e_{hik}$, and $\bar{e}_h = (1/n_h) \sum_i e_{hi}$. Consistency of $v_L(\bar{y}_{N,GREG})$ follows from consistency of $v_L^*(\bar{y}_{N,GREG})$ using techniques from Chapter 2 to show that the difference converges to zero in probability. Again, we note that $v_L^*(\bar{y}_{N,GREG})$ is not a true variance estimator. Asymptotic consistency of $v_L^*(\bar{y}_{N,GREG})$ is now established in the following theorem.

Theorem 4.6 *Under conditions C4, C10, C11, C12 and the uniform response mechanism*

$$n_T(v_L^*(\bar{y}_{N,GREG}) - V(\bar{y}_{N,GREG})) \longrightarrow_p 0.$$

Proof

We can rewrite $v_L^*(\bar{y}_{N,GREG})$ as

$$\begin{aligned} v_L^*(\bar{y}_{N,GREG}) &= d^2 \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 \\ &= \left(\frac{1}{p}\right)^2 \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 + \\ &\quad \left(d - \frac{1}{p}\right)^2 \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 + \\ &\quad 2\frac{1}{p} \left(d - \frac{1}{p}\right) \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 \\ &= \left(\frac{1}{p}\right)^2 v_1 + \left(d - \frac{1}{p}\right)^2 v_1 + 2\frac{1}{p} \left(d - \frac{1}{p}\right) v_1 \end{aligned}$$

Following the proof of Theorem 4.3, we rewrite v_1 as

$$\begin{aligned} v_1 &= \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - E(e_{hi}))^2 - \sum_h \frac{n_h^2}{n_h - 1} (\bar{e}_h - E(e_{hi}))^2 \\ &= v_{11} - v_{12}. \end{aligned}$$

Under conditions C10 and C11, we have

$$n_T v_{11} \longrightarrow_p n_T \sum_h \frac{n_h^2}{n_h - 1} \sigma_{eh}^2$$

and

$$n_T v_{12} \longrightarrow_p n_T \sum_h \frac{n_h}{n_h - 1} \sigma_{eh}^2$$

where $\sigma_{eh}^2 = V(e_{hi})$. Combining these two results gives

$$n_T v_1 \longrightarrow_p n_T \sigma_{er}^2.$$

Also, under C4, see (4.2.6),

$$d \longrightarrow_p \frac{1}{p}.$$

Thus,

$$\begin{aligned} n_T v_L^*(\bar{y}_{N,GREG}) &= d^2 n_T v_1 \\ &= \left(\frac{1}{p}\right)^2 n_T v_1 + \left(d - \frac{1}{p}\right)^2 n_T v_1 + 2\frac{1}{p} \left(d - \frac{1}{p}\right) n_T v_1 \\ &\longrightarrow_p n_T \frac{\sigma_{er}^2}{p^2} \approx n_T V(\bar{y}_{N,GREG}). \end{aligned}$$

■

To construct a jackknife variance estimator, we first define the nonresponse adjustment when the $(gj)^{th}$ cluster has been deleted. Let,

$$d_{(gj)} = \frac{\hat{U}_{(gj)}}{\hat{R}_{(gj)}}$$

where $\hat{U}_{(gj)} = \sum_{(hik) \in s} \hat{w}_{hik(gj)} a_{hik} \mathbf{x}_{hik}^T \mathbf{y}_{hik}$ and $\hat{w}_{hik(gj)}$ are the normalized jackknife weights. Also, let

$$\hat{\mathbf{A}}_{(gj)} = \sum_{(hik) \in s} d_{(gj)} \hat{w}_{hik(gj)} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T$$

and

$$\hat{\mathbf{b}}_{(gj)} = \sum_{(hik) \in s} d_{(gj)} \hat{w}_{hik(gj)} a_{hik} \mathbf{x}_{hik} \mathbf{y}_{hik}.$$

The regression estimator with the $(gj)^{th}$ cluster deleted is

$$\bar{y}_{N,GREG(gj)} = \hat{\beta}_{(gj)}^T \bar{\mathbf{X}}$$

where $\hat{\beta}_{(g)} = \tilde{\mathbf{A}}_{(g)}^{-1} \hat{\mathbf{b}}_{(g)}$. The jackknife variance estimator is then defined as

$$v_J(\bar{y}_{N,GREG}) = \sum_h \frac{n_h - 1}{n_h} \sum_j (\bar{y}_{N,GREG(g)} - \bar{y}_{N,GREG})^2.$$

Asymptotic consistency of $v_J(\bar{y}_{N,GREG})$ is established in the following theorem.

Theorem 4.7 *Under condition C4, C10, C11, C12 and the uniform response mechanism,*

$$n_T(v_J(\bar{y}_{N,GREG}) - V(\bar{y}_{N,GREG})) \longrightarrow_p 0.$$

Proof

Noting that $\bar{y}_{N,GREG(g)} - \bar{y}_{N,GREG} = (\hat{\beta}_{(g)} - \hat{\beta})^T \bar{\mathbf{X}}$, we wish to evaluate the difference $\hat{\beta}_{(g)} - \hat{\beta}$. First we need to approximate the nonresponse adjustment, $d_{(g)}$, in a Taylor series expansion around the full sample adjustment, d ,

$$\begin{aligned} d_{(g)} &= \frac{\hat{U}_{(g)}}{\hat{R}_{(g)}} \\ &= d + \frac{1}{\hat{R}} (\hat{U}_{(g)} - \hat{U} - d(\hat{R}_{(g)} - \hat{R})) \\ &= d + \Delta_{gJ} \quad \text{say.} \end{aligned}$$

Using this expression in $\hat{\mathbf{A}}_{(g)}$ and $\hat{\mathbf{b}}_{(g)}$ and ignoring lower order terms gives

$$\begin{aligned} \hat{\mathbf{A}}_{(g)} &= \sum_{(hik)cs} d_{(g)} \tilde{w}_{hik(g)} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T \\ &= \hat{\mathbf{A}} + d\mathbf{D}_{gJ} + \Delta_{gJ} \mathbf{A}^* \end{aligned} \tag{4.3.8}$$

and

$$\begin{aligned} \hat{\mathbf{b}}_{(g)} &= \sum_{(hik)cs} d_{(g)} \hat{w}_{hik(g)} a_{hik} \mathbf{x}_{hik} y_{hik} \\ &= \hat{\mathbf{b}} + d\mathbf{c}_{gJ} + \Delta_{gJ} \mathbf{b}^* \end{aligned}$$

where

$$\begin{aligned}
 \mathbf{D}_{gj} &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} a_{gik} \mathbf{x}_{gik} \mathbf{x}_{gik}^T - \sum_{(gjk) \in s} \hat{w}_{gjk} a_{gjk} \mathbf{x}_{gjk} \mathbf{x}_{gjk}^T \right) \\
 \mathbf{c}_{gj} &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} a_{gik} \mathbf{x}_{gik} y_{gik} - \sum_{(gjk) \in s} \hat{w}_{gjk} a_{gjk} \mathbf{x}_{gjk} y_{gjk} \right) \\
 \mathbf{A}^* &= \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T \\
 &= \frac{1}{d} \hat{\mathbf{A}}
 \end{aligned}$$

and

$$\begin{aligned}
 \mathbf{b}^* &= \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} \mathbf{x}_{hik} y_{hik} \\
 &= \frac{1}{d} \hat{\mathbf{b}}.
 \end{aligned}$$

Applying the matrix equality (3.3.9) to equation (4.3.8) and ignoring lower order terms, we obtain

$$\hat{\mathbf{A}}_{(gj)}^{-1} \cong \hat{\mathbf{A}}^{-1} - \hat{\mathbf{A}}^{-1} (d\mathbf{D}_{gj} + \Delta_{gj} \mathbf{A}^*) \hat{\mathbf{A}}^{-1}.$$

Using this expression in $\hat{\beta}_{(gj)}$, we have

$$\begin{aligned}
 \hat{\beta}_{(gj)} &= \hat{\mathbf{A}}_{(gj)}^{-1} \hat{\mathbf{b}}_{(gj)} \\
 &\cong \left(\hat{\mathbf{A}}_{(gj)}^{-1} - \hat{\mathbf{A}}^{-1} - \hat{\mathbf{A}}^{-1} (d\mathbf{D}_{gj} + \Delta_{gj} \mathbf{A}^*) \hat{\mathbf{A}}^{-1} \right) (\hat{\mathbf{b}} + d\mathbf{c}_{gj} + \Delta_{gj} \mathbf{b}^*) \\
 &= \hat{\beta} + \hat{\mathbf{A}}^{-1} d(\mathbf{c}_{gj} - \mathbf{D}_{gj} \beta) + \Delta_{gj} \hat{\mathbf{A}}^{-1} (\mathbf{b}^* - \mathbf{A}^* \hat{\beta}) + \text{lower order terms.}
 \end{aligned}$$

Ignoring the lower order terms, we can now rewrite $\hat{y}_{N,GREG(gj)} - \bar{y}_{N,GREG}$ as,

$$\begin{aligned}
 \hat{y}_{N,GREG(gj)} - \bar{y}_{N,GREG} &= (\hat{\beta}_{(gj)} - \hat{\beta})^T \bar{\mathbf{X}} \\
 &\cong d(\mathbf{c}_{gj} - \mathbf{D}_{gj} \beta)^T \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}} + \Delta_{gj} (\mathbf{b}^* - \mathbf{A}^* \hat{\beta})^T \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}}. \quad (4.3.9)
 \end{aligned}$$

Noting that $\mathbf{b}^* = (1/d)\tilde{\mathbf{b}}$ and $\mathbf{A}^* = (1/d)\tilde{\mathbf{A}}$, the last term in equation (4.3.9) can be simplified to

$$\begin{aligned}\Delta_{gj}(\mathbf{b}^* - \mathbf{A}^*\hat{\beta})^T \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{X}} &= \Delta_{gj} \left(\frac{1}{d} \hat{\beta}^T \tilde{\mathbf{X}} - \frac{1}{d} \hat{\beta}^T \tilde{\mathbf{X}} \right) \\ &= 0.\end{aligned}$$

Therefore we have

$$\begin{aligned}\tilde{y}_{N,GREG(gj)} - \bar{y}_{N,GREG} &\cong d(\mathbf{c}_{gj} - \mathbf{D}_{gj}\beta)^T \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{X}} \\ &= d \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gjk) \in s} \hat{w}_{gjk} a_{gjk} e_{gjk}^* \mathbf{x}_{gjk}^T \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{X}} - \sum_{(gjk) \in s} \hat{w}_{gjk} a_{gjk} e_{gjk}^* \mathbf{x}_{gjk}^T \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{X}} \right) \\ &= d \frac{n_g}{n_g - 1} (\hat{e}_g^* - \hat{e}_{gj}^*)\end{aligned}$$

where $\hat{e}_{hi}^* = \sum_k \hat{w}_{hik}^* a_{hik} e_{hik}^*$, and $\hat{w}_{hik}^* = \hat{w}_{hik} \mathbf{x}_{hik}^T \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{X}}$. Putting this expression for $\tilde{y}_{N,GREG(gj)} - \bar{y}_{N,GREG}$ into the jackknife variance estimator gives

$$\begin{aligned}v_J(\tilde{y}_{N,GREG}) &= \sum_g \frac{n_g - 1}{n_g} \sum_j (\tilde{y}_{N,GREG(gj)} - \bar{y}_{N,GREG})^2 \\ &\cong \sum_g \frac{n_g}{n_g - 1} \sum_j (d(\hat{e}_g^* - \hat{e}_{gj}^*))^2 \\ &= v_{JL}(\tilde{y}_{N,GREG}) \\ &= v(r_{hi})\end{aligned}$$

where

$$\begin{aligned}r_{hi} &= d \sum_k \hat{w}_{hik}^* a_{hik} e_{hik}^* \\ &= \sum_k \hat{w}_{hik} a_{hik} e_{hik}^* \mathbf{x}_{hik}^T \mathbf{A}^{*-1} \tilde{\mathbf{X}}.\end{aligned}$$

This linearization variance estimator, $v_{JL}(\tilde{y}_{N,GREG})$, depends only on the cluster totals of the estimated residuals weighted by the design weights adjusted for generalized regression estimation. This linearization variance estimator differs from

$v_L(\bar{y}_{N,GREG})$ in that the weight has been adjusted for the regression estimation method. This is analogous to the difference between the customary linearization variance estimators and the robust variance estimators presented earlier for the one-way poststratified estimator.

To establish consistency of $v_{JL}(\bar{y}_{N,GREG})$, and thereby for $v_J(\bar{y}_{N,GREG})$, note that

$$\begin{aligned} r_{hi} &= d \sum_k \hat{w}_{hik} a_{hik} \epsilon_{hik}^* + d \sum_k \hat{w}_{hik} a_{hik} \epsilon_{hik} (\mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \mathbf{X} - 1) \\ &\approx d \epsilon_{hi}^* \end{aligned}$$

since $\mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \mathbf{X} \rightarrow_p 1$. To show this, we use the fact that the vector $\mathbf{1}$ is in the column space of $\mathbf{X}_d$. This condition is equivalent to the existence of a vector $\mathbf{c}$ such that $\mathbf{x}_{hik}^T \mathbf{c} = 1$. Assuming the existence of such a $\mathbf{c}$, we have

$$\begin{aligned} \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \mathbf{X} &= \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \mathbf{X} \\ &= \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \sum_u \mathbf{x}_{hik} \mathbf{x}_{hik}^T \mathbf{c} \\ &= \mathbf{x}_{hik}^T \mathbf{c} = 1 \end{aligned}$$

since $\hat{\mathbf{A}}^{-1} \sum_u \mathbf{x}_{hik} \mathbf{x}_{hik}^T \rightarrow_p \mathbf{I}$ (Särndal, 1989), where $\sum_u$ denote summing over the population and $\mathbf{X}$ is the vector of population totals. We can now rewrite $v_{JL}(\bar{y}_{N,GREG})$ as

$$v_{JL}(\bar{y}_{N,GREG}) \approx v_L(\bar{y}_{N,GREG}).$$

In Theorem 4.6, we showed that

$$n_T(v_L(\bar{y}_{N,GREG}) - V(\bar{y}_{N,GREG})) \rightarrow_p 0.$$

Therefore, $v_{JL}(\bar{y}_{GREG})$ is consistent for $V(\bar{y}_{GREG})$ along with $v_J(\bar{y}_{GREG})$. ■

4.3.2 Multiple Imputation Classes

Suppose we have $J \geq 2$ imputation classes with a uniform response mechanism within each imputation class. The nonresponse adjustment factor for the j^{th} impu-

tation class is

$$d_j = \frac{\sum_{(hik) \in s_j} \hat{w}_{hik}}{\sum_{(hik) \in s_j} \hat{w}_{hik} a_{hik}} = \frac{\hat{U}_j}{\hat{R}_j},$$

and the regression adjustment factor is

$$g_{hik} = \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} \bar{\mathbf{X}}$$

where

$$\hat{\mathbf{A}} = \sum_j \sum_{(hik) \in s_j} d_j \hat{w}_{hik} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T.$$

The regression estimator is given by

$$\begin{aligned} \bar{y}_{N,GREG} &= \sum_j \sum_{(hik) \in s_j} d_j \hat{w}_{hik} g_{hik} a_{hik} y_{hik} \\ &= \hat{\beta}^T \bar{\mathbf{X}} \end{aligned}$$

where

$$\hat{\beta} = \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}}$$

and

$$\hat{\mathbf{b}} = \sum_j \sum_{(hik) \in s_j} d_j \hat{w}_{hik} a_{hik} y_{hik} \mathbf{x}_{hik}.$$

It is easy to show that $\bar{\mathbf{x}}_{N,GREG} = \bar{\mathbf{X}}$, showing consistency with the known population totals. Under a uniform response mechanism in each imputation class, it follows that $\bar{y}_{N,GREG}$ is approximately unbiased for $\bar{Y}$,

$$\begin{aligned} E(\bar{y}_{N,GREG}) &\approx \sum_j \sum_{(hik) \in s_j} \frac{1}{p_j} p_j E \left[\hat{w}_{hik} y_{hik} \mathbf{x}_{hik}^T \left(\sum_{(hik) \in s_j} \hat{w}_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T \right)^{-1} \bar{\mathbf{X}} \right] \\ &= E(\bar{y}_{GREG}) = \bar{Y}. \end{aligned}$$

To construct a jackknife variance estimator of $\bar{y}_{N,GREG}$, we first define the following terms when the $(gl)^{th}$ cluster has been deleted,

$$\begin{aligned} d_{j(gl)} &= \frac{\hat{U}_{j(gl)}}{\hat{R}_{j(gl)}} \\ &= \frac{\sum_{(hik) \in s_j} \hat{w}_{hik(gl)}}{\sum_{(hik) \in s_j} \hat{w}_{hik(gl)} a_{hik}} \end{aligned}$$

and

$$\hat{\mathbf{A}}_{(gl)} = \sum_j \sum_{(hik) \in s_j} d_{j(gl)} \hat{w}_{hik(gl)} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T$$

where $\hat{w}_{hik(gl)}$ are the normalized jackknife weights. The generalized regression estimator, with the $(gl)^{th}$ cluster deleted, is

$$\begin{aligned} \bar{y}_{N,GREG(gl)} &= \sum_j \sum_{(hik) \in s_j} d_{j(gl)} \hat{w}_{hik(gl)}^* a_{hik} y_{hik} \\ &= \hat{\beta}_{(gl)}^T \bar{\mathbf{X}} \end{aligned}$$

where $\hat{w}_{hik(gl)}^* = \hat{w}_{hik(gl)} \mathbf{x}_{hik}^T \hat{\mathbf{A}}_{(gl)}^{-1} \bar{\mathbf{X}}$.

$$\hat{\beta}_{(gl)} = \hat{\mathbf{A}}_{(gl)}^{-1} \hat{\mathbf{b}}_{(gl)}$$

and

$$\hat{\mathbf{b}}_{(gl)} = \sum_j \sum_{(hik) \in s_j} d_{j(gl)} \hat{w}_{hik(gl)} a_{hik} y_{hik} \mathbf{x}_{hik}.$$

The jackknife variance estimator is then defined as

$$v_J(\bar{y}_{N,GREG}) = \sum_g \frac{n_g - 1}{n_g} \sum_l (\bar{y}_{N,GREG(gl)} - \bar{y}_{N,GREG})^2.$$

Consistency of $v_J(\bar{y}_{N,GREG})$ can be established along the same lines as Theorem 4.7.

To obtain a linearization variance estimator, noting that

$$\bar{y}_{N,GREG(gl)} - \bar{y}_{N,GREG} = (\hat{\beta}_{(gl)} - \hat{\beta})^T \bar{\mathbf{X}}$$

we wish to approximate the difference $\hat{\beta}_{(gl)} - \hat{\beta}$. Using a Taylor approximation to $d_{j(gl)}$ we have,

$$\begin{aligned} d_{j(gl)} &= \frac{\hat{U}_{j(gl)}}{\hat{R}_{j(gl)}} \\ &\cong d_j + \frac{1}{\bar{R}} (\hat{U}_{j(gl)} - \bar{U}_j - d_j (\hat{R}_{j(gl)} - \bar{R}_j)) \\ &= d_j + \Delta_{jgl} \quad \text{say.} \end{aligned}$$

Using this expression in $\tilde{\mathbf{A}}_{(gl)}$ and $\tilde{\mathbf{b}}_{(gl)}$ and ignoring lower order terms we have,

$$\begin{aligned}\tilde{\mathbf{A}}_{(gl)} &\cong \sum_j \sum_{(hik) \in s_j} (d_j + \Delta_{jgl}) \tilde{w}_{hik(gl)} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T \\ &= \tilde{\mathbf{A}} + \sum_j d_j \mathbf{D}_{jgl} + \sum_j \Delta_{jgl} \mathbf{A}_j^* + \text{lower order terms}\end{aligned}$$

and

$$\begin{aligned}\tilde{\mathbf{b}}_{(gl)} &\cong \sum_j \sum_{(hik) \in s_j} (d_j + \Delta_{jgl}) \tilde{w}_{hik(gl)} a_{hik} y_{hik} \mathbf{x}_{hik} \\ &= \tilde{\mathbf{b}} + \sum_j d_j \mathbf{c}_{jgl} + \sum_j \Delta_{jgl} \mathbf{b}_j^* + \text{lower order terms}\end{aligned}$$

where

$$\begin{aligned}\mathbf{D}_{jgl} &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s_j} \tilde{w}_{gik} a_{gik} \mathbf{x}_{gik} \mathbf{x}_{gik}^T - \sum_{(gik) \in s_j} \tilde{w}_{gik} a_{gik} \mathbf{x}_{gik} \mathbf{x}_{gik}^T \right) \\ \mathbf{c}_{jgl} &= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s_j} \tilde{w}_{gik} a_{gik} y_{gik} \mathbf{x}_{gik} - \sum_{(gik) \in s_j} \tilde{w}_{gik} a_{gik} y_{gik} \mathbf{x}_{gik} \right) \\ \mathbf{A}_j^* &= \sum_{(hik) \in s_j} \tilde{w}_{hik} a_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T\end{aligned}$$

and

$$\mathbf{b}_j^* = \sum_{(hik) \in s_j} \tilde{w}_{hik} a_{hik} y_{hik} \mathbf{x}_{hik}.$$

Using the matrix equality (3.3.9), we have

$$\tilde{\mathbf{A}}_{(gl)}^{-1} \cong \tilde{\mathbf{A}}^{-1} - \tilde{\mathbf{A}}^{-1} \left(\sum_j d_j \mathbf{D}_{jgl} + \sum_j \Delta_{jgl} \mathbf{A}_j^* \right) \tilde{\mathbf{A}}^{-1}$$

and

$$\begin{aligned}\hat{\beta}_{(gl)} &= \tilde{\mathbf{A}}_{(gl)}^{-1} \tilde{\mathbf{b}}_{(gl)} \\ &\cong \left[\tilde{\mathbf{A}}^{-1} - \tilde{\mathbf{A}}^{-1} \left(\sum_j d_j \mathbf{D}_{jgl} + \sum_j \Delta_{jgl} \mathbf{A}_j^* \right) \tilde{\mathbf{A}}^{-1} \right] \times \\ &\quad \left(\tilde{\mathbf{b}} + \sum_j d_j \mathbf{c}_{jgl} + \sum_j \Delta_{jgl} \mathbf{b}_j^* \right)\end{aligned}$$

$$\begin{aligned}
&= \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}} + \hat{\mathbf{A}}^{-1} \sum_j d_j \mathbf{c}_{jgl} + \hat{\mathbf{A}}^{-1} \sum_j \Delta_{jgl} \mathbf{b}_j^* - \hat{\mathbf{A}}^{-1} \sum_j d_j \mathbf{D}_{jgl} \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}} - \\
&\quad \hat{\mathbf{A}}^{-1} \sum_j d_j \mathbf{D}_{jgl} \hat{\mathbf{A}}^{-1} \sum_j d_j \mathbf{c}_{jgl} - \hat{\mathbf{A}}^{-1} \sum_j d_j \mathbf{D}_{jgl} \hat{\mathbf{A}}^{-1} \sum_j \Delta_{jgl} \mathbf{b}_j^* - \\
&\quad \hat{\mathbf{A}}^{-1} \sum_j \Delta_{jgl} \mathbf{A}_j^* \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}} - \hat{\mathbf{A}}^{-1} \sum_j \Delta_{jgl} \mathbf{A}_j^* \hat{\mathbf{A}}^{-1} \sum_j d_j \mathbf{c}_{jgl} - \\
&\quad \hat{\mathbf{A}}^{-1} \sum_j \Delta_{jgl} \mathbf{A}_j^* \hat{\mathbf{A}}^{-1} \sum_j \Delta_{jgl} \mathbf{b}_j^* \\
&= \hat{\beta} + \hat{\mathbf{A}}^{-1} \sum_j d_j (\mathbf{c}_{jgl} - \mathbf{D}_{jgl} \hat{\beta}) + \hat{\mathbf{A}}^{-1} \sum_j \Delta_{jgl} (\mathbf{b}_j^* - \mathbf{A}_j^* \hat{\beta}) + \\
&\quad \text{lower order terms.}
\end{aligned}$$

Note that

$$\mathbf{c}_{jgl} - \mathbf{D}_{jgl} \hat{\beta} = \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s_j} \hat{w}_{gik} a_{gik} e_{gik} \mathbf{x}_{gik}^T - \sum_{(gik) \in s_j} \hat{w}_{gik} a_{gik} e_{gik} \mathbf{x}_{gik}^T \right) \quad (4.3.10)$$

and

$$\mathbf{b}_j^* - \mathbf{A}_j^* \hat{\beta} = \sum_{(hik) \in s_j} \hat{w}_{hik} a_{hik} e_{hik} \mathbf{x}_{hik}^T. \quad (4.3.11)$$

Therefore, the jackknife variance estimator can be approximated by

$$\begin{aligned}
v_{JL}(\tilde{y}_{N,GREG}) &= \sum_g \frac{n_g}{n_g - 1} \sum_l \left[\sum_j d_j (\mathbf{c}_{jgl} - \mathbf{D}_{jgl} \hat{\beta})^T \hat{\mathbf{A}}^{-1} \mathbf{X} + \right. \\
&\quad \left. \sum_j \Delta_{jgl} (\mathbf{b}_j^* - \mathbf{A}_j^* \hat{\beta})^T \hat{\mathbf{A}}^{-1} \mathbf{X} \right]^2 \\
&= v(r_{hi})
\end{aligned}$$

where, using equations (4.3.10) and (4.3.11),

$$r_{hi} = \sum_j d_j \sum_k \hat{w}_{hik}^* a_{hik} e_{hik}^* \eta_{hik}^j + \sum_j \tilde{e}_j^* \sum_k \hat{w}_{hik} (1 - d_j a_{hik}) \eta_{hik}^j$$

with $\tilde{e}_j^* = \sum_{(hik) \in s_j} \hat{w}_{hik}^* a_{hik} e_{hik}^*$. That is, by linearizing the jackknife variance estimator, we obtain a linearization type variance estimator which uses the correctly weighted residuals and includes an additional term due to the nonresponse adjustment.

4.4 Simulation Study

We performed a simulation study to evaluate the effectiveness of the jackknife and the linearized jackknife variance estimators for the poststratified estimator under a weighting adjustment for unit nonresponse. The 1988 CPS sample described in Section 3.4 was used as a finite population with eight poststrata defined on the basis of age, race and sex. Due to the lack of auxiliary information, only the single weighting class case is considered in the simulation study.

Using the CPS population two stratified populations were created with 50 design strata and 100 design strata respectively. In each population, the strata were created to have approximately the same total number of households. Within each stratum, we selected $n_h = 2$ segments with probability proportional to size. A simple random sample of size 4 was selected without replacement from each sampled segment with all persons selected if the segment contained fewer than 4 individuals.

To simulate the nonresponse, for each sampled individual a Bernoulli trial was performed with probability p for the outcome "response" where $p = 0.9, 0.8$ or 0.7 . Once the nonrespondents were identified, a weighting adjustment was calculated as outlined in Section 4.2.1. Two independent sets of 10,000 samples were selected, one set for the 50 strata design and another set for the 100 strata design.

In each sample we computed the poststratified estimator, the jackknife, v_J , the linearized jackknife, v_{JL} , a "naive" jackknife, v_{JN} , and the corresponding error rates associated with the normal theory confidence intervals. In the weighting adjustment case, the naive jackknife variance estimator fails to recalculate the nonresponse adjustment factor each time a cluster is deleted, but in the special case of the poststratified estimator and a single weighting class, the "naive" jackknife is equivalent to the correct jackknife. Relative bias, error rates and average lengths of the confidence intervals were calculated as detailed in Section 3.4.

Table 4.1 presents results, averaged over 10,000 samples, for the 50 strata design.

Table 4.1: Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Weighting Adjustment for Unit Nonresponse (50 Strata Design)

Performance Measure	$v_J(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_{JN}(\hat{Y}_{PS})$
$p = 0.9$			
Relative Bias	-2.42 %	-3.15 %	-2.42 %
Error Rate	6.04 %	6.10 %	6.04 %
Lower Error Rate	2.40 %	2.41 %	2.40 %
Upper Error Rate	3.64 %	3.69 %	3.64 %
Average Length	5.68	5.66	5.68
$p = 0.8$			
Relative Bias	-2.75 %	-3.62 %	-2.75 %
Error Rate	6.17 %	6.36 %	6.17 %
Lower Error Rate	2.67 %	2.71 %	2.67 %
Upper Error Rate	3.50 %	3.59 %	3.50 %
Average Length	5.98	5.95	5.98
$p = 0.7$			
Relative Bias	-3.04 %	-4.16 %	-3.04 %
Error Rate	6.05 %	6.21 %	6.05 %
Lower Error Rate	2.63 %	2.69 %	2.63 %
Upper Error Rate	3.42 %	3.52 %	3.42 %
Average Length	6.30	6.27	6.30

As expected, v_J and v_{JN} are identical while v_{JL} differs slightly from v_J and v_{JN} . In general, all three variance estimators appear to be performing well with relative biases less than 5% and error rates slightly above the nominal 5%. Comparing the average lengths of the confidence intervals, we see little differences due to the asymptotic equivalence of v_J and v_{JL} .

Results from the 100 strata design appear in Table 4.2. Again v_J and v_{JN} are identical and all three variance estimators perform well. We note that the asymptotic equivalence between v_J and v_{JL} is more evident here than in the 50 strata design as indicated by the similarity of the average lengths of the confidence intervals.

Table 4.2: Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Weighting Adjustments for Unit Nonresponse (100 Strata Design)

Performance Measure	$v_J(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_{JN}(\hat{Y}_{PS})$
<u>$p = 0.9$</u>			
Relative Bias	-3.64 %	-3.80 %	-3.64 %
Error Rate	6.11 %	6.09 %	6.11 %
Lower Error Rate	2.68 %	2.66 %	2.68 %
Upper Error Rate	3.43 %	3.43 %	3.43 %
Average Length	3.99	3.99	3.99
<u>$p = 0.8$</u>			
Relative Bias	-4.61 %	-4.61 %	-4.61 %
Error Rate	6.17 %	6.16 %	6.17 %
Lower Error Rate	2.72 %	2.71 %	2.72 %
Upper Error Rate	3.45 %	3.45 %	3.45 %
Average Length	4.20	4.19	4.20
<u>$p = 0.7$</u>			
Relative Bias	-2.40 %	-2.65 %	-2.40 %
Error Rate	5.61 %	5.63 %	5.61 %
Lower Error Rate	2.50 %	2.50 %	2.50 %
Upper Error Rate	3.11 %	3.13 %	3.11 %
Average Length	4.43	4.43	4.43

Chapter 5

Deterministic Imputation

5.1 Introduction

Item nonresponse is usually handled by some form of imputation. Although there are many ways of determining imputed values, there are two broad categories of imputation: (a) deterministic imputation which includes mean, ratio, regression and nearest neighbor imputation. The imputed values are deterministic given the sample of respondents and any auxiliary information on the nonrespondents. (b) Hot deck imputation which employs a sample drawn from the respondent values. In the simplest form of hot deck imputation, a simple random sample is drawn with replacement from the sample respondents to an item y and the associated item values are used as donors. Regardless of the imputation method, it is common to treat the imputed values as if they are true values, and then compute the variance estimates using standard formulae. This procedure, however, can lead to serious underestimation of the true variance of the estimates when the proportion of missing values for the item of interest is appreciable.

In order to correctly estimate the variance of an estimator under imputation, Rao and Shao (1992) proposed a new jackknife variance estimation method for a particular hot deck imputation procedure based on adjusted imputed values. Assuming

a uniform response mechanism, this new method leads to design consistent variance estimators. Using the same method, Rao and Sitter (1992) obtained “correct” design based variance estimators for the expansion estimator under several deterministic imputation procedures for simple random sampling. They also presented linearized versions of the proposed jackknife variance estimators obtained by Taylor approximations to the jackknife variance estimator. These resulting linearization variance estimators retain desirable properties of the jackknife, unlike some other linearization variance estimators.

In this chapter, the work of Rao and Sitter (1992) will be extended to the poststratified estimator under mean imputation for a stratified multi-stage survey design. The poststratified estimator under hot deck imputation will be considered in Chapter 6. When performing imputation with the poststratified estimator, one has the choice of performing the imputation with the design weights and then performing the poststratification adjustment or to perform the poststratification adjustment first and use the poststratified weights in the imputation procedure. By using the poststratified weights for imputation, the resulting estimator may be more efficient than if the design weights were used during the imputation procedure. Only the second case, where the poststratified weights are used in the imputation procedure, will be presented in this chapter. Section 5.2 will consider the poststratified estimator under mean imputation with one imputation class and with multiple imputation classes. Asymptotic normality of the poststratified estimator will be established under the one imputation class only, for simplicity. We present correct jackknife variance estimators and linearization variance estimators obtained by linearizing the jackknife variance estimator. Asymptotic consistency of the jackknife variance estimator and the linearization variance estimator will be shown in the one imputation class only. Results from a simulation study investigating the finite sample properties of the jackknife and the linearized jackknife variance estimators under mean imputation will be presented in Section 5.3.

5.2 Mean Imputation

Suppose we have a stratified multi-stage design with the sample, s , divided into respondents, s_r , and nonrespondents, s_m . Also, suppose there are C poststrata with ${}_c s_r$ denoting the respondents in the c^{th} poststratum and ${}_c s_m$ denoting the nonrespondents in the c^{th} poststratum. Throughout this thesis we will assume that an individual's poststratum is known regardless of whether the individual responds or not.

5.2.1 One Imputation Class

Given that we know which poststratum each individual belongs to, we can use the usual poststratified weight

$${}_c \tilde{w}_{hik} = \frac{{}_c M}{M} \frac{\tilde{w}_{hik}}{{}_c \hat{M}}$$

where ${}_c \hat{M} = \sum_{(hik) \in s} \tilde{w}_{hik} {}_c \delta_{hik}$, ${}_c M$ is the known poststratum total, $\tilde{w}_{hik}$ is the normalized design weight, and ${}_c \delta_{hik}$ is the poststratum indicator variable. Now, for each element in s_m we will impute the mean value. That is for $(hik) \in s_m$, we impute

$$y_{hik}^* = \frac{\hat{S}}{\hat{T}}$$

where

$$\begin{aligned} \hat{S} &= \sum_c \sum_{(hik) \in s} {}_c \tilde{w}_{hik} a_{hik} y_{hik} {}_c \delta_{hik} \\ \hat{T} &= \sum_c \sum_{(hik) \in s} {}_c \tilde{w}_{hik} a_{hik} {}_c \delta_{hik} \end{aligned}$$

and a_{hik} is the response indicator variable. Note that in the definition of $\hat{S}$ and $\hat{T}$, the poststratified adjusted weight, ${}_c \tilde{w}_{hik}$, is used. That is, the poststratification adjustment is performed before the imputation procedure.

The poststratified estimator of the mean, $\bar{Y}$, under mean imputation is

$$\begin{aligned}\bar{y}_{PS}^{MI} &= \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} a_{hik} y_{hik} c \delta_{hik} + \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} (1 - a_{hik}) y_{hik}^* c \delta_{hik} \quad (5.2.1) \\ &= \frac{\hat{S}}{\hat{T}} \hat{U}\end{aligned}$$

where

$$\hat{U} = \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} c \delta_{hik}.$$

Under the assumption that each individual's poststratum is known, the imputed poststratified estimator will still benchmark to the known poststrata totals. To see this, substitute

$$d\delta_{hik} = \begin{cases} 1 & \text{if the } (hik)^{th} \text{ unit is in the } d^{th} \text{ poststratum} \\ 0 & \text{otherwise} \end{cases}$$

for y_{hik} and y_{hik}^* in equation (5.2.1) This gives

$$\begin{aligned}{}_d\hat{M}_{PS}^{MI} &= \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} a_{hik} d\delta_{hik} c \delta_{hik} + \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} (1 - a_{hik}) d\delta_{hik} c \delta_{hik} \\ &= \sum_{(hik) \in s} d \hat{w}_{hik} a_{hik} d\delta_{hik} + \sum_{(hik) \in s} d \hat{w}_{hik} (1 - a_{hik}) d\delta_{hik} \\ &= \sum_{(hik) \in s} d \hat{w}_{hik} d\delta_{hik} \\ &= \frac{{}_dM}{M} \frac{\hat{M}}{\hat{M}} = \frac{{}_dM}{M}\end{aligned}$$

which is the normalized poststratum total. From this, it follows that $\hat{U} = 1$ since

$$\begin{aligned}\hat{U} &= \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} c \delta_{hik} \\ &= \sum_c \frac{{}_cM}{M} \\ &= 1.\end{aligned}$$

Further more, we have

$$\begin{aligned}\bar{y}_{PS}^{MI} &= \frac{\hat{S}}{\hat{T}} \bar{U} \\ &= \frac{\hat{S}}{\hat{T}}.\end{aligned}$$

Assuming a uniform response mechanism, that is $E(a_{hik}) = p > 0$ and $\{a_{hik}\}$ are independent of each other and of y_{hik} , $\bar{y}_{PS}^{MI}$ is an approximately unbiased estimator of the mean since

$$\begin{aligned}E(\hat{S}) &= pE\left(\sum_c \sum_{(hik) \in s} c \hat{w}_{hik} y_{hik} c \delta_{hik}\right) \\ &= p\bar{Y},\end{aligned}$$

and

$$\begin{aligned}E(\hat{T}) &= pE\left(\sum_c \sum_{(hik) \in s} c \hat{w}_{hik} c \delta_{hik}\right) \\ &= p.\end{aligned}$$

Using a ratio approximation, we have

$$E(\bar{y}_{PS}^{MI}) = E\left(\frac{\hat{S}}{\hat{T}}\right) \approx \frac{p\bar{Y}}{p} = \bar{Y}.$$

To investigate the large sample properties of $\bar{y}_{PS}^{MI}$, note that

$$\begin{aligned}\bar{y}_{PS}^{MI} - \bar{Y} &= \frac{\hat{S}}{\hat{T}} - \bar{Y} \\ &= \frac{1}{\hat{T}}(\hat{S} - \bar{Y}\hat{T}) \\ &= \frac{1}{\hat{T}} \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} a_{hik} c \delta_{hik} (y_{hik} - \bar{Y}) \\ &= \frac{1}{\hat{T}} \hat{R}_{PS}\end{aligned}\tag{5.2.2}$$

where $\hat{R}_{PS}$ is the poststratified estimator of the mean of the variable $r_{hik} = y_{hik} - \bar{Y}$. Further, $\bar{Y} = \sum_c (cM/M) \bar{Y}$ implies that $\sum_c (cM/M)(c\bar{Y} - \bar{Y}) = 0$. Therefore we can subtract

$$\frac{p}{\bar{T}} \sum_c \frac{cM}{M} (c\bar{Y} - \bar{Y})$$

from $\bar{y}_{PS}^{MI}$ giving

$$\begin{aligned} \bar{y}_{PS}^{MI} - \bar{Y} &= \frac{1}{\bar{T}} \hat{R}_{PS} - \frac{p}{\bar{T}} \sum_c \frac{cM}{M} (c\bar{Y} - \bar{Y}) \\ &= \frac{1}{\bar{T}} \left(\sum_c \frac{cM}{M} \frac{1}{c\bar{M}} ({}_c\hat{R}^r - p({}_c\bar{Y} - \bar{Y})({}_c\hat{M})) \right) \\ &= \frac{1}{\bar{T}} \sum_c \frac{cM}{M} \frac{1}{c\bar{M}} \sum_{(hik) \in s} \hat{w}_{hik} {}_c\delta_{hik} (a_{hik}(y_{hik} - \bar{Y}) - p({}_c\bar{Y} - \bar{Y})) \\ &= \frac{1}{\bar{T}} \sum_c \frac{cM}{c\bar{M}} {}_c\bar{e} \\ &= \frac{1}{\bar{T}} \bar{e} + \frac{1}{\bar{T}} \sum_c \left(\frac{cM}{c\bar{M}} - 1 \right) {}_c\bar{e} \end{aligned} \quad (5.2.3)$$

where

$$\begin{aligned} {}_c\bar{e} &= \sum_{(hik) \in s} \hat{w}_{hik} {}_c\delta_{hik} {}_c\epsilon_{hik}, \\ {}_c\epsilon_{hik} &= a_{hik}(y_{hik} - \bar{Y}) - p({}_c\bar{Y} - \bar{Y}), \\ \bar{e} &= \sum_{(hik) \in s} \hat{w}_{hik} \epsilon_{hik}, \\ \epsilon_{hik} &= a_{hik}(y_{hik} - \bar{Y}) - p \sum_c ({}_c\bar{Y} - \bar{Y}) {}_c\delta_{hik} \end{aligned}$$

and

$${}_c\hat{R}^r = \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_c\delta_{hik} (y_{hik} - \bar{Y}).$$

To establish the asymptotic normality of $\bar{y}_{PS}^{MI}$, we need the following regularity conditions.

C13 $n_T^{1+\delta} \sum_h \sum_i E|e_{hi} - E(e_{hi})|^{2+\delta} = O(1)$ for some $\delta > 0$ and $e_{hi} = \sum_k \hat{w}_{hik} \epsilon_{hik}$

C14 $n_TV({}_c\bar{e}) = O(1)$ for all c

$$C15 \quad n_T V(\bar{e}) \rightarrow \sigma_e^2.$$

The asymptotic normality of $\bar{y}_{PS}^{MI}$ is now established in the following theorem.

Theorem 5.1 *Under conditions C4, C13, C14, C15 and the uniform response mechanism,*

$$\sqrt{n_T}(\bar{y}_{PS}^{MI} - \bar{Y}) \longrightarrow_d N(0, \sigma_e^2/p^2)$$

where p is the probability of a response under the uniform response mechanism.

Proof

Using equation (5.2.3), we have

$$\sqrt{n_T}(\bar{y}_{PS}^{MI} - \bar{Y}) = \frac{1}{\bar{T}} \sqrt{n_T} \bar{e} + \frac{1}{\bar{T}} \sqrt{n_T} \sum_c \left(\frac{{}_c M}{{}_c \tilde{M}} - 1 \right) {}_c \bar{e}. \quad (5.2.4)$$

Working with the first term of (5.2.4), we will show

$$\frac{1}{\bar{T}} \sqrt{n_T} \bar{e} \longrightarrow_d N(0, \sigma_e^2/p^2).$$

Now,

$$\begin{aligned} \bar{T} &= \sum_c \sum_{(hik)_{cs}} {}_c \tilde{w}_{hik} a_{hik} {}_c \delta_{hik} \\ &= \frac{{}_c M}{{}_c \tilde{M}} \frac{{}_c \tilde{M}^r}{{}_c \tilde{M}} \end{aligned}$$

where ${}_c \tilde{M}^r = \sum_{(hik)_{cs}} {}_c \tilde{w}_{hik} a_{hik} {}_c \delta_{hik}$ and ${}_c \tilde{M} = \sum_{(hik)_{cs}} {}_c \tilde{w}_{hik} {}_c \delta_{hik}$. We have previously shown that $n_T V({}_c \tilde{M}) = O(1)$ (Lemma 2.7), and $n_T V({}_c \tilde{M}^r) = O(1)$ (equation (4.2.2)). Therefore, by Chebyshev's inequality we have

$${}_c \tilde{M}^r \longrightarrow_p E({}_c \tilde{M}^r) = p \frac{{}_c M}{{}_c \tilde{M}}$$

and

$${}_c \tilde{M} \longrightarrow_p \frac{{}_c M}{{}_c \tilde{M}}.$$

By continuity of the function y/x for $x \neq 0$, we have by Slutsky's Theorem

$$\frac{\hat{M}^r}{\hat{M}} \longrightarrow_p p.$$

Therefore,

$$\begin{aligned} \hat{T} &= \sum_c \frac{cM}{\hat{M}} \frac{\hat{M}^r}{\hat{M}} \\ &\longrightarrow_p p \end{aligned}$$

and again by continuity,

$$\frac{1}{\hat{T}} \longrightarrow_p \frac{1}{p}. \quad (5.2.5)$$

Next, we show that $\sqrt{n_T} \bar{\epsilon} \longrightarrow_d N(0, \sigma_\epsilon^2)$. Let

$$X_{hi} = n_T(\epsilon_{hi} - E(\epsilon_{hi}))$$

where $\epsilon_{hi} = \sum_k \tilde{w}_{hik} \epsilon_{hik}$. Now, under condition C15,

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i V(X_{hi}) &= n_T \sum_h \sum_i V(\epsilon_{hi}) \\ &= n_T V(\bar{\epsilon}) \rightarrow \sigma_\epsilon^2 \end{aligned}$$

and under condition C13,

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i E|X_{hi}|^{2+\delta} &= n_T^{1+\delta} \sum_h \sum_i E|\epsilon_{hi} - E(\epsilon_{hi})|^{2+\delta} \\ &= O(1) \end{aligned}$$

satisfying the conditions of the Central Limit Theorem, Lemma 2.1. Therefore,

$$\sqrt{n_T} \bar{\epsilon} \longrightarrow_d N(0, \sigma_\epsilon^2).$$

Applying Slutsky's Theorem, Lemma 2.3, gives

$$\frac{1}{\hat{T}} \sqrt{n_T} \bar{\epsilon} \longrightarrow_d N(0, \sigma_\epsilon^2 / p^2).$$

It remains to show that

$$\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \sqrt{n_T} {}_c \bar{e} \rightarrow_p 0.$$

By Lemma 2.7 we have

$$\left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \rightarrow_p 0$$

for all c , and under condition C14

$$\sqrt{n_T}({}_c \bar{e} - E({}_c \bar{e})) = \sqrt{n_T} {}_c \bar{e} = O_p(1)$$

for all c . Applying Lemma 2.8 gives

$$\left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \sqrt{n_T} {}_c \bar{e} \rightarrow_p 0$$

for all c . Thus, by Slutsky's Theorem

$$\begin{aligned} \sqrt{n_T}(\bar{y}_{PS}^{MI} - \bar{Y}) &= \frac{1}{\bar{T}} \sqrt{n_T} \bar{e} + \frac{1}{\bar{T}} \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \sqrt{n_T} {}_c \bar{e} \\ &\rightarrow_d N(0, \sigma_e^2 / p^2) \end{aligned}$$

■

Turning to variance estimation, equation (5.2.2) suggests the following linearization variance estimator

$$v_L(\bar{y}_{PS}^{MI}) = \frac{1}{\bar{T}^2} \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^* - \bar{e}_h^*)^2$$

where

$$\begin{aligned} e_{hi}^* &= \sum_k \hat{w}_{hik} e_{hik}^* \\ e_{hik}^* &= a_{hi} \left(y_{hik} - \frac{\hat{S}}{\bar{T}} \right) - \sum_c {}_c \bar{r} {}_c \delta_{hik}, \\ {}_c \bar{r} &= \frac{1}{{}_c \hat{M}} \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_c \delta_{hik} \left(y_{hik} - \frac{\hat{S}}{\bar{T}} \right) \\ &= \frac{1}{{}_c \hat{M}} \left({}_c \bar{Y}^r - \frac{\hat{S}}{\bar{T}} {}_c \hat{M}^r \right) \end{aligned}$$

with

$${}_c\hat{Y}^r = \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} y_{hik} {}_c\delta_{hik}$$

and

$${}_c\hat{M}^r = \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} {}_c\delta_{hik}.$$

Asymptotic consistency of $v_L(\bar{y}_{PS}^{MI})$ is established in the following theorem.

Theorem 5.2 *Under conditions C4, C13, C14, C15, the uniform response mechanism and*

$$n_TV({}_c\hat{Y}^r) = O(1) \quad (5.2.6)$$

for all c , we have

$$n_T(v_L(\bar{y}_{PS}^{MI}) - V(\bar{y}_{PS}^{MI})) \longrightarrow_p 0.$$

Proof

In order to recover the true residuals, we add and subtract $p \sum_c ({}_c\bar{Y} - \bar{Y}) {}_c\delta_{hik}$ and $\bar{Y} - \hat{S}/\hat{T}$ to the estimated residual,

$$\begin{aligned} e_{hik}^* &= a_{hik} \left(y_{hik} - \bar{Y} + \bar{Y} - \frac{\hat{S}}{\hat{T}} \right) - \sum_c \left(p({}_c\bar{Y} - \bar{Y}) - p({}_c\bar{Y} - \bar{Y}) + \frac{{}_c\bar{r}}{{}_c\hat{M}} \right) {}_c\delta_{hik} \\ &= e_{hik} + a_{hik} \left(\bar{Y} - \frac{\hat{S}}{\hat{T}} \right) - \sum_c \left(\frac{{}_c\bar{r}}{{}_c\hat{M}} - p({}_c\bar{Y} - \bar{Y}) \right) {}_c\delta_{hik} \end{aligned}$$

where $e_{hik} = a_{hik}(y_{hik} - \hat{S}/\hat{T}) - p \sum_c ({}_c\bar{Y} - \bar{Y}) {}_c\delta_{hik}$. After straight forward algebra, we have

$$\begin{aligned} e_{hi}^* - \bar{e}_h^* &= e_{hi} - \bar{e}_h + \left(\bar{Y} - \frac{\hat{S}}{\hat{T}} \right) (\hat{M}_{hi}^r - \hat{M}_h^r) - \\ &\quad \sum_c \left(\frac{{}_c\bar{r}}{{}_c\hat{M}} - p({}_c\bar{Y} - \bar{Y}) \right) ({}_c\hat{M}_{hi}^r - {}_c\hat{M}_h^r) \end{aligned}$$

where

$$\begin{aligned}\hat{M}_{hi}^r &= \sum_k \hat{w}_{hik} a_{hik}, \\ \hat{M}_h^r &= \frac{1}{n_h} \sum_i \hat{M}_{hi}^r, \\ {}_c\hat{M}_{hi}^r &= \sum_k \hat{w}_{hik} a_{hik} {}_c\delta_{hik}\end{aligned}$$

and

$${}_c\hat{M}_h^r = \frac{1}{n_h} \sum_i {}_c\hat{M}_{hi}^r.$$

The linearization variance estimator, $v_L(\bar{y}_{PS}^{MJ})$, can now be rewritten as

$$\begin{aligned}v_L(\bar{y}_{PS}^{MJ}) &= \frac{1}{\hat{T}^2} \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi}^* - \bar{e}_h^*)^2 \\ &= \frac{1}{\hat{T}^2} \sum_h \frac{n_h}{n_h - 1} \sum_i (e_{hi} - \bar{e}_h)^2 + \\ &\quad \frac{1}{\hat{T}^2} \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\left(\bar{Y} - \frac{\hat{S}}{\hat{T}} \right) (\hat{M}_{hi}^r - \hat{M}_h^r) - \right. \\ &\quad \left. \sum_c \left(\frac{{}_c\hat{r}}{{}_c\hat{M}} - p({}_c\bar{Y} - \bar{Y}) \right) ({}_c\hat{M}_{hi}^r - {}_c\hat{M}_h^r) \right]^2 + 2 \text{ cross-product terms} \\ &= \frac{1}{\hat{T}^2} v_L^*(\bar{y}_{PS}^{MJ}) + \frac{1}{\hat{T}^2} v_1 + 2 \text{ cross-product terms, say.} \quad (5.2.7)\end{aligned}$$

To show

$$n_T(v_L(\bar{y}_{PS}^{MJ}) - V(\bar{y}_{PS}^{MJ})) \longrightarrow_p 0,$$

we will show

$$\begin{aligned}n_T \left(\frac{1}{\hat{T}^2} v_L^*(\bar{y}_{PS}^{MJ}) - V(\bar{y}_{PS}^{MJ}) \right) &\longrightarrow_p 0, \\ n_T v_1 &\longrightarrow_p 0\end{aligned}$$

and

$$n_T \text{ cross-product terms} \longrightarrow_p 0.$$

Following the proof of Theorem 4.3 with C7 and C9 replaced by C13 and C15, we have

$$n_T(v_L^*(\bar{y}_{PS}^{MJ}) - \sigma_e^2) \longrightarrow_p 0.$$

Under C4, we have

$$\frac{1}{\hat{T}} \longrightarrow_p \frac{1}{p}.$$

see equation (5.2.5). Combining these two results gives

$$\frac{1}{\hat{T}^2} n_T v_L^*(\bar{y}_{PS}^{MI}) \longrightarrow_p n_T \frac{\sigma_\epsilon^2}{p^2} \approx n_T V(\bar{y}_{PS}^{MI}). \quad (5.2.8)$$

Now, turning to v_1 we have

$$\begin{aligned} v_1 &= \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\left(\bar{Y} - \frac{\hat{S}}{\hat{T}} \right) (\hat{M}_{hi}^r - \hat{M}_h^r) \right]^2 + \\ &\quad \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \left(\frac{c\hat{r}}{c\hat{M}} - p(c\bar{Y} - \bar{Y}) \right) ({}_c\hat{M}_{hi}^r - {}_c\hat{M}_h^r) \right]^2 - \\ &\quad 2 \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\left(\bar{Y} - \frac{\hat{S}}{\hat{T}} \right) (\hat{M}_{hi}^r - \hat{M}_h^r) \right] \times \\ &\quad \left[\sum_c \left(\frac{c\hat{r}}{c\hat{M}} - p(c\bar{Y} - \bar{Y}) \right) ({}_c\hat{M}_{hi}^r - {}_c\hat{M}_h^r) \right] \\ &= v_{11} + v_{12} - 2v_{13} \quad \text{say.} \end{aligned}$$

Working with v_{11} note that

$$\begin{aligned} n_T v_{11} &= \left(\bar{Y} - \frac{\hat{S}}{\hat{T}} \right)^2 n_T \sum_h \frac{n_h}{n_h - 1} \sum_i (\hat{M}_{hi}^r - \hat{M}_h^r)^2 \\ &= \left(\bar{Y} - \frac{\hat{S}}{\hat{T}} \right)^2 n_T v_L(\hat{M}^r). \end{aligned}$$

We have shown previously that $n_T V(\hat{M}^r) = O(1)$, see equation (4.2.5), and since $v_L(\hat{M}^r)$ is consistent for $V(\hat{M}^r)$, we have

$$n_T v_L(\hat{M}^r) = O_p(1).$$

Therefore, if we can show $\bar{Y} - \hat{S}/\hat{T} \longrightarrow_p 0$, we will have $n_T v_{11} \longrightarrow_p 0$. But $\hat{S}/\hat{T}$ is precisely $\bar{y}_{PS}^{MI}$ and under condition C13 we can show

$$\bar{y}_{PS}^{MI} \longrightarrow_p \bar{Y}. \quad (5.2.9)$$

Note, to show equation (5.2.9) we need only $1 + \delta$ bounded moments in condition C13 which is implied by $2 + \delta$ bounded moments. Therefore, applying Lemma 2.8 gives

$$n_T v_{11} = \left(\bar{Y} - \frac{\hat{S}}{\bar{T}} \right)^2 n_T v_L(\hat{M}^r) \rightarrow_p 0.$$

Next we show $n_T v_{12} \rightarrow_p 0$. By the Schwarz inequality we have

$$\begin{aligned} n_T v_{12} &\leq n_T \sum_h \frac{n_h}{n_h - 1} \sum_i \left[\sum_c \left(\frac{{}_c \bar{r}}{{}_c \hat{M}} - p({}_c \bar{Y} - \bar{Y}) \right)^2 \right] \left[\sum_c ({}_c \hat{M}_h^r - {}_c \hat{M}_h^r)^2 \right] \\ &= \sum_c \left(\frac{{}_c \bar{r}}{{}_c \hat{M}} - p({}_c \bar{Y} - \bar{Y}) \right)^2 n_T v_L({}_c \hat{M}). \end{aligned}$$

From Lemma 2.7, $n_T V({}_c \hat{M}) = O(1)$, which implies that for all c

$$n_T v_L({}_c \hat{M}) = O_p(1). \quad (5.2.10)$$

Now consider

$$\frac{{}_c \bar{r}}{{}_c \hat{M}} = \frac{{}_c \hat{Y}^r}{{}_c \hat{M}} - \frac{\hat{S}}{{\bar{T}}} \frac{{}_c \hat{M}^r}{{}_c \hat{M}}.$$

We have previously shown that $n_T V({}_c \hat{M}^r) = O(1)$, see equation (4.2.2). From Chebyshev's inequality and Lemma 2.7 we have

$${}_c \hat{M}^r \rightarrow_p p_c M/M,$$

$${}_c \hat{M} \rightarrow_p {}_c M/M,$$

and from equation (5.2.9),

$$\hat{S}/\bar{T} \rightarrow_p \bar{Y}.$$

Finally, by equation (5.2.6) and Chebyshev's inequality,

$${}_c \hat{Y}^r \rightarrow_p p_c Y/M.$$

Combining these results gives

$$\begin{aligned} \frac{{}_c \bar{r}}{{}_c \hat{M}} &\rightarrow_p p \frac{{}_c Y}{{}_c M} - \bar{Y} p \frac{{}_c M}{{}_c M} \\ &= p({}_c \bar{Y} - \bar{Y}) \end{aligned}$$

That is,

$$\frac{c\bar{r}}{c\bar{M}} - p(c\bar{Y} - \bar{Y}) \longrightarrow_p 0 \quad (5.2.11)$$

for all c . Applying Lemma 2.8 to equations (5.2.10) and (5.2.11) gives

$$n_T v_{12} \longrightarrow_p 0.$$

An application of the Schwarz inequality implies that $n_T v_{13} \longrightarrow_p 0$ so that

$$n_T v_1 \longrightarrow_p 0.$$

Noting that $1/\hat{T}^2 \longrightarrow_p 1/p^2$, we have

$$\frac{1}{\hat{T}^2} n_T v_1 \longrightarrow_p 0. \quad (5.2.12)$$

It remains to show that the cross-product terms in equation (5.2.7) converge to zero in probability. In light of (5.2.8) and (5.2.12), an application of the Schwarz inequality implies the cross-product terms converge to zero in probability. Thus

$$\begin{aligned} n_T v_L(\bar{y}_{PS}^{MI}) &= n_T \frac{1}{\hat{T}^2} v_L^*(\bar{y}_{PS}^{MI}) + n_T \frac{1}{\hat{T}^2} v_1 + 2n_T \text{ cross-product terms} \\ &\longrightarrow_p n_T V(\bar{y}_{PS}^{MI}) \end{aligned}$$

by equations (5.2.8) and (5.2.12). ■

Turning to the jackknife method, the standard jackknife variance estimator is

$$v_{JN}(\bar{y}_{PS}^{MI}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (\bar{y}_{PS(gj)}^{MI} - \bar{y}_{PS}^{MI})^2 \quad (5.2.13)$$

where

$$\bar{y}_{PS(gj)}^{MI} = \sum_c \sum_{(hsk) \in s} c \hat{w}_{hsk(gj)} a_{hsk} y_{hsk} c \delta_{hsk} + \sum_c \sum_{(hsk) \in s} c \hat{w}_{hsk(gj)} (1 - a_{hsk}) y_{hsk}^* c \delta_{hsk}.$$

That is, equation (5.2.13) treats the imputed values as true values. Rao and Shao (1992) have shown that this “naive” jackknife leads to severe underestimation of the true variance if the response rate is small.

When calculating a jackknife variance estimator, the entire estimation procedure should be repeated whenever a cluster is deleted. Thus, the imputed values will be affected by the deletion of the respondents in the deleted cluster. The imputed values should be adjusted to reflect the influence on the mean from these deleted respondents. We now define the adjusted imputed values for $(hik) \in s_m$, when the $(gj)^{th}$ cluster has been deleted, as

$$\begin{aligned} z_{hik(gj)}^* &= y_{hik}^* + \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} \\ &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} \end{aligned}$$

where

$$\begin{aligned} \hat{S}_{(gj)} &= \sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik(gj)} a_{hik} y_{hik} {}_c \delta_{hik} \\ \hat{T}_{(gj)} &= \sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik(gj)} a_{hik} {}_c \delta_{hik} \end{aligned}$$

and ${}_c \hat{w}_{hik(gj)}$ are the poststratified jackknife weights. Using these adjusted imputed values, we define the adjusted poststratified estimator, with the $(gj)^{th}$ cluster deleted, as

$$\begin{aligned} \bar{y}_{PS(gj)}^{MJa} &= \sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik(gj)} a_{hik} y_{hik} {}_c \delta_{hik} + \sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik(gj)} (1 - a_{hik}) z_{hik(gj)}^* {}_c \delta_{hik} \\ &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} \hat{t}_{(gj)} \\ &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} \end{aligned}$$

since $\hat{t}_{(gj)} = \sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik(gj)} a_{hik} {}_c \delta_{hik} = 1$. The jackknife variance estimator is then defined as

$$v_J(\bar{y}_{PS}^{MJ}) = \sum_g \frac{n_g - 1}{n_g} \sum_i (\bar{y}_{PS(gj)}^{MJa} - \bar{y}_{PS}^{MJ})^2.$$

We now establish the asymptotic consistency of the jackknife variance estimator.

Theorem 5.3 *Under conditions C4, C13, C14, C15 and the uniform response mechanism,*

$$n_T(v_J(\bar{y}_{PS}^{MI}) - V(\bar{y}_{PS}^{MI})) \longrightarrow_p 0.$$

Proof

We have

$$\begin{aligned} \bar{y}_{PS(gj)}^{MIa} - \bar{y}_{PS}^{MI} &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} \\ &= \frac{1}{\hat{T}_{(gj)}} \left(\hat{S}_{(gj)} - \hat{S} - \frac{\hat{S}}{\hat{T}} (\hat{T}_{(gj)} - \hat{T}) \right). \end{aligned} \quad (2.14)$$

First, working with $\hat{S}_{(gj)} - \hat{S}$ we have

$$\begin{aligned} \hat{S}_{(gj)} - \hat{S} &= \sum_c \frac{{}_cM}{\hat{M}} \left(\frac{{}_c\hat{Y}_{(gj)}^r}{\hat{M}_{(gj)}} - \frac{{}_c\hat{Y}^r}{\hat{M}} \right) \\ &= \sum_c \frac{{}_cM}{\hat{M}} \frac{1}{{}_c\hat{M}_{(gj)}} \left({}_c\hat{Y}_{(gj)}^r - {}_c\hat{Y}^r - \frac{{}_c\hat{Y}^r}{{}_c\hat{M}} ({}_c\hat{M}_{(gj)} - {}_c\hat{M}) \right) \\ &= \sum_c \frac{{}_cM}{\hat{M}} \frac{1}{{}_c\hat{M}_{(gj)}} {}_cS_{gj}^* \end{aligned} \quad (5.2.15)$$

where

$$\begin{aligned} {}_cS_{gj}^* &= {}_c\hat{Y}_{(gj)}^r - {}_c\hat{Y}^r - \frac{{}_c\hat{Y}^r}{{}_c\hat{M}} ({}_c\hat{M}_{(gj)} - {}_c\hat{M}), \\ {}_c\hat{Y}^r &= \sum_{(hik) \in s} \hat{w}_{hik} a_{hik} y_{hik} {}_c\delta_{hik}, \\ {}_c\hat{M} &= \sum_{(hik) \in s} \hat{w}_{hik} {}_c\delta_{hik}, \end{aligned}$$

and ${}_c\hat{Y}_{(gj)}^r$ and ${}_c\hat{M}_{(gj)}$ are obtained from ${}_c\hat{Y}^r$ and ${}_c\hat{M}$ by replacing $\hat{w}_{hik}$ with $\hat{w}_{hik(gj)}$

Expanding the function ${}_c\hat{M}_{(gj)}^{-1}$ around the point ${}_c\hat{M}$, gives

$$\frac{1}{{}_c\hat{M}_{(gj)}} = \frac{1}{{}_c\hat{M}} - \frac{1}{{}_c\hat{M}^2} ({}_c\hat{M}_{(gj)} - {}_c\hat{M}) \quad (5.2.16)$$

where ${}_c x_{gj}$ lies between ${}_c \hat{M}_{(gj)}$ and ${}_c \hat{M}$. Putting expression (5.2.16) into equation (5.2.15) gives

$$\hat{S}_{(gj)} - \hat{S} = \sum_c \frac{{}_c M}{M} \frac{1}{{}_c \hat{M}} {}_c S_{gj}^* - \sum_c \frac{{}_c M}{M} \frac{1}{{}_c x_{gj}^2} {}_c S_{gj}^* ({}_c \hat{M}_{(gj)} - {}_c \hat{M}). \quad (5.2.17)$$

Similarly, we have

$$\hat{T}_{(gj)} - \hat{T} = \sum_c \frac{{}_c M}{M} \frac{1}{{}_c \hat{M}} {}_c T_{gj}^* - \sum_c \frac{{}_c M}{M} \frac{1}{{}_c x_{gj}^2} {}_c T_{gj}^* ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \quad (5.2.18)$$

where

$$\begin{aligned} {}_c T_{gj}^* &= {}_c \hat{M}_{(gj)}^r - {}_c \hat{M}^r - \frac{{}_c \hat{M}^r}{{}_c \hat{M}} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \\ {}_c \hat{M}^r &= \sum_{(hik)es} \hat{w}_{hik} a_{hik} {}_c \delta_{hik} \end{aligned}$$

and

$${}_c \hat{M}_{(gj)}^r = \sum_{(hik)es} \hat{w}_{hik(gj)} a_{hik} {}_c \delta_{hik}.$$

Using equations (5.2.17) and (5.2.18) in equation (5.2.14) gives

$$\begin{aligned} y_{PS(hi)}^{MIa} - \bar{y}_{PS}^{MI} &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} \\ &= \frac{1}{\hat{T}_{(gj)}} \left(\hat{S}_{(gj)} - \hat{S} - \frac{\hat{S}}{\hat{T}} (\hat{T}_{(gj)} - \hat{T}) \right) \\ &= \frac{1}{\hat{T}_{(gj)}} \left[\sum_c \frac{{}_c M}{M} \frac{1}{{}_c \hat{M}} \left({}_c S_{gj}^* - \frac{\hat{S}}{\hat{T}} {}_c T_{gj}^* \right) - \right. \\ &\quad \left. \sum_c \frac{{}_c M}{M} \frac{1}{{}_c x_{gj}^2} \left({}_c S_{gj}^* - \frac{\hat{S}}{\hat{T}} {}_c T_{gj}^* \right) ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right] \\ &= \frac{1}{\hat{T}_{(gj)}} \left[\sum_c \frac{{}_c M}{M} \frac{1}{{}_c \hat{M}} ({}_c \bar{e}_g^* - {}_c e_{gj}^*) - \sum_c \frac{{}_c M}{M} \frac{1}{{}_c x_{gj}^2} ({}_c \bar{e}_g^* - {}_c e_{gj}^*) \times \right. \\ &\quad \left. ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right] \end{aligned}$$

where

$${}_c S_{gj}^* - \frac{\hat{S}}{\hat{T}} {}_c T_{gj}^*$$

$$\begin{aligned}
&= {}_c\hat{Y}_{(g)}^r - {}_c\hat{Y}^r - \frac{\hat{S}}{\hat{T}}({}_c\hat{M}_{(g)}^r - {}_c\hat{M}^r) - \left(\frac{{}_c\hat{Y}^r}{{}_c\hat{M}} - \frac{\hat{S}}{\hat{T}} \frac{{}_c\hat{M}^r}{{}_c\hat{M}} \right) ({}_c\hat{M}_{(g)} - {}_c\hat{M}) \\
&= \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik) \in s} \hat{w}_{gik} {}_c\delta_{gik} [a_{gik}(\hat{y}_{gik} - \hat{S}/\hat{T}) - {}_c\bar{r}] - \right. \\
&\quad \left. \sum_{(gjk) \in s} \hat{w}_{gjk} {}_c\delta_{gjk} [a_{gjk}(\hat{y}_{gjk} - \hat{S}/\hat{T}) - {}_c\bar{r}] \right) \\
&= \frac{n_g}{n_g - 1} ({}_c\bar{\epsilon}_g^* - {}_c\epsilon_{gg}^*).
\end{aligned}$$

Now, expanding the function $\hat{T}_{(g)}^{-1}$ around the point $\hat{T}$ gives

$$\frac{1}{\hat{T}_{(g)}} = \frac{1}{\hat{T}} - \frac{1}{x_{gg}^2}(\hat{T}_{(g)} - \hat{T})$$

where x_{gg} lies between $\hat{T}_{(g)}$ and $\hat{T}$. This gives

$$\begin{aligned}
&\bar{y}_{PS(g)}^{MIa} - \bar{y}_{PS}^{MI} \\
&= \frac{1}{\hat{T}} \left[\sum_c \frac{{}_cM}{{}_c\hat{M}} ({}_c\bar{\epsilon}_g^* - {}_c\epsilon_{gg}^*) - \sum_c \frac{{}_cM}{\hat{M}} \frac{1}{{}_c x_{gg}^2} ({}_c\bar{\epsilon}_g^* - {}_c\epsilon_{gg}^*) ({}_c\hat{M}_{(g)} - {}_c\hat{M}) \right] - \\
&\quad \frac{1}{x_{gg}^2} (\hat{T}_{(g)} - \hat{T}) \left[\sum_c \frac{{}_cM}{{}_c\hat{M}} ({}_c\bar{\epsilon}_g^* - {}_c\epsilon_{gg}^*) - \sum_c \frac{{}_cM}{\hat{M}} \frac{1}{{}_c x_{gg}^2} ({}_c\bar{\epsilon}_g^* - {}_c\epsilon_{gg}^*) ({}_c\hat{M}_{(g)} - {}_c\hat{M}) \right].
\end{aligned}$$

Putting this expression in the jackknife variance estimator gives

$$\begin{aligned}
v_J(\bar{y}_{PS}^{MI}) &= \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\frac{\hat{S}_{(g)}}{\hat{T}_{(g)}} - \frac{\hat{S}}{\hat{T}} \right)^2 \\
&= \sum_g \frac{n_g - 1}{n_g} \sum_j \frac{1}{\hat{T}^2} \left[\sum_c ({}_c\bar{\epsilon}_g^* - {}_c\epsilon_{gg}^*) \left(\frac{{}_cM}{{}_c\hat{M}} - \frac{{}_cM}{\hat{M}} \frac{1}{{}_c x_{gg}^2} ({}_c\hat{M}_{(g)} - {}_c\hat{M}) \right) \right]^2 + \\
&\quad \sum_g \frac{n_g - 1}{n_g} \sum_j \frac{1}{x_{gg}^2} (\hat{T}_{(g)} - \hat{T})^2 \left[\sum_c ({}_c\bar{\epsilon}_g^* - {}_c\epsilon_{gg}^*) \times \right. \\
&\quad \left. \left(\frac{{}_cM}{{}_c\hat{M}} - \frac{{}_cM}{\hat{M}} \frac{1}{{}_c x_{gg}^2} ({}_c\hat{M}_{(g)} - {}_c\hat{M}) \right) \right]^2 - 2 \text{ cross-product terms} \\
&= v_1 + v_2 - 2 \text{ cross-product terms, say.}
\end{aligned}$$

We establish consistency of $v_J(\bar{y}_{PS}^{MI})$ by showing

$$n_T v_1 \longrightarrow_p n_T V(\bar{y}_{PS}^{MI}) \quad (5.2.19)$$

$$n_T v_2 \longrightarrow_p 0 \quad (5.2.20)$$

and

$$n_T \text{ cross-product terms } \longrightarrow_p 0. \quad (5.2.21)$$

Proofs of equations (5.2.20) and (5.2.21) are given in the appendix at the end of this chapter. To show (5.2.19), note that

$$\begin{aligned} v_1 &= \frac{1}{\hat{T}^2} \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_c M}{{}_c \hat{M}} ({}_c \bar{e}_g^* - {}_c e_{gj}^*) \right]^2 + \\ &\quad \frac{1}{\hat{T}^2} \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \frac{{}_c M}{{}_c \hat{M}} \frac{1}{{}_c x_{gj}^2} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) ({}_c \bar{e}_g^* - {}_c e_{gj}^*) \right]^2 - \\ &\quad \text{2 cross-product terms} \\ &= v_{11} + v_{12} - 2v_{13} \text{ say.} \end{aligned}$$

Working with v_{11} , we have

$$\begin{aligned} n_T v_{11} &= \frac{1}{\hat{T}^2} n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[(\bar{e}_g^* - e_{gj}^*) + \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) ({}_c \bar{e}_g^* - {}_c e_{gj}^*) \right]^2 \\ &= \frac{1}{\hat{T}^2} n_T \sum_g \frac{n_g}{n_g - 1} \sum_j (\bar{e}_g^* - e_{gj}^*)^2 + \\ &\quad \frac{1}{\hat{T}^2} n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) ({}_c \bar{e}_g^* - {}_c e_{gj}^*) \right]^2 + \\ &\quad \text{2 cross-product terms} \\ &= n_T v_L(\bar{y}_{PS}^{MI}) + \frac{1}{\hat{T}^2} n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) ({}_c \bar{e}_g^* - {}_c e_{gj}^*) \right]^2 + \\ &\quad \text{2 cross-product terms.} \end{aligned} \quad (5.2.22)$$

From Theorem 5.2,

$$n_T v_L(\bar{y}_{PS}^{MI}) \longrightarrow_p n_T V(\bar{y}_{PS}^{MI}).$$

We now show that the second term in (5.2.22) converges to zero in probability. From the Schwarz inequality

$$n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) ({}_c \bar{e}_g^* - {}_c e_{gj}^*) \right]^2$$

$$\begin{aligned}
&\leq n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right)^2 \right] \left[\sum_c ({}_c \bar{e}_g^* - {}_c e_{gj}^*)^2 \right] \\
&= \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right)^2 n_T v_L({}_c \bar{e}).
\end{aligned}$$

By Lemma 2.7,

$$\left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \longrightarrow_p 0$$

for all c , and under condition C14,

$$n_T v_L({}_c \bar{e}) = O_p(1)$$

for all c . Noting that $\hat{T}^{-1} \longrightarrow_p p^{-1}$, we have

$$\frac{1}{\hat{T}^2} n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) ({}_c \bar{e}_g^* - {}_c e_{gj}^*) \right]^2 \longrightarrow_p 0.$$

An application of the Schwarz inequality implies that the cross-product terms in equation (5.2.22) converge to zero in probability. Therefore, we have

$$n_T v_{11} \longrightarrow_p n_T V(y_{PS}^{MI}).$$

Note that the term v_{11} is a consistent estimator of the variance of y_{PS}^{MI} and can be used as a linearization variance estimator. This expression was obtained by using a Taylor linearization of the jackknife variance estimator and will be denoted as the linearized jackknife variance estimator,

$$v_{JL}(\bar{y}_{PS}^{MI}) = v(r_{ht})$$

where $r_{ht} = \hat{T}^{-1} \sum_k {}_c \hat{w}_{hik} {}_c e_{hik} {}_c \delta_{hik}$. This linearized jackknife variance estimator is similar to the robust variance estimator proposed by Rao (1985) in the context of one-way poststratification.

Next consider v_{12} . By the Schwarz inequality,

$$n_T v_{12} \leq n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} \right)^2 ({}_c \bar{e}_g^* - {}_c e_{gj}^*)^2 \right] \times$$

$$\begin{aligned}
& \left[\sum_c \left(\frac{1}{c x_{gj}^2} \right)^2 ({}_c \hat{M}_{(gj)} - {}_c \hat{M})^2 \right]^2 \\
& \leq \max_{g,j,c} ({}_c \hat{M}_{(gj)} - {}_c \hat{M})^2 n_T \sum_g \frac{n_g}{n_g - 1} \sum_j \left[\sum_c ({}_c \bar{e}_g^* - {}_c e_{gj}^*)^2 \right] \times \\
& \quad \left[\sum_c \left(\frac{1}{c x_{gj}^2} - \frac{1}{{}_c \hat{M}} + \frac{1}{{}_c \bar{M}} \right)^2 \right] \\
& \leq \max_{g,j,c} ({}_c \hat{M}_{(gj)} - {}_c \hat{M})^2 \left[\sum_c \left(\frac{{}_c M}{{}_c \hat{M}} \right)^4 \left(\frac{M}{{}_c \bar{M}} \right)^4 \right] \sum_c n_T v_L({}_c \bar{e}) + \\
& \quad \max_{g,j,c} ({}_c \hat{M}_{(gj)} - {}_c \hat{M})^2 \max_{g,j,c} \left(\frac{1}{c x_{gj}^2} - \frac{1}{{}_c \hat{M}} \right)^2 \sum_c n_T v_L({}_c \bar{e}) + \\
& \quad \text{2 cross-product terms} \\
& \longrightarrow_p 0
\end{aligned}$$

since, for the first term, $\max_{g,j,c} |{}_c \hat{M}_{(gj)} - {}_c \hat{M}| \longrightarrow_p 0$, see Lemma 2.9, ${}_c M / {}_c \hat{M} \longrightarrow_p 1$, see Lemma 2.7, $M / {}_c \bar{M} \longrightarrow 1 / {}_c \lambda$, and $n_T v_L({}_c \bar{e}) = O_p(1)$ for all c . For the second term, we further note that, see equation (2.4.11),

$$\max_{g,j,c} \left| \frac{1}{c x_{gj}^2} - \frac{1}{{}_c \hat{M}} \right|^2 \longrightarrow_p 0.$$

The Schwarz inequality implies that the cross-products terms converge to zero in probability. Finally, $n_T v_{13} \longrightarrow_p 0$ by the Schwarz inequality. Thus

$$\begin{aligned}
n_T v_1 &= n_T v_{11} + n_T v_{12} - 2n_T v_{13} \\
&\longrightarrow_p n_T V(\bar{y}_{PS}^{MI}).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
n_T v_J(\bar{y}_{PS}^{MI}) &= n_T v_1 + n_T v_2 - 2n_T \text{ cross-product terms} \\
&\longrightarrow_p n_T V(\bar{y}_{PS}^{MI})
\end{aligned}$$

or

$$n_T (v_J(\bar{y}_{PS}^{MI}) - V(\bar{y}_{PS}^{MI})) \longrightarrow_p 0$$

as required. ■

5.2.2 Multiple Imputation Classes

Suppose that we have J imputation classes which can cut across the poststrata. For a nonrespondent, we will impute the mean of the imputation class which the individual falls into. That is, for $(hik) \in s_{jm}$ let

$$\begin{aligned} y_{hik}^* &= \frac{\sum_c \sum_{(hik) \in s_j} {}_c \hat{w}_{hik} a_{hik} y_{hik}}{\sum_c \sum_{(hik) \in s_j} {}_c \hat{w}_{hik} a_{hik}} \\ &= \frac{\hat{S}_j}{\hat{T}_j} \end{aligned}$$

where ${}_c \hat{w}_{hik}$ is the usual poststratified adjusted weight, s_{jm} denotes the nonrespondents in the j^{th} imputation class and s_j denotes the sample in the j^{th} imputation class. The imputed poststratified estimator is given by

$$\begin{aligned} \bar{y}_{PS}^{MJ} &= \sum_j \sum_c \sum_{(hik) \in s_j} {}_c \hat{w}_{hik} a_{hik} y_{hik} {}_c \delta_{hik} + \sum_j \sum_c \sum_{(hik) \in s_j} {}_c \hat{w}_{hik} (1 - a_{hik}) y_{hik}^* {}_c \delta_{hik} \\ &= \sum_j \frac{\hat{S}_j}{\hat{T}_j} \hat{U}_j \end{aligned}$$

where

$$\hat{U}_j = \sum_c \sum_{(hik) \in s_j} {}_c \hat{w}_{hik} {}_c \delta_{hik}.$$

Assuming a uniform response mechanism in each imputation class, the imputed poststratified estimator is approximately unbiased for $\bar{Y}$ since

$$\begin{aligned} E(\hat{S}_j) &= p_j E \left(\sum_c \sum_{(hik) \in s_j} {}_c \hat{w}_{hik} y_{hik} {}_c \delta_{hik} \right) \quad \text{and} \\ E(\hat{T}_j) &= p_j E(\hat{U}_j) \end{aligned}$$

which gives

$$\begin{aligned} E \left(\sum_j \frac{\hat{S}_j}{\hat{T}_j} \hat{U}_j \right) &\approx \sum_j \frac{p_j E \left(\sum_c \sum_{(hik) \in s_j} {}_c \hat{w}_{hik} y_{hik} {}_c \delta_{hik} \right)}{p_j E(\hat{U}_j)} E(\hat{U}_j) \\ &= E(\bar{y}_{PS}) = \bar{Y}. \end{aligned}$$

The adjusted imputed values for $(hik)\epsilon s_{jm}$, with the $(gl)^{th}$ cluster deleted, are

$$z_{hik(gl)}^* = y_{hik}^* + \frac{\hat{S}_{j(gl)}}{\hat{T}_{j(gl)}} - \frac{\hat{S}_j}{\hat{T}_j}$$

and the corresponding adjusted imputed poststratified estimator is

$$\begin{aligned} \bar{y}_{PS(gl)}^{M/a} &= \sum_j \sum_c \sum_{(hik)\epsilon s_j} c \hat{w}_{hik(gl)} a_{hik} y_{hik} c \delta_{hik} + \\ &\quad \sum_j \sum_c \sum_{(hik)\epsilon s_j} c \hat{w}_{hik(gl)} (1 - a_{hik}) z_{hik(gl)}^* c \delta_{hik} \\ &= \sum_j \frac{\hat{S}_{j(gl)}}{\hat{T}_{j(gl)}} \hat{U}_{j(gl)} \end{aligned}$$

where $\hat{S}_{(gj)}$, $\hat{T}_{(gj)}$ and $\hat{U}_{(gj)}$ are obtained from $\hat{S}, \hat{T}$ and $\hat{U}$, respectively, by replacing $\hat{w}_{hik}$ with $\hat{w}_{hik(gj)}$. The jackknife variance estimator is then given by

$$v_J(\bar{y}_{PS}^{MI}) = \sum_g \frac{n_g - 1}{n_g} \sum_l \left(\bar{y}_{PS(gl)}^{M/a} - \bar{y}_{PS}^{MI} \right)^2.$$

Asymptotic consistency of $v_J(\bar{y}_{PS}^{MI})$ can be established along the lines of Theorem 5.3 but is not included here for simplicity.

A linearization variance estimator can be obtained by linearizing the jackknife variance estimator. The calculations are lengthy but straightforward and are not presented here for simplicity. The linearized jackknife variance estimator is

$$v_{JL}(\bar{y}_{PS}^{MI}) = v(r_{hi})$$

where

$$r_{hi} = \sum_j \sum_c \left[\frac{\hat{U}}{\hat{T}} \sum_k c \hat{w}_{hik} c \epsilon_{hik}^j + \frac{\hat{S}_j}{\hat{T}_j} \sum_k c \hat{w}_{hik} c \delta_{hik} \left(\eta_{hik}^j - \frac{c \hat{M}_j}{c \hat{M}} \right) \right],$$

with

$$c \epsilon_{hik}^j = a_{hik} \eta_{hik}^j \left(y_{hik} - \frac{\hat{S}_j}{\hat{T}_j} \right) - \left(\frac{c \hat{Y}_j^r}{c \hat{M}} - \frac{\hat{S}_j}{\hat{T}_j} \frac{c \hat{M}_j^r}{c \hat{M}} \right),$$

$$\begin{aligned}
{}_c\hat{M}_j &= \sum_{(hik) \in s_j} \hat{w}_{hik} {}_c\delta_{hik}, \\
{}_c\hat{M}_j^r &= \sum_{(hik) \in s_j} \hat{w}_{hik} a_{hik} {}_c\delta_{hik}, \\
{}_c\hat{Y}_j^r &= \sum_{(hik) \in s_j} \hat{w}_{hik} a_{hik} y_{hik} {}_c\delta_{hik}
\end{aligned}$$

and η_{hik}^j is the j^{th} imputation class indicator variable defined as

$$\eta_{hik}^j = \begin{cases} 1 & \text{if the } (hik)^{th} \text{ unit is in the } j^{th} \text{ imputation class} \\ 0 & \text{otherwise.} \end{cases}$$

Sometimes the poststrata are used as imputation classes. That is, for $(hik) \in s_m$, let

$$\begin{aligned}
y_{hik}^* &= \frac{\sum_{(hik) \in s} {}_c\hat{w}_{hik} a_{hik} y_{hik} {}_c\delta_{hik}}{\sum_{(hik) \in s} {}_c\hat{w}_{hik} a_{hik} {}_c\delta_{hik}} \\
&= \frac{{}_c\hat{S}}{{}_c\hat{T}}.
\end{aligned}$$

In this special case, the linearized jackknife variance estimator is

$$v_{JL}(\bar{y}_{PS}^{MI}) = v(r_{h_1}^*)$$

where

$$\begin{aligned}
r_{h_1}^* &= \sum_C \sum_k {}_c\hat{w}_{hik}^N a_{hik} {}_c\epsilon_{hik}^* {}_c\delta_{hik} \\
{}_c\epsilon_{hik}^* &= y_{hik} - \frac{{}_c\hat{Y}^r}{{}_c\hat{M}^r}
\end{aligned}$$

and

$${}_c\hat{w}_{hik}^N = \frac{{}_cM}{{}_c\hat{M}^r} \hat{w}_{hik}.$$

That is, $v_{JL}(\bar{y}_{PS}^{MI})$ coincides with the linearization variance estimator obtained under a nonresponse adjustment with a single weighting class, see equation (4.2.9).

5.3 Simulation Study

A limited simulation study was performed to evaluate the effectiveness of the jackknife and the linearized jackknife variance estimators for the poststratified estimator under mean imputation. Using the same population and sampling scheme as described in Section 4.4, we used mean imputation to compensate for the missing values. In each sample we computed the poststratified estimator, $\hat{Y}_{PS}$, three variance estimators, v_J , v_{JL} , and v_{JN} , and the corresponding error rates associated with the normal theory confidence intervals. The "naive" jackknife variance estimator, v_{JN} , treats the imputed values as true values and leads to underestimation of the true variance. The performance measures, relative bias, error rates, and average length of the confidence intervals, were calculated as outlined in Section 3.4.

Table 5.1 presents results, averaged over 10,000 samples, for the 50 strata design. The correct jackknife variance estimators, v_J and v_{JL} , appear to be estimating the true variance of $\hat{Y}_{PS}$ with relative biases less than 5% and error rates just slightly over the nominal 5%. On the other hand, the naive jackknife, v_{JN} , is underestimating the true variance with the underestimation becoming more severe as the response rate decreases. With a response rate of 70%, v_{JN} suffers a 65% underestimation. The error rates corresponding to v_{JN} are also at unacceptable levels varying from 11% to 25% for a nominal 5% level, as the response rate decreases from 0.9 to 0.7. Turning to the average length of the confidence intervals, we see that the average length corresponding to v_J and v_{JL} are increasing as the response rate decreases while for v_{JN} the average length actually decreases. This indicates that v_J and v_{JL} are correctly tracking the increasing MSE as the response rate decreases while v_{JN} is remaining relatively stable as p decreases.

Results from the 100 strata design are presented in Table 5.2. Again, v_J and v_{JL} perform well in terms of both relative biases and error rates while the naive jackknife, v_{JN} , severely underestimates the true variance of $\hat{Y}_{PS}$. We also note that

Table 5.1: Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Mean Imputation (50 Strata Design)

Performance Measure	$v_J(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_{JN}(\hat{Y}_{PS})$
<u>$p = 0.9$</u>			
Relative Bias	-4.06 %	-3.45 %	-29.49 %
Error Rate	6.19 %	6.16 %	10.84 %
Lower Error Rate	2.42 %	2.42 %	4.67 %
Upper Error Rate	3.76 %	3.74 %	6.17 %
Average Length	5.77	5.75	4.93
<u>$p = 0.8$</u>			
Relative Bias	-4.10 %	-3.49 %	-49.78 %
Error Rate	6.44 %	6.35 %	17.18 %
Lower Error Rate	2.68 %	2.61 %	8.03 %
Upper Error Rate	3.76 %	3.74 %	9.15 %
Average Length	6.14	6.12	4.43
<u>$p = 0.7$</u>			
Relative Bias	-4.88 %	-4.25 %	-64.52 %
Error Rate	6.37 %	6.33 %	25.33 %
Lower Error Rate	2.53 %	2.51 %	11.82 %
Upper Error Rate	3.84 %	3.82 %	13.51 %
Average Length	6.60	6.58	4.01

Table 5.2: Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Mean Imputation (100 Strata Design)

Performance Measure	$v_J(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_{JN}(\hat{Y}_{PS})$
<u>$p = 0.9$</u>			
Relative Bias	-1.77 %	-1.63 %	-28.19 %
Error Rate	5.76 %	5.75 %	10.05 %
Lower Error Rate	2.42 %	2.42 %	4.30 %
Upper Error Rate	3.34 %	3.33 %	5.75 %
Average Length	4.06	4.05	3.47
<u>$p = 0.8$</u>			
Relative Bias	-4.05 %	-3.91 %	-50.03 %
Error Rate	6.07 %	6.07 %	16.90 %
Lower Error Rate	2.73 %	2.73 %	8.39 %
Upper Error Rate	3.34 %	3.34 %	8.51 %
Average Length	4.33	4.33	3.12
<u>$p = 0.7$</u>			
Relative Bias	-1.48 %	-1.33 %	-63.52 %
Error Rate	5.60 %	5.60 %	24.14 %
Lower Error Rate	2.37 %	2.37 %	11.49 %
Upper Error Rate	3.23 %	3.23 %	12.65 %
Average Length	4.66	4.66	2.83

v_J and v_{JL} are almost exactly the same in the 100 strata design case as indicated by the similarity of the average length of the confidence intervals. We also note that the average length of the v_{JN} confidence intervals is decreasing as p decreases indicating that v_{JN} is not tracking the MSE properly.

5.4 Appendix

Proof of equation (5.2.20)

We wish to show $n_T v_2 \longrightarrow_p 0$, where

$$v_2 = \sum_g \frac{n_g}{n_g - 1} \sum_j \frac{1}{x_{gj}^2} (\hat{T}_{(gj)} - \hat{T})^2 \left[\sum_c \frac{{}_c M}{{}_c \hat{M}} ({}_c \bar{\epsilon}_g^* - {}_c \epsilon_{gj}^*) - \sum_c \frac{{}_c M}{{}_c \hat{M}} \frac{1}{{}_c x_{gj}^2} \times \right. \\ \left. ({}_c \bar{\epsilon}_g^* - {}_c \epsilon_{gj}^*) ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right]^2. \quad (5.4.1)$$

We first show

$$\max_{g,j} |\hat{T}_{(gj)} - \hat{T}| \longrightarrow_p 0$$

which, by continuity, will imply

$$\max_{g,j} \left| \frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c \hat{M}} \right| \longrightarrow_p 0.$$

Now, using equation (5.2.18)

$$\begin{aligned} & \max_{g,j} |\hat{T}_{(gj)} - \hat{T}| \\ &= \max_{g,j} \left| \sum_c {}_c T_{gj}^* \left(\frac{{}_c M}{{}_c \hat{M}} - \frac{{}_c M}{{}_c \hat{M}} \frac{1}{{}_c x_{gj}^2} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right) \right| \\ &\leq \sum_c \max_{g,j} \left| {}_c T_{gj}^* \left(\frac{{}_c M}{{}_c \hat{M}} - \frac{{}_c M}{{}_c \hat{M}} \frac{1}{{}_c x_{gj}^2} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right) \right| \\ &\leq \sum_c \max_{g,j} |{}_c T_{gj}^*| \left(\frac{{}_c M}{{}_c \hat{M}} \right) + \sum_c \frac{{}_c M}{{}_c \hat{M}} \max_{g,j} |{}_c T_{gj}^*| \left| \frac{1}{{}_c x_{gj}^2} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right|. \end{aligned}$$

Working with the first term in (5.4.1),

$$\begin{aligned} & \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} \right) \max_{g,j} |{}_c T_{gj}^*| \\ &= \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} \right) \max_{g,j} \left| ({}_c \hat{M}_{(gj)}^r - {}_c \hat{M}^r) - \frac{{}_c \hat{M}^r}{{}_c \hat{M}} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right| \\ &\leq \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} \right) \left[\max_{g,j} |{}_c \hat{M}_{(gj)}^r - {}_c \hat{M}^r| + \left(\frac{{}_c \hat{M}^r}{{}_c \hat{M}} \right) \max_{g,j} |{}_c \hat{M}_{(gj)} - {}_c \hat{M}| \right] \\ &\longrightarrow_p 0 \end{aligned}$$

since

$$\begin{aligned}\frac{{}_c\hat{M}^r}{{}_c\hat{M}} &\longrightarrow_p p, \\ \frac{{}_c\hat{M}}{{}_c\hat{M}} &\longrightarrow_p 1, \\ \max_{g,j} |{}_c\hat{M}_{(gj)} - {}_c\hat{M}| &\longrightarrow_p 0 \text{ (see Lemma 2.9)}\end{aligned}$$

and, as will be shown,

$$\max_{g,j} |{}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r| \longrightarrow_p 0.$$

Now, consider the second term in (5.4.1),

$$\begin{aligned}& \sum_c \frac{{}_cM}{\hat{M}} \max_{g,j} |{}_cT_g^*| \left| \frac{1}{{}_c x_{gj}^2} ({}_c\hat{M}_{(gj)} - {}_c\hat{M}) \right| \\ & \leq \sum_c \frac{{}_cM}{\hat{M}} \max_{g,j} |{}_cT_g^*| \max_{g,j} \left| \left(\frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} + \frac{1}{{}_c\hat{M}^2} \right) ({}_c\hat{M}_{(gj)} - {}_c\hat{M}) \right| \\ & \leq \sum_c \frac{{}_cM}{\hat{M}} \left[\max_{g,j} |{}_cT_g^*| \right] \left[\max_{g,j} \left| \frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right| \right] \left[\frac{1}{{}_c\hat{M}^2} \max_{g,j} |{}_c\hat{M}_{(gj)} - {}_c\hat{M}| \right] \\ & \longrightarrow_p 0\end{aligned}$$

since ${}_cM/\hat{M} < 1$,

$$\begin{aligned}\frac{1}{{}_c\hat{M}^2} &\longrightarrow_p \frac{1}{{}_c\lambda} < \infty, \\ \max_{g,j} |{}_cT_g^*| &\longrightarrow_p 0, \\ \max_{g,j} \left| \frac{1}{{}_c x_{gj}^2} - \frac{1}{{}_c\hat{M}^2} \right| &\longrightarrow_p 0 \text{ see equation (2.4.11)}\end{aligned}$$

and

$$\max_{g,j} |{}_c\hat{M}_{(gj)} - {}_c\hat{M}| \longrightarrow_p 0.$$

It remains to show that

$$\max_{g,j} |{}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r| \longrightarrow_p 0.$$

Noting that

$${}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r = \frac{n_g}{n_g - 1} \left(\frac{1}{n_g} \sum_{(gik)^{\neq s}} \hat{w}_{gik} a_{gik} \delta_{gik} - \sum_{(gj^k)^{\neq s}} \hat{w}_{gj^k} a_{gj^k} \delta_{gj^k} \right),$$

we have

$$\begin{aligned}
 P \left\{ \max_{g,j} |{}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r| > \epsilon \right\} &\leq \sum_g \sum_j P \{ |{}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r| > \epsilon \} \\
 &\leq \sum_g \sum_j \frac{1}{\epsilon^2} E({}_c\hat{M}_{(gj)}^r - {}_c\hat{M}^r)^2 \\
 &\leq \frac{2}{\epsilon^2} V({}_c\hat{M}^r) \rightarrow 0
 \end{aligned}$$

from equation (4.2.2). Therefore,

$$\max_{g,j} |\hat{T}_{(gj)} - \hat{T}| \longrightarrow_p 0. \quad (5.4.2)$$

Next, we show that

$$\max_{g,j} \left| \frac{1}{x_{gj}^2} - \frac{1}{\hat{T}} \right| \longrightarrow_p 0.$$

Noting that the function $1/x^2$ is continuous for $x \neq 0$, for every $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that

$$\left| \frac{1}{x^2} - \frac{1}{y^2} \right| < \epsilon$$

whenever $|x - y| < \delta_\epsilon$. Thus, since $\max_{g,j} |\hat{T}_{(gj)} - \hat{T}| \longrightarrow_p 0$,

$$\begin{aligned}
 P \left\{ \max_{g,j} \left| \frac{1}{x_{gj}^2} - \frac{1}{\hat{T}} \right| < \epsilon \right\} &\geq P \left\{ \max_{g,j} |\hat{T}_{(gj)} - \hat{T}| < \delta_\epsilon \right\} \\
 &\rightarrow 1.
 \end{aligned}$$

That is,

$$\max_{g,j} \left| \frac{1}{x_{gj}^2} - \frac{1}{\hat{T}} \right| \longrightarrow_p 0 \quad (5.4.3)$$

for all c . Using equations (5.4.2) and (5.4.3), we can show $n_T v_2 \longrightarrow_p 0$ by following the proof of $n_T v_{12} \longrightarrow_p 0$ in Theorem 5.3.

Proof of equation (5.2.21)

We wish to show

$$2n_T \text{ cross-product term } \longrightarrow_p 0.$$

We have

$$n_T v_1 \longrightarrow_p n_T V(\bar{y}_{PS}^{M_I})$$

and

$$n_T v_2 \longrightarrow_p 0.$$

An application of the Schwartz inequality will imply equation (5.2.21).

Chapter 6

Hot Deck Imputation

6.1 Introduction

Hot deck imputation is commonly employed for item nonresponse for the following reasons: (a) It preserves the distribution of the item values unlike mean imputation which leads to a “spike” at the the mean of the respondents for the item y . (b) Results obtained from different analyses are consistent with one another unlike results from analyses from an incomplete data set. (c) It permits the use of the same survey weight for all items, unlike the weighting adjustment method.

In the simplest form of hot deck imputation, a simple random sample is selected with replacement from the sample respondents to an item y and the associated item values are used as donors. While this simple hot deck provides unbiased estimates for totals and means for simple random sampling under uniform response, the resulting estimator under a stratified multi-stage design will be biased. Unbiased estimation can be achieved by using a weighted hot deck procedure. Donors are selected with replacement with probabilities proportional to their sample weights. That is, a donor $(gjl) \in s_r$ is selected with probability $\hat{w}_{gjl} / \sum_{(hik) \in s_r} \hat{w}_{hik}$ and y_{gjl} is imputed for $(hik) \in s_m$. In Chapter 5, we noted that when performing mean imputation, one could use the design weights or the poststratified weights when calculating the mean.

In the hot deck case, one could use the design weights when selecting a donor, as outlined above, or one could select donors with probabilities proportional to their poststratified adjusted weights. In this case, a donor is selected with probability $c\hat{w}_{gjl} / \sum_c \sum_{(gjl) \in s} c\hat{w}_{gjl}a_{gjl}c\delta_{gjl}$ where the donor is in the c^{th} poststratum. It is felt that using the poststratified weights will lead to more efficient estimation.

It is common practice to treat the imputed values as if they are true values and use standard estimation procedures. This will lead to valid point estimates but variance estimates will severely underestimate the true variance when the proportion of missing values is appreciable. Rao and Shao (1992) proposed a jackknife variance estimator for the basic expansion estimator under the weighted hot deck imputation scheme. They showed consistency of the new jackknife variance estimator under a uniform response mechanism. In this chapter, the poststratified estimator will be considered under the weighted hot deck imputation scheme using the poststratified weights. We consider only the one imputation class situation but note that the extension to multiple imputation classes is fairly straightforward, see Section 5.2.2.

In Section 6.2 we will define the imputed poststratified estimator and then establish its asymptotic normality. In Section 6.3, we will construct a jackknife variance estimator using the idea of adjusting the imputed values whenever a cluster is deleted. The jackknife variance estimator will be shown to be asymptotically consistent for the variance of the imputed poststratified estimator. We also present a linearization variance estimator obtained by linearizing the “correct” jackknife variance estimator. Finally, results of a limited simulation study investigating the effectiveness of the jackknife and linearized jackknife variance estimators will be presented in Section 6.4.

6.2 Imputed Poststratified Estimator

Since we have assumed that the poststrata of all sampled units are known, we can calculate the poststratified weights before we impute for the nonrespondents. Letting ${}_c\hat{w}_{hik}$ be the usual normalized poststratified weights,

$${}_c\hat{w}_{hik} = \frac{{}_cM}{M} \frac{\hat{w}_{hik}}{\hat{M}},$$

we select donors, y_{gjl} , with replacement from the respondents with probabilities ${}_c\hat{w}_{gjl} / \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik} a_{hik} {}_c\delta_{hik}$, i.e. for $(hik) \in s_m$, we impute $y_{hik}^* = y_{gjl}$. Letting E_* denote expectation with respect to the hot deck imputation, given the sample of respondents, s_r , then

$$\begin{aligned} E_*(y_{hik}^*) &= \frac{\sum_c \sum_{(gjl) \in s} {}_c\hat{w}_{gjl} a_{gjl} y_{gjl} {}_c\delta_{gjl}}{\sum_c \sum_{(gjl) \in s} {}_c\hat{w}_{gjl} a_{gjl} {}_c\delta_{gjl}} \\ &= \frac{\hat{S}}{\hat{T}} \end{aligned}$$

where

$$\hat{S} = \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik} a_{hik} y_{hik} {}_c\delta_{hik}$$

and

$$\hat{T} = \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik} a_{hik} {}_c\delta_{hik}.$$

The totals, $\hat{S}$ and $\hat{T}$, are the same as the totals used in mean imputation showing that the hot deck imputation is a stochastic counterpart to the mean imputation presented in Chapter 5.

The imputed poststratified estimator of the mean is

$$y_{PS}^I = \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik} a_{hik} y_{hik} {}_c\delta_{hik} + \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik} (1 - a_{hik}) y_{hik}^* {}_c\delta_{hik}$$

where y_{hik}^* are the imputed values. Under the weighted hot deck scheme, we have

$$E_*(y_{PS}^I) = \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik} a_{hik} y_{hik} {}_c\delta_{hik} + \sum_c \sum_{(hik) \in s} {}_c\hat{w}_{hik} (1 - a_{hik}) E_*(y_{hik}^*) {}_c\delta_{hik}$$

$$\begin{aligned}
&= \sum_c \sum_{(hsk) \in s} c \hat{w}_{hsk} a_{hsk} y_{hsk} c \delta_{hsk} + \sum_c \sum_{(hsk) \in s} c \hat{w}_{hsk} (1 - a_{hsk}) c \delta_{hsk} \frac{\hat{S}}{\hat{T}} \\
&= \frac{\hat{S}}{\hat{T}} \hat{t} \\
&= \frac{\hat{S}}{\hat{T}}
\end{aligned}$$

since

$$\begin{aligned}
\hat{t} &= \sum_c \sum_{(hsk) \in s} c \hat{w}_{hsk} c \delta_{hsk} \\
&= 1
\end{aligned}$$

by the benchmarking property of the poststratified weights. Assuming a uniform response mechanism, y_{hs}^I is approximately unbiased for the population mean, since

$$\begin{aligned}
E(\hat{S}) &= pE \left(\sum_c \sum_{(hsk) \in s} c \hat{w}_{hsk} y_{hsk} c \delta_{hsk} \right) \\
&= p\bar{Y}
\end{aligned}$$

and

$$\begin{aligned}
E(\hat{T}) &= pE \left(\sum_c \sum_{(hsk) \in s} c \hat{w}_{hsk} c \delta_{hsk} \right) \\
&= p.
\end{aligned}$$

Thus, using a ratio approximation

$$\begin{aligned}
E(\hat{y}_{PS}^I) &\approx \frac{p\bar{Y}}{p} \\
&= \bar{Y}.
\end{aligned}$$

We now show that the imputed poststratified estimator is asymptotically normally distributed. Noting that $E_*(\hat{y}_{PS}^I) = \hat{S}/\hat{T}$, we can write

$$\begin{aligned}
\hat{y}_{PS}^I - Y &= \hat{y}_{PS}^I - E_*(\hat{y}_{PS}^I) + E_*(\hat{y}_{PS}^I) - Y \\
&= \left(\hat{y}_{PS}^I - \frac{\hat{S}}{\hat{T}} \right) + \left(\frac{\hat{S}}{\hat{T}} - Y \right).
\end{aligned} \tag{6.2.1}$$

3 of/de 3

PM-1 3½" x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

Working with the first term in equation (6.2.1) and noting that $\bar{U} = 1$,

$$\begin{aligned}\bar{y}_{PS}^I - \frac{\hat{S}}{\hat{T}} &= \hat{S} + \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} (1 - a_{hik}) y_{hik}^* c \delta_{hik} - \frac{\hat{S}}{\hat{T}} \bar{U} \\ &= \hat{S} - \frac{\hat{S}}{\hat{T}} \hat{T} + \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) c \delta_{hik} \\ &= \hat{Z}^*\end{aligned}$$

where $\hat{Z}^* = \sum_c \sum_{(hik) \in s} c \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) c \delta_{hik}$. Therefore, we have

$$\bar{y}_{PS}^I - \bar{Y} = \frac{\hat{S}}{\hat{T}} - \bar{Y} + \hat{Z}^*. \quad (6.2.2)$$

Before establishing the asymptotic normality of the imputed poststratified estimator, we need additional regularity conditions:

C16 $n_T V_*(t^*) \rightarrow \sigma_z^2$, where $t^* = \sum_{(hik) \in s} \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T})$ and V_* denotes the variance with respect to the hot deck imputation given the sample of respondents, s_r ,

$$\text{C17 } \sum_{(hik) \in s} \hat{w}_{hik} |y_{hik} - \bar{Y}|^{2+\delta} = O_p(1)$$

and the following lemma from Rao and Shao (1991),

Lemma 6.1 (*Rao and Shao (1991)*)

Suppose that $\{X_n\}$ are random variables and $\{a_n\}$ are positive constants such that $X_n/a_n \rightarrow_d N(0, 1)$. Let $\{W_n\}$ be random variables, $\{b_n\}$ be positive constants and $\{Z_n\}$ be random variables such that

$$\frac{W_n}{b_n} \rightarrow_{d|Z_n} N(0, 1).$$

Assume also that X_n is a function of Z_n only. Then

$$c_n^{-1}(X_n + W_n) \rightarrow_d N(0, 1)$$

where $c_n^2 = a_n^2 + b_n^2$.

The asymptotic normality of the imputed poststratified estimator is now established in the following theorem.

Theorem 6.1 *Under conditions C4, C13, C14, C15, C16, C17 and the uniform response mechanism,*

$$\sqrt{n_T} (\bar{y}_{PS}^I - \bar{Y}) \longrightarrow_d N(0, c_n^2)$$

where $c_n^2 = \sigma_\epsilon^2/p^2 + b_n^2$ and

$$\begin{aligned} b_n^2 &= EV_*(\bar{y}_{PS}^I) \\ &= (1-p)E \left[\sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik}^2 {}_c \delta_{hik} \right] E \left[\sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik} {}_c \delta_{hik} (y_{hik} - \hat{S}/\hat{T})^2 \right]. \end{aligned}$$

Proof

From equation (6.2.2),

$$\sqrt{n_T} (\bar{y}_{PS}^I - \bar{Y}) = \sqrt{n_T} \left(\frac{\hat{S}}{\hat{T}} - \bar{Y} \right) + \sqrt{n_T} \hat{Z}^*.$$

In Theorem 5.1 we showed that

$$\sqrt{n_T} \left(\frac{\hat{S}}{\hat{T}} - \bar{Y} \right) \longrightarrow_p N(0, \sigma_\epsilon^2/p^2). \quad (6.2.3)$$

It remains to show

$$\sqrt{n_T} \hat{Z}^* \longrightarrow_{d|y_{obs}, a} N(0, \sigma_z^2)$$

where $y_{obs} = \{y_{hik}, (hik) \in s_r\}$ and $a = \{a_{hik}, (hik) \in s_m\}$. We have

$$\begin{aligned} \hat{Z}^* &= \sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) {}_c \delta_{hik} \\ &= \sum_{(hik) \in s} \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) + \\ &\quad \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \sum_{(hik) \in s} \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) {}_c \delta_{hik} \\ &= \sum_h \sum_i t_{hi}^* + \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \sum_h \sum_i {}_c t_{hi}^* \quad \text{say,} \end{aligned} \quad (6.2.4)$$

where

$${}_c t_{hi}^* = \sum_k \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) {}_c \delta_{hik}$$

and

$$t_{hi}^* = \sum_k \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}).$$

Working with the first term of (6.2.4), by condition C16

$$\begin{aligned} \frac{1}{n_T} \sum_h \sum_i V_*(n_T t_{hi}^*) &= n_T \sum_h \sum_i V_*(t_{hi}^*) \\ &\rightarrow \sigma_z^2 \text{ as } n_T \rightarrow \infty. \end{aligned}$$

In order to show $\sqrt{n_T} (\sum_h \sum_i i_{hi}^*) \rightarrow_{d|y_{obs,a}} N(0, \sigma_z^2)$, it suffices to show

$$\frac{1}{n_T} \sum_h \sum_i E_* |n_T t_{hi}^*|^{2+\delta} = O(1)$$

since $E_*(t_{hi}^*) = 0$. First consider

$$\begin{aligned} E_* |t_{gj}^*|^{2+\delta} &= \frac{\sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik} a_{hik} {}_c \delta_{hik} \left| \sum_l \tilde{w}_{gjl} (1 - a_{gjl}) (y_{hik} - \hat{S}/\hat{T}) \right|^{2+\delta}}{\hat{T}} \\ &\leq \left(\sum_l \tilde{w}_{gjl} \right)^{2+\delta} \frac{\sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik} {}_c \delta_{hik} \left| y_{hik} - \hat{S}/\hat{T} \right|^{2+\delta}}{\hat{T}} \end{aligned}$$

since a_{gjl} is an indicator variable. Now,

$$\begin{aligned} &\sum_c \sum_{(hik) \in s} {}_c \hat{w}_{hik} {}_c \delta_{hik} \left| y_{hik} - \hat{S}/\hat{T} \right|^{2+\delta} \\ &= \sum_{(hik) \in s} \hat{w}_{hik} \left| y_{hik} - \hat{S}/\hat{T} \right|^{2+\delta} + \\ &\quad \sum_c \left(\frac{{}_c M}{{}_c \hat{M}} - 1 \right) \sum_{(hik) \in s} \hat{w}_{hik} {}_c \delta_{hik} \left| y_{hik} - \hat{S}/\hat{T} \right|^{2+\delta} \\ &= A_1 + A_2 \text{ say.} \end{aligned}$$

Noting that $\tilde{S}/\tilde{T} - \bar{Y} \rightarrow_p 0$, see equation (5.2.9), we have

$$A_1 = \sum_{(hik) \in s} \hat{w}_{hik} \left| y_{hik} - \tilde{S}/\tilde{T} \right|^{2+\delta}$$

$$\begin{aligned}
&\leq 2^{1+\delta} \left(\sum_{(hik) \in \mathcal{S}} \hat{w}_{hik} |y_{hik} - \bar{Y}|^{2+\delta} + \sum_{(hik) \in \mathcal{S}} \hat{w}_{hik} |\bar{Y} - \hat{S}/\hat{T}|^{2+\delta} \right) \\
&= O_p(1)
\end{aligned}$$

by C17. Also,

$$\sum_{(hik) \in \mathcal{S}} \hat{w}_{hik} {}_c\delta_{hik} |y_{hik} - \hat{S}/\hat{T}|^{2+\delta} \leq \sum_{(hik) \in \mathcal{S}} \hat{w}_{hik} |y_{hik} - \hat{S}/\hat{T}|^{2+\delta}$$

since ${}_c\delta_{hik}$ is either 0 or 1. That is,

$$\sum_{(hik) \in \mathcal{S}} \hat{w}_{hik} {}_c\delta_{hik} |y_{hik} - \hat{S}/\hat{T}|^{2+\delta} = O_p(1).$$

Therefore, since ${}_cM/{}_c\hat{M} \rightarrow_p 1$ we have

$$\begin{aligned}
A_2 &= \sum_c \left(\frac{{}_cM}{{}_c\hat{M}} - 1 \right) \sum_{(hik) \in \mathcal{S}} \hat{w}_{hik} {}_c\delta_{hik} |y_{hik} - \hat{S}/\hat{T}|^{2+\delta} \\
&= \sum_c \left(\frac{{}_cM}{{}_c\hat{M}} - 1 \right) O_p(1) \\
&\rightarrow_p 0.
\end{aligned}$$

Thus,

$$\sum_c \sum_{(hik) \in \mathcal{S}} {}_c\hat{w}_{hik} {}_c\delta_{hik} |y_{hik} - \hat{S}/\hat{T}|^{2+\delta} = O_p(1) \quad (6.2.5)$$

and

$$\begin{aligned}
E_* |t_{gj}^*|^{2+\delta} &\leq \left(\sum_l \hat{w}_{gjl} \right)^{2+\delta} \frac{O_p(1)}{\hat{T}} \\
&\leq \left(\sum_l \hat{w}_{gjl} \right)^{2+\delta} O_p(1)
\end{aligned}$$

since $\hat{T}^{-1} \rightarrow_p p^{-1}$; see equation (5.2.5). Now, this gives

$$\begin{aligned}
\frac{1}{n_T} \sum_h \sum_i E_* |n_T t_{hi}^*|^{2+\delta} &= n_T^{1+\delta} \sum_h \sum_i E_* |t_{hi}^*|^{2+\delta} \\
&\leq n_T^{1+\delta} \sum_h \sum_i \left(\sum_l \hat{w}_{hik} \right)^{2+\delta} O_p(1) \\
&\leq \left(n_T \max_{h,i,k} m_{hi} \hat{w}_{hik} \right)^{2+\delta} O_p(1) \\
&= O_p(1)
\end{aligned} \quad (6.2.6)$$

by C4 . Therefore, by the Central Limit Theorem, Lemma 2.1, we have

$$\sqrt{n_T} \sum_h \sum_i t_{hi}^* \longrightarrow_{d|y_{obs},a} N(0, \sigma_z^2).$$

It remains to show that the second term in equation (6.2.4) converges to 0 in probability. We know that ${}_cM/{}_c\hat{M} \longrightarrow_p 1$, so we would like to show that $\sqrt{n_T} \cdot t^* = O_p(1)$ for all c where ${}_c t^* = \sum_{(hik)\epsilon s} \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) {}_c \delta_{hik}$. Noting that $E_*({}_c t_{hi}^*) = 0$, it suffices to show $n_T V_*({}_c t^*) = O_p(1)$ for all c . We have

$$\begin{aligned} n_T V_*({}_c t^*) &= n_T \sum_h \sum_i V_*({}_c t_{hi}^*) \\ &= n_T \sum_h \sum_i E_*({}_c t_{hi}^*)^2 \\ &\leq n_T \sum_h \sum_i \left(\sum_k \hat{w}_{hik} \right)^2 O_p(1) \\ &\leq \left(n_T \max_{h,i,k} m_{hi} \hat{w}_{hik} \right)^2 O_p(1) \\ &= O_p(1) \end{aligned} \tag{6.2.7}$$

if we can show

$$E_*({}_c t_{hi}^*)^2 \leq \left(\sum_k \hat{w}_{hik} \right)^2 O_p(1).$$

Now,

$$\begin{aligned} E_*({}_c t_{gj}^*)^2 &= \frac{\sum_c \sum_{(hik)\epsilon s} {}_c \hat{w}_{hik} a_{hik} {}_c \delta_{hik} \left(\sum_l \hat{w}_{gjl} (1 - a_{gjl}) (y_{hik} - \hat{S}/\hat{T}) {}_c \delta_{gjl} \right)^2}{\hat{T}} \\ &\leq \left(\sum_l \hat{w}_{gjl} \right)^2 \frac{\sum_c \sum_{(hik)\epsilon s} {}_c \hat{w}_{hik} {}_c \delta_{hik} (y_{hik} - \hat{S}/\hat{T})^2}{\hat{T}}. \end{aligned}$$

Noting that for $\delta > 0$

$$(y_{hik} - \hat{S}/\hat{T})^2 \leq 1 + |y_{hik} - \hat{S}/\hat{T}|^{2+\delta}$$

we have

$$\sum_c \sum_{(hik)\epsilon s} {}_c \hat{w}_{hik} {}_c \delta_{hik} (y_{hik} - \hat{S}/\hat{T})^2$$

$$\begin{aligned}
&\leq \sum_c \sum_{(hik) \in s} {}_c \tilde{w}_{hik} {}_c \delta_{hik} + \sum_c \sum_{(hik) \in s} {}_c \tilde{w}_{hik} {}_c \delta_{hik} |y_{hik} - \tilde{S}/\tilde{T}|^{2+\delta} \\
&= 1 + \sum_c \sum_{(hik) \in s} {}_c \tilde{w}_{hik} {}_c \delta_{hik} |y_{hik} - \tilde{S}/\tilde{T}|^{2+\delta} \\
&= O_p(1)
\end{aligned}$$

by equation (6.2.5). It now follows that

$$E_*({}_c t_{hi}^*)^2 \leq \left(\sum_k \tilde{w}_{hik} \right)^2 O_p(1). \quad (6.2.8)$$

Thus we have $n_T V_*({}_c t^*) = O_p(1)$ and Lemma (2.8) gives

$$\sum_c \left(\frac{{}_c M}{\hat{{}_c M}} - 1 \right) \sqrt{n_T} {}_c t^* \longrightarrow_p 0.$$

Therefore,

$$\begin{aligned}
\sqrt{n_T} \hat{Z}^* &= \sqrt{n_T} \sum_h \sum_i t_{hi}^* + \sum_c \left(\frac{{}_c M}{\hat{{}_c M}} - 1 \right) \sqrt{n_T} {}_c t^* \\
&\longrightarrow_{d|y_{obs}, a} N(0, \sigma_z^2)
\end{aligned} \quad (6.2.9)$$

or

$$\hat{Z}^* \longrightarrow_{d|y_{obs}, a} N(0, \sigma_z^2/n_T).$$

This is equivalent to

$$\hat{Z}^* \longrightarrow_{d|y_{obs}, a} N(0, V_*(\hat{Z}^*))$$

since

$$\lim_{n_T \rightarrow \infty} \frac{V_*(\hat{Z}^*)}{(\sigma_z^2/n_T)} = \lim_{n_T \rightarrow \infty} \left(\frac{n_T V_*(\hat{Z}^*)}{\lim_{n_T \rightarrow \infty} n_T V_*(\hat{Z}^*)} \right) = 1.$$

Now, letting $b_n^2 = EV_*(\hat{Z}^*)$ gives

$$\begin{aligned}
\frac{\hat{Z}^*}{b_n} &= \frac{\hat{Z}^*}{\sqrt{V_*(\hat{Z}^*)}} \frac{\sqrt{V_*(\hat{Z}^*)}}{b_n} \\
&\longrightarrow_{d|y_{obs}, a} N(0, 1)
\end{aligned} \quad (6.2.10)$$

or

$$\hat{Z}^* \longrightarrow_{d|y_{obs}, a} N(0, b_n^2).$$

Finally applying Lemma (6.1) to equations (6.2.3) and (6.2.10), we have

$$\begin{aligned} \sqrt{n_T}(\bar{y}_{PS}^I - \bar{Y}) &= \sqrt{n_T} \frac{\hat{S}}{\hat{T}} + \sqrt{n_T} \hat{Z}^* \\ &\longrightarrow_d N(0, c_n^2) \end{aligned}$$

where $c_n^2 = \sigma_e^2/p^2 + b_n^2$. ■

6.3 Jackknife Variance Estimation

Turning to variance estimation, in order to construct a jackknife variance estimator we first define the adjusted imputed values, when the $(gj)^{th}$ cluster has been deleted, as

$$z_{hik(gj)}^* = y_{hik}^* + \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}}$$

where

$$\begin{aligned} \hat{S}_{(gj)} &= \sum_c \sum_{(hik) \in s} {}_c \tilde{w}_{hik(gj)} a_{hik} y_{hik} {}_c \delta_{hik}, \\ \hat{T}_{(gj)} &= \sum_c \sum_{(hik) \in s} {}_c \tilde{w}_{hik(gj)} a_{hik} {}_c \delta_{hik} \end{aligned}$$

and ${}_c \tilde{w}_{hik(gj)}$ are the poststratified jackknife weights. The adjusted imputed values are simply the original imputed values adjusted by an amount $\hat{S}_{(gj)}/\hat{T}_{(gj)} - \hat{S}/\hat{T}$ when the $(gj)^{th}$ cluster is deleted. Noting that $E_*(y_{hik}^*) = \hat{S}/\hat{T}$, it follows that

$$E_*(z_{hik(gj)}^*) = \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}}.$$

Therefore, we see that on average we impute the properly adjusted mean of the respondents. Using the adjusted imputed values, the adjusted poststratified estimator

when the $(gj)^{th}$ cluster has been deleted is

$$\begin{aligned}\bar{y}_{PS(gj)}^{Ia} &= \sum_c \sum_{(hik) \in s} c \tilde{w}_{hik(gj)} a_{hik} y_{hik} c \delta_{hik} + \sum_c \sum_{(hik) \in s} c \tilde{w}_{hik(gj)} (1 - a_{hik}) \tilde{y}_{hik(gj)}^* c \delta_{hik} \\ &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} \hat{U}_{(gj)} + \hat{Z}_{(gj)}^* \\ &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} + \hat{Z}_{(gj)}^*\end{aligned}$$

where

$$\hat{U}_{(gj)} = \frac{1}{1} \sum_c \sum_{(hik) \in s} c \tilde{w}_{hik(gj)} c \delta_{hik}$$

and

$$\hat{Z}_{(gj)}^* = \sum_c \sum_{(hik) \in s} c \tilde{w}_{hik(gj)} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) c \delta_{hik}.$$

The jackknife variance estimator is then defined as

$$v_J(\bar{y}_{PS}^I) = \sum_g \frac{n_g - 1}{n_g} \sum_j (\bar{y}_{PS(gj)}^{Ia} - \bar{y}_{PS}^I)^2.$$

Asymptotic consistency of $v_J(\bar{y}_{PS}^I)$ is now established in the following theorem.

Theorem 6.2 *Under conditions C4, C13, C14, C15, C16, C17 and the uniform response mechanism,*

$$n_T(v_J(\bar{y}_{PS}^I) - V(\bar{y}_{PS}^I)) \rightarrow_p 0.$$

Proof

First, we note that $\bar{y}_{PS}^I$ can be written as

$$\begin{aligned}\bar{y}_{PS}^I &= \hat{S} + \sum_c \sum_{(hik) \in s} c \tilde{w}_{hik} (1 - a_{hik}) y_{hik}^* c \delta_{hik} \\ &= \hat{S} + \sum_c \sum_{(hik) \in s} c \tilde{w}_{hik} (1 - a_{hik}) c \delta_{hik} (y_{hik}^* - \hat{S}/\hat{T} + \hat{S}/\hat{T}) \\ &= \frac{\hat{S}}{\hat{T}} + \hat{Z}^*.\end{aligned}$$

Therefore,

$$\begin{aligned}\bar{y}_{PS(gj)}^I - \bar{y}_{PS}^I &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \hat{Z}_{(gj)}^* + \frac{\hat{S}}{\hat{T}} - \hat{Z}^* \\ &= \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} + \hat{Z}_{(gj)}^* - \hat{Z}^*\end{aligned}$$

and the jackknife variance estimator becomes

$$\begin{aligned}v_J(\bar{y}_{PS}^I) &= \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} \right)^2 + \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\hat{Z}_{(gj)}^* - \hat{Z}^* \right)^2 + \\ &\quad 2 \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} \right) \left(\hat{Z}_{(gj)}^* - \hat{Z}^* \right).\end{aligned}\quad (6.3.1)$$

We establish the consistency of $v_J(\bar{y}_{PS}^I)$ by showing

$$n_T(v_J(\bar{y}_{PS}^I) - V(\bar{y}_{PS}^I)) \longrightarrow_p 0.$$

The variance of $\bar{y}_{PS}^I$ is

$$\begin{aligned}V(\bar{y}_{PS}^I) &= VE_*(\bar{y}_{PS}^I) + EV_*(\bar{y}_{PS}^I) \\ &= V\left(\frac{\hat{S}}{\hat{T}}\right) + EV_*(\hat{Z}^*).\end{aligned}$$

In Theorem 5.3, we have shown that the first term in the right hand side of equation (6.3.1) is consistent for $V(\hat{S}/\hat{T})$, so it remains to show

$$n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\hat{Z}_{(gj)}^* - \hat{Z}^* \right) \longrightarrow_p n_T EV_*(\hat{Z}^*) \quad (6.3.2)$$

and

$$2n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} \right) \left(\hat{Z}_{(gj)}^* - \hat{Z}^* \right) \longrightarrow_p 0. \quad (6.3.3)$$

The proof of equation (6.3.3) is given in the appendix at the end of this chapter.

To show equation (6.3.2), we note that

$$\hat{Z}_{(gj)}^* - \hat{Z}^* = \sum_c \frac{{}_c M}{M} \left(\frac{{}_c t_{(gj)}^*}{{}_c \hat{M}_{(gj)}} - \frac{{}_c t^*}{{}_c \hat{M}} \right)$$

$$\begin{aligned}
&= \sum_c \frac{cM}{c\hat{M}} \frac{1}{c\hat{M}} \left({}_c t_{(gj)}^* - {}_c t^* - \frac{{}_c t^*}{c\hat{M}} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right) - \\
&\quad \sum_c \frac{cM}{c\hat{M}} \frac{1}{c\hat{M}_{gj}} \left({}_c t_{(gj)}^* - {}_c t^* - \frac{{}_c t^*}{c\hat{M}} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right) \quad (6.3.4)
\end{aligned}$$

where

$$\begin{aligned}
{}_c t^* &= \sum_{(hik) \in \mathcal{S}} \hat{w}_{hik} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) {}_c \delta_{hik}, \\
{}_c t_{(gj)}^* &= \sum_{(hik) \in \mathcal{S}} \hat{w}_{hik(gj)} (1 - a_{hik}) (y_{hik}^* - \hat{S}/\hat{T}) {}_c \delta_{hik}
\end{aligned}$$

and ${}_c x_{gj}$ lies between ${}_c \hat{M}_{(gj)}$ and ${}_c \hat{M}$. The second term in the right hand side of equation (6.3.4) can be handled using techniques in the proof of Theorem 5.3 and thus we will consider only the first term in the right hand side of equation (6.3.4).

We now have

$$\begin{aligned}
&n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\hat{Z}_{(gj)}^* - \hat{Z}^* \right)^2 \\
&= n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\sum_c \frac{cM}{c\hat{M}} ({}_c t_{(gj)}^* - {}_c t^*) \right)^2 + \\
&\quad n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\sum_c \frac{cM}{c\hat{M}} \frac{{}_c t^*}{c\hat{M}} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right)^2 - \\
&\quad 2n_T \text{ cross-product terms.} \quad (6.3.5)
\end{aligned}$$

We next show that the first term in the right hand side of equation (6.3.5) is a consistent estimator for $EV_*(\bar{y}_{PS}^I)$. The second term and the cross-product terms are shown to converge to zero in probability in the appendix at the end of this chapter. Now, we can rewrite the first term in the right hand side of equation (6.3.5) as

$$\begin{aligned}
&n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left[\sum_c \frac{cM}{c\hat{M}} ({}_c t_{(gj)}^* - {}_c t^*) \right]^2 \\
&= n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left[(t_{(gj)}^* - t^*) + \sum_c \left(\frac{cM}{c\hat{M}} - 1 \right) ({}_c t_{(gj)}^* - {}_c t^*) \right]^2
\end{aligned}$$

$$\begin{aligned}
&= n_T \sum_g \frac{n_g - 1}{n_g} \sum_j (t_{(gj)}^* - t^*)^2 + \\
&\quad n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left[\sum_c \left(\frac{cM}{c\hat{M}} - 1 \right) ({}_c t_{(gj)}^* - {}_c t^*) \right]^2 + \\
&\quad 2n_T \sum_g \frac{n_g - 1}{n_g} \sum_j (t_{(gj)}^* - t^*) \left[\sum_c \left(\frac{cM}{c\hat{M}} - 1 \right) ({}_c t_{(gj)}^* - {}_c t^*) \right] \\
&= n_T v_1 + n_T v_2 + 2n_T v_3 \quad \text{say,}
\end{aligned}$$

where $t_{(gj)}^* = \sum_c {}_c t_{(gj)}^*$ and $t^* = \sum_c {}_c t^*$. Using the definition of the jackknife weights, it is straight forward to show

$$t_{(gj)}^* - t^* = \frac{n_g}{n_g - 1} (t_g^* - t_{gj}^*)$$

where $t_{gj}^* = \sum_k \hat{w}_{gjk} (1 - a_{gjk}) (y_{gjk}^* - \hat{S}/\hat{T})$ and $t_g^* = (1/n_g) \sum_j t_{gj}^*$.

Putting this expression for $t_{(gj)}^* - t^*$ in v_1 gives

$$v_1 = \sum_g \frac{n_g}{n_g - 1} \sum_j (t_g^* - t_{gj}^*)^2.$$

Noting that

$$\begin{aligned}
E_* v_1 &= \sum_g \sum_j V_*(t_{gj}^*) \\
&= V_*(t^*) \approx V_*(\hat{Z}^*) = V_*(\hat{y}_{PS}^I)
\end{aligned}$$

see equation (6.2.10), we can appeal to the Law of Large Numbers, Lemma 2.2, to show

$$n_T v_1 \longrightarrow_p n_T V_*(\hat{y}_{PS}^I).$$

Following the proof of Theorem 2.3, noting that $E_*(t_{gj}^*) = 0$ it suffices to show

$$\frac{1}{n_T} \sum_g \sum_j E_* |n_T t_{gj}^*|^{2+b} = O_p(1).$$

This was shown in the proof of Theorem 6.1, see equation (6.2.6). Again, following the proof of Theorem 2.3, Lemma 2.2 gives

$$n_T v_1 \longrightarrow_p n_T V_*(\hat{y}_{PS}^I).$$

This result, along with

$$n_T(V_*(\bar{y}_{PS}^I) - EV_*(\bar{y}_{PS}^I)) \longrightarrow_p 0$$

gives

$$n_T(v_1 - n_T EV_*(\bar{y}_{PS}^I)) \longrightarrow_p 0$$

as required.

Next we show that $n_T v_2 \longrightarrow_p 0$. Using the Schwarz inequality,

$$\begin{aligned} n_T v_2 &\leq n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \dot{M}} - 1 \right)^2 \right] \left[\sum_c ({}_c t_{(gj)}^* - {}_c t^*)^2 \right] \\ &= \sum_c \left(\frac{{}_c M}{{}_c \dot{M}} - 1 \right)^2 n_T v_L({}_c t^*) \\ &\longrightarrow_p 0 \end{aligned}$$

since from Lemma 2.7,

$$\left(\frac{{}_c M}{{}_c \dot{M}} - 1 \right) \longrightarrow_p 0,$$

for all c , and $n_T v_L({}_c t^*) = O_p(1)$ since

$$n_T v_L({}_c t^*) \longrightarrow_p n_T V_*({}_c t^*) = O_p(1)$$

from equation (6.2.7). An application of the Schwarz inequality implies

$$2n_T v_3 \longrightarrow_p 0.$$

Combining these results gives

$$\begin{aligned} n_T \sum_g \frac{n_g - 1}{n_g} \sum_j (\hat{Z}_{(gj)}^* - \hat{Z}^*)^2 &= n_T v_1 + n_T v_2 + 2n_T v_3 \\ &\longrightarrow_p n_T EV_*(\bar{y}_{PS}^I). \end{aligned}$$

Hence we have

$$n_T v_J(\bar{y}_{PS}^I) = n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} \right)^2 +$$

$$\begin{aligned}
& n_T \sum_g \frac{n_g - 1}{n_g} \sum_j (\dot{Z}_{(gj)}^* - \dot{Z}^*)^2 + \\
& 2n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\frac{\dot{S}_{(gj)}}{\dot{T}_{(gj)}} - \frac{\dot{S}}{\dot{T}} \right) (\dot{Z}_{(gj)}^* - \dot{Z}^*) \\
& \longrightarrow_p n_T V\left(\frac{\dot{S}}{\dot{T}}\right) + n_T EV_*(y_{PS}^I) = n_T V(y_{PS}^I).
\end{aligned}$$

That is,

$$n_T(v_J(y_{PS}^I) - V(y_{PS}^I)) \longrightarrow_p 0$$

as required. ■

A linearization variance estimator can be obtained by using a Taylor approximation to the jackknife variance estimator. The linearized jackknife variance estimator is

$$v_{JL}(y_{PS}^I) = v(r_{hi})$$

where

$$r_{hi} = \frac{1}{\dot{T}} \sum_k {}_c \hat{w}_{hik} {}_c \epsilon_{hik}^* {}_c \delta_{hik} + \sum_k {}_c \hat{w}_{hik} {}_c \delta_{hik} \left[(1 - a_{hik})(y_{hik}^* - \dot{S}/\dot{T}) - \frac{{}_c t^*}{{}_c \dot{M}} \right]$$

with

$$\begin{aligned}
{}_c \epsilon_{hik}^* &= a_{hik}(y_{hik} - \dot{S}/\dot{T}) - {}_c \bar{r} \\
{}_c \bar{r} &= \frac{1}{{}_c \dot{M}} \sum_{(hik), s} \hat{w}_{hik} a_{hik} {}_c \delta_{hik} (y_{hik} - \dot{S}/\dot{T})
\end{aligned}$$

and

$${}_c t^* = \sum_{(hik), s} \hat{w}_{hik} (1 - a_{hik}) {}_c \delta_{hik} (y_{hik}^* - \dot{S}/\dot{T}).$$

As a result of the consistency of the jackknife variance estimator, the linearized jackknife variance estimator is also consistent for $V(\bar{y}_{PS}^I)$.

6.4 Simulation Study

A limited simulation study was performed to evaluate the effectiveness of the jackknife and the linearized jackknife variance estimators for the poststratified estimator

under hot deck imputation. We used the same population and sampling scheme as in Section 5.4 except we used the weighted hot deck imputation detailed in Section 6.2, instead of mean imputation. As was the case under mean imputation, the “naive” jackknife under hot deck imputation treats the imputed values as true values.

Table 6.1 reports the results, average over 10,000 samples, for the 50 strata design. We see that v_J and v_{JL} perform well with relative biases less than 4% and error rates slightly higher than the nominal 5%. We note that v_J and v_{JL} are expected to behave similarly due to their asymptotic equivalence. Turning to v_{JN} , we see that v_{JN} severely underestimates the true variance of $\hat{Y}_{PS}$ with the difference increasing as the nonresponse rate increases. The error rates corresponding to v_{JN} are indicative of the underestimation as they are much larger than the nominal 5%. Turning to the average length of the confidence intervals, we see that the average length of the v_{JN} confidence intervals are constant while the lengths for the v_J and v_{JL} confidence intervals are increasing as the response rate decreases. This indicates that v_J and v_{JL} are tracking the increasing MSE while v_{JN} is remaining stable.

Results from the 100 strata design are presented in Table 6.2. The findings are similar to the 50 strata design. The correct jackknife variance estimators, v_J and v_{JL} , perform well while the naive jackknife severely underestimates the true variance of $\hat{Y}_{PS}$. We note, as expected, the discrepancy between v_J and v_{JL} is much smaller for the 100 strata design than for the 50 strata design. The behavior of the average length of the confidence intervals is similar for the 50 design case.

Table 6.1: Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Hot Deck Imputation (50 Strata Design)

Performance Measure	$v_J(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_{JN}(\hat{Y}_{PS})$
<u>$p = 0.9$</u>			
Relative Bias	-3.35 %	-3.90 %	-20.37 %
Error Rate	6.14 %	6.21 %	8.93 %
Lower Error Rate	2.66 %	2.71 %	3.99 %
Upper Error Rate	3.48 %	3.50 %	4.94 %
Average Length	6.07	6.05	5.51
<u>$p = 0.8$</u>			
Relative Bias	-2.92 %	-3.43 %	-33.47 %
Error Rate	6.09 %	6.18 %	11.68 %
Lower Error Rate	2.58 %	2.62 %	5.28 %
Upper Error Rate	3.51 %	3.56 %	6.40 %
Average Length	6.71	6.69	5.55
<u>$p = 0.7$</u>			
Relative Bias	-1.66 %	-2.13 %	-43.74 %
Error Rate	5.95 %	6.01 %	15.14 %
Lower Error Rate	2.10 %	2.13 %	6.64 %
Upper Error Rate	3.85 %	3.88 %	8.50 %
Average Length	7.38	7.36	5.58

Table 6.2: Relative Bias, Error Rates and Average Length of Confidence Intervals for The Poststratified Estimator under Hot Deck Imputation (100 Strata Design)

Performance Measure	$v_J(\hat{Y}_{PS})$	$v_{JL}(\hat{Y}_{PS})$	$v_{JN}(\hat{Y}_{PS})$
<u>$p = 0.9$</u>			
Relative Bias	-0.77 %	-0.90 %	-18.24 %
Error Rate	5.57 %	5.58 %	8.05 %
Lower Error Rate	2.36 %	2.36 %	3.50 %
Upper Error Rate	3.21 %	3.22 %	4.55 %
Average Length	4.27	4.27	3.87
<u>$p = 0.8$</u>			
Relative Bias	-0.92 %	-1.04 %	-32.05 %
Error Rate	5.65 %	5.67 %	10.73 %
Lower Error Rate	2.43 %	2.44 %	4.87 %
Upper Error Rate	3.22 %	3.23 %	5.86 %
Average Length	4.72	4.72	3.91
<u>$p = 0.7$</u>			
Relative Bias	-2.61 %	-2.71 %	-44.15 %
Error Rate	5.65 %	5.66 %	14.64 %
Lower Error Rate	2.39 %	2.39 %	6.54 %
Upper Error Rate	3.26 %	3.27 %	8.10 %
Average Length	5.21	5.20	3.94

6.5 Appendix

Proof of equation (6.3.3)

Letting

$$\hat{R}_{gj} = \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}}$$

and using equation (6.3.4), we can rewrite the left hand side of equation (6.3.3) as

$$\begin{aligned} 2n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \hat{R}_{gj} (\hat{Z}_{(gj)}^* - \hat{Z}^*) &= 2n_T \sum_g \sum_j \hat{R}_{gj} \left(\sum_c \frac{{}_c M}{{}_c \hat{M}} {}_c t_g^* \right) - \\ & 2n_T \sum_g \sum_j \hat{R}_{gj} \left(\sum_c \frac{{}_c M}{{}_c \hat{M}} {}_c t_{gj}^* \right) - \\ & 2n_T \sum_g \sum_j \hat{R}_{gj} \left(\sum_c \frac{{}_c M}{{}_c \hat{M}} \frac{{}_c t^*}{{}_c \hat{M}} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right) \\ &= 2n_T r_1 - 2n_T r_2 - 2n_T r_3 \quad \text{say.} \end{aligned} \quad (6.5.1)$$

Noting that $E_*(n_T r_1) = 0$, $E_*(n_T r_2) = 0$ and $E_*(n_T r_3) = 0$, we can appeal to the Law of Large Numbers, Lemma (2.2), with $\delta = 1$ to show that each term in equation (6.5.1) converges to zero in probability. Working with r_1 ,

$$\begin{aligned} & \frac{1}{n_T} E_* \left[\sum_g \sum_j \left(n_T^2 \hat{R}_{gj} \sum_c ({}_c M / {}_c \hat{M}) {}_c t_g^* \right)^2 \right] \\ &= \frac{1}{n_T} E_* \left[\sum_g \sum_j \left(n_T^2 (1/n_g) \sum_i \hat{R}_{gi} \sum_c ({}_c M / {}_c \hat{M}) {}_c t_{gj}^* \right)^2 \right] \\ &= n_T^3 \sum_g \sum_j \left((1/n_g) \sum_i \hat{R}_{gi} \right)^2 E_* \left[\sum_c ({}_c M / {}_c \hat{M}) {}_c t_{gj}^* \right] \\ &\leq n_T^3 \sum_g \sum_j \left(\sum_g \sum_i \hat{R}_{gi} \right)^2 \left[\sum_c ({}_c M / {}_c \hat{M})^2 \right] \left[\sum_c E_* ({}_c t_{gj}^*)^2 \right] \\ &\leq C n_T^2 \max_{g,j,c} E_* ({}_c t_{gj}^*)^2 n_T \left(\sum_g \sum_i \hat{R}_{gi} \right)^2 \left[\sum_c ({}_c M / {}_c \hat{M})^2 \right] \\ &= O_p(1) \end{aligned}$$

by C4 and equation (6.2.8). Turning to r_2 , we have

$$\begin{aligned}
& \frac{1}{n_T} E_* \left[\sum_g \sum_j \left(n_T^2 \hat{R}_{gj} \sum_c ({}_c M / {}_c \hat{M}) {}_c t_{gj}^* \right) \right]^2 \\
&= n_T^3 \sum_g \sum_j \hat{R}_{gj}^2 E_* \left[\sum_c ({}_c M / {}_c \hat{M}) {}_c t_{gj}^* \right]^2 \\
&= n_T^3 \sum_g \sum_j \hat{R}_{gj}^2 \left[\sum_c ({}_c M / {}_c \hat{M})^2 \right] \left[\sum_c E_* ({}_c t_{gj}^{*2}) \right] \\
&\leq C n_T^2 \max_{g,j,c} E_* ({}_c t_{gj}^*)^2 n_T \sum_g \sum_j \hat{R}_{gj}^2 \left[\sum_c ({}_c M / {}_c \hat{M})^2 \right] \\
&= O_p(1).
\end{aligned}$$

Finally, working with r_3 ,

$$\begin{aligned}
& \frac{1}{n_T} E_* \left[\sum_g \frac{n_g - 1}{n_g} \sum_j \left(n_T^2 \hat{R}_{gj} \left\{ \sum_c ({}_c M / {}_c \hat{M}) ({}_c t^* / {}_c \hat{M}) ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right\} \right) \right]^2 \\
&= n_T^3 \sum_g \frac{n_g - 1}{n_g} \sum_j \hat{R}_{gj}^2 E_* \left[\sum_c ({}_c M / {}_c \hat{M}) ({}_c t^* / {}_c \hat{M}) ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right] \\
&\leq n_T^3 \sum_g \frac{n_g - 1}{n_g} \sum_j \hat{R}_{gj}^2 \left[\sum_c ({}_c M / {}_c \hat{M})^2 \right] \left[\sum_c ({}_c \hat{M}_{(gj)} - {}_c \hat{M})^2 \right] \left[\sum_c E_* ({}_c t^* / {}_c \hat{M})^2 \right] \\
&\leq C n_T^3 \max_{g,j,c} ({}_c \hat{M}_{(gj)} - {}_c \hat{M})^2 \sum_g \frac{n_g - 1}{n_g} \sum_j \hat{R}_{gj}^2 \left[\sum_c ({}_c M / {}_c \hat{M})^2 \right] \sum_c (1 / {}_c \hat{M})^2 V_* ({}_c t^*) \\
&= O_p \left(\frac{1}{n_T} \right)
\end{aligned}$$

since C , the number of poststrata, is finite, $\max_{g,j,c} ({}_c \hat{M}_{(gj)} - {}_c \hat{M})^2 = O_p(n_T^{-2})$, from equation (3.2.7), $n_T V_* ({}_c t^*) = O_p(1)$, from equation (6.2.7), and

$$n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \hat{R}_{gj}^2 = O_p(1).$$

Proof of equation (6.3.5)

We wish to show

$$n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\sum_c \frac{{}_c M}{{}_c \hat{M}} \frac{{}_c t^*}{{}_c \hat{M}} ({}_c \hat{M}_{(gj)} - {}_c \hat{M}) \right)^2 \rightarrow_p 0.$$

By the Schwarz inequality,

$$\begin{aligned}
 & n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\sum_c \frac{{}_c M}{{}_c \dot{M}} \frac{{}_c t^*}{{}_c \dot{M}} ({}_c \dot{M}_{(gj)} - {}_c \dot{M}) \right)^2 \\
 & \leq n_T \sum_g \frac{n_g - 1}{n_g} \sum_j \left[\sum_c \left(\frac{{}_c M}{{}_c \dot{M}} \right)^2 \left(\frac{{}_c t^*}{{}_c \dot{M}} \right)^2 \right] \left[\sum_c ({}_c \dot{M}_{(gj)} - {}_c \dot{M})^2 \right] \\
 & = \sum_c \left(\frac{{}_c M}{{}_c \dot{M}} \right)^2 \left(\frac{{}_c t^*}{{}_c \dot{M}} \right)^2 n_T v_L({}_c \dot{M}) \\
 & \xrightarrow{p} 0
 \end{aligned}$$

since $n_T v_L({}_c \dot{M}) = O_p(1)$, ${}_c M / {}_c \dot{M} \xrightarrow{p} 1$, and ${}_c t^* \xrightarrow{p} 0$ because $n_T V_*({}_c t^*) = O_p(1)$, see equation (6.2.7). The cross product term in equation (6.3.5) converges to zero in probability by appealing to the Schwarz inequality.

Chapter 7

Conclusions and Future Research

7.1 Conclusions

The commonly employed poststratified estimator is systematically studied along with the corresponding jackknife variance estimator. We first consider the complete data case, then proceed to the missing data case. Weighting adjustments are used to compensate for unit nonresponse, while for item nonresponse we consider two imputation methods, mean imputation and hot deck imputation. In each situation, we study the large sample properties of the poststratified estimator and its jackknife variance estimator. We also propose a new linearization variance estimator obtained by using a Taylor approximation to the jackknife variance estimator.

In the full response case, we have established the asymptotic normality of the poststratified estimator and the asymptotic consistency of the corresponding jackknife variance estimator. These results are derived under weaker regularity conditions than those given in Krewski and Rao (1981) for smooth functions of means. Our regularity conditions are similar to those of Scott and Wu (1981) for ratio and regression estimators under simple random sampling. Equivalence of the jackknife variance estimator and a robust linearization variance estimator are also established. This equivalence is shown to hold to higher order terms in the important special

case of two sampled clusters per stratum, thus generalizing the result of Rao and Wu (1985) to poststratification.

It has been argued by Holt and Smith (1979) and others, that inference based on the poststratified estimator should be made conditional on the realized sample of units. Using a model-based approach, Valliant (1993) has shown that the jackknife, balanced repeated replication and a robust linearization variance estimator correctly estimate the model variance of the poststratified estimator, conditional on the realized sample. These results are extended to the generalized regression (GREG) estimator used in the context of benchmarking to the marginal totals of a two-way table.

In the case of weighting adjustment for unit nonresponse with a single weighting class, both the poststratified and the GREG estimators reduce to the situation where only the respondents are used for estimation. Asymptotic normality of the estimators and consistency of the corresponding jackknife variance estimators readily follow from the complete data case. With multiple weighting classes, the estimators do not simplify but large sample properties can be obtained along the lines of the single weighting class case.

Turning to imputation for item nonresponse, it is well known that treating imputed values as true values at the stage of variance estimation can lead to severe underestimation of the true variance when the nonresponse rate is appreciable. Using the idea of adjusting the imputed values whenever a cluster is deleted (see Rao and Shao, 1992, Rao and Sitter, 1993, Rao, 1996), we construct "correct" jackknife variance estimators for the poststratified estimator under mean imputation and hot deck imputation. Large sample properties of the poststratified estimator and the corresponding jackknife variance estimator are established under both imputation methods.

Since the jackknife can become computationally expensive, we have derived linearization variance estimators by using a Taylor approximation to the jackknife

variance estimators. These linearized jackknife variance estimators retain the desirable properties of the jackknife. A drawback of obtaining linearization variance estimators in this way is that the method is only valid for designs in which the jackknife is valid. This is not seen as a major disadvantage as the stratified multi-stage design studied here is commonly used in large-scale surveys. Empirical results indicate that for designs with as few as 50 strata, the linearized jackknife performs as well as the true jackknife variance estimator.

7.2 Future Research

While it is hoped that the results obtained in this thesis are useful in practice, there are many areas in which further research could prove profitable. We now discuss two such areas.

Imputation

In this thesis we considered only mean imputation and hot deck imputation for the poststratified estimator. Rao and Sitter (1993) and Rao (1996) have used the Rao-Shao jackknife method to construct jackknife variance estimators and linearized jackknife variance estimators for the basic estimator under other imputation methods such as ratio, ratio hot deck, regression and regression hot deck. These results should be extended to the poststratified estimator. Also, results under all of the above mentioned imputation methods should be obtained for the GREG estimator for both indicator auxiliary variables and continuous auxiliary variables.

Conditional Analysis

While the poststratified and the GREG estimator were shown to be model unbiased, in the design-based framework both are conditionally biased when conditioned on the estimated auxiliary variable totals, $\hat{X}$, which may be regarded as ancillary.

Rao and Lui (1992), Rao (1994a) and Casady and Valliant (1993) proposed the optimal estimator

$$\hat{Y}_{opt} = \hat{Y} + \hat{\beta}_{opt}(\mathbf{X} - \hat{\mathbf{X}})$$

where $\hat{\beta}_{opt} = \hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xy}$ with $\hat{\Sigma}_{xx}$ and $\hat{\Sigma}_{xy}$ being unbiased estimators of Σ_{xx} and Σ_{xy} which respectively denote the $p \times p$ covariance matrix of $\hat{\mathbf{X}}$ and the p -vector of covariances, $cov(\hat{Y}, \hat{\mathbf{X}}_t)$. The optimal estimator is approximately conditionally unbiased given $\hat{\mathbf{X}}$. Neither Rao (1994a) nor Casady and Valliant (1993) address variance estimation although Casady and Valliant note that the methods of Särndal, Swensson and Wretman (1989) can be used to obtain a design-based variance estimator. It is felt that the jackknife variance estimator would be a conditionally correct variance estimator and the resulting linearized jackknife variance estimator would also be conditionally valid. Lui (1992) obtained a variance estimator using asymptotic conditional analysis. Comparison of the linearized jackknife variance estimator with Lui's variance estimator would also be useful.

References

- Bartle, R.G. (1976). *The Elements of Real Analysis* (Second Edition), New York: John Wiley and Sons, Inc.
- Beebakhee, R.M. (1995). A Comparison of Two Variance Estimation Techniques: The Jackknife Method and The Linearized Jackknife Method, Methodology Branch working paper HSMD-95-005-E, Ottawa: Statistics Canada.
- Bickel, P.J. and Doksom, K.A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*, California: Holden-Day, Inc.
- Billingsley, P. (1986). *Probability and Measure* (Second Edition), New York: John Wiley and Sons, Inc.
- Casady, R.J. and Valliant, R. (1993). Conditional Properties of Poststratified Estimators under Normal Theory, *Survey Methodology*, **19**, 183-192.
- Chang, D.S. (1984). On the Asymptotic Distribution of the Multivariate Regression Estimator, *Scandinavian Journal of Statistics*, **11**, 247-255.
- Cochran, W.G. (1977). *Sampling Techniques* (Third Edition), New York: John Wiley and Sons, Inc.
- Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, London: Chapman and Hall.
- Deville, J.C. and Särndal, C.E. (1992). Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, **87**, 376-383.
- Durbin, J. (1969). Inferential Aspects of the Randomness of Sample Size in Survey Sampling, in *New Developments in Survey Sampling* (N.L. Johnson, and H. Smith, Eds.), New York: Wiley-Interscience.
- Fay, R.E. (1993). Valid Inferences from Imputed Survey Data, in *Proceedings of the Section on Survey Research Methods*, Washington, D.C.: American Statistical Association, pp. 41-48.

- Feller, W. (1977) *An Introduction to Probability Theory and Its Applications*, Volume 2, New York: John Wiley and Sons, Inc.
- Flyer, P., Rust, K. and Morganstein, D. (1989). Complex Survey Variance Estimation and Contingency Table Analysis using Replication, internal Westat, Inc. document.
- Fuller, W.A., Loughlin, M.M. and Baker, H.D. (1994). Regression Weighting in the Presence of Nonresponse with Application to the 1987-88 Nationwide Food Consumption Survey, *Survey Methodology*, **20**, 75-85.
- Holt, D. and Smith, T.M.F. (1979). Poststratification, *Journal of the Royal Statistical Society, Series A*, **142**, 33-46.
- Huang, E.T. and Fuller, W.A. (1978). Nonnegative Regression Estimation for Sample Survey Data, in *Proceedings of the Social Statistics Section*, Washington, D.C.: American Statistical Association, pp. 300-305.
- Jones, S.M. and Chromy, J.R. (1982). Improved Variance Estimators using Weighting Class Adjustments for Sample Survey Nonresponse, In *Proceedings of the Section on Survey Research Methods*, Washington, D.C.: American Statistical Association, pp. 105-110.
- Kalton, G. and Kasprzyk, D. (1986). The Treatment of Missing Data, *Survey Methodology*, **12**, 1-16.
- Kovar, J.G., Rao, J.N.K. and Wu, C.F.J. (1988). Measuring Errors in Survey Estimates, *The Canadian Journal of Statistics*, **16**, 25-45.
- Krewski, D. R. (1977). Linearization and Replication Methods in Finite Population Sampling, unpublished Ph.D. thesis, Carleton University, Ottawa, Canada.
- Krewski, D.R. and Rao, J.N.K. (1981). Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods, *The Annals of Statistics*, **9**, 1010-1019.

- Lui, J. (1992). Inference from Stratified Samples: Application of Edgeworth Expansions, unpublished Ph.D. thesis, Carleton University, Ottawa, Canada.
- McCarthy, P.J. (1969). Pseudo-replication: Half-Samples, *Review of the International Statistical Institute*, **37**, 239-264.
- Quenouille, M.H. (1949). Approximate Tests of Correlation in Time-Series, *Journal of the Royal Statistical Society, Series B*, **11**, 68-84.
- Rao, J.N.K. (1985). Conditional Inference in Survey Sampling, *Survey Methodology*, **11**, 15-31.
- Rao, J.N.K. (1988). Variance Estimation in Sample Surveys, in *Handbook of Statistics*, Volume 6 (P.R. Krishnaiah and C.R. Rao, Eds.), Amsterdam: Elsevier Science, pp. 427-447.
- Rao, J.N.K. (1993). Linearization Variance Estimators under Imputation for Missing Data, unpublished manuscript, Carleton University, Ottawa, Canada.
- Rao, J.N.K. (1994a). Estimating Totals and Distribution Functions using Auxiliary Information at the Estimation Stage, *Journal of Official Statistics*, **10**, 153-165.
- Rao, J.N.K. (1994b). Resampling Methods for Complex Surveys, to appear in *Proceedings of the Section on Survey Research Methods*, Washington, D.C.: American Statistical Association.
- Rao, J.N.K. (1996). On Variance Estimation with Imputed Survey Data, to appear in the *Journal of the American Statistical Association*.
- Rao, J.N.K. and Lui, J. (1992). On Estimating Distribution Functions from Sample Survey Data Using Supplementary Information at the Estimation Stage, in *Nonparametric Statistics and Related Topics* (A.K.Md.E Saleh, Ed.), Amsterdam: Elsevier Science Publishers, pp. 399-407.
- Rao, J.N.K. and Shao, J. (1991). Jackknife Variance Estimation with Survey Data under Hot Deck Imputation, Technical Report No. 172, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada.

- Rao, J.N.K. and Shao, J. (1992). Jackknife Variance Estimation with Survey Data under Hot Deck Imputation, *Biometrika*, **79**, 811-822.
- Rao, J.N.K. and Sitter, R.R. (1993). Jackknife Variance Estimation under Imputation for Missing Survey Data, Technical Report No. 214, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada.
- Rao, J.N.K. and Wu, C.F.J. (1985). Inference from Stratified Samples: Second-Order Analysis of Three Methods for Nonlinear Statistics, *Journal of the American Statistical Association*, **80**, 620-630.
- Rao, J.N.K. and Wu, C.F.J. (1988). Resampling Inference with Complex Survey Data, *Journal of the American Statistical Association*, **83**, 231-241.
- Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some Recent Work on Resampling Methods for Complex Surveys, *Survey Methodology*, **18**, 209-217.
- Royall, R.M. (1986). The Prediction Approach to Robust Variance Estimation in Two-Stage Cluster Sampling, *Journal of the American Statistical Association*, **81**, 119-123.
- Royall, R.M. and Cumberland, J. G. (1981). An Empirical Study of the Ratio Estimator and Estimators of Its Variance, *Journal of the American Statistical Association*, **76**, 66-77.
- Royden, H.L. (1988). *Real Analysis* (Third Edition), New York: Macmillan publishing Co.
- Rubin, D.B. (1978). Multiple Imputation in Sample Surveys - A Phenomenological Bayesian Approach to Nonresponse, in *Proceedings of the Section on Survey Research Methods*, Washington, D.C.: American Statistical Association, pp. 20-34.
- Rubin, D.B. and Schenker, N. (1986). Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse, *Journal of the American Statistical Association*, **81**, 366-374.

- Särndal, C.E. (1992). Methods for Estimating the Precision of Survey Estimates when Imputation has Been used, *Survey Methodology*, **18**, 241-252.
- Särndal, C.E., Swensson, B. and Wretman, J.H. (1989). The Weighted Residual Technique for Estimating the Variance of the General Regression Estimator of the Finite Population Total, *Biometrika*, **76**, 527-537.
- Särndal, C.E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag New York, Inc.
- Scott, A. and Wu, C.F.J. (1981). On the Asymptotic Distribution of Ratio and Regression Estimators, *Journal of the American Statistical Association*, **76**, 98-102.
- Shao, J. (1996). Resampling Methods in Sample Surveys, to appear in *Statistics*.
- Shao, J. and Rao, J.N.K. (1993). Standard Errors for Low Income Proportions Estimated from Stratified Multi-Stage Samples, *Sankhyā: The Indian Journal of Statistics*, **55-B**, 393-414.
- Shao, J. and Tu, D.S. (1995). *The Jackknife and Bootstrap*, New York: Springer-Verlag New York, Inc.
- Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, New York: John Wiley and Sons, Inc.
- Smith, T.M.F. (1976). The Foundations of Survey Sampling: A Review, *Journal of the Royal Statistical Society, Series A*, **139**, 183-204.
- Statistics Canada (1992). *Family Expenditure in Canada*, Catalogue Number 62-555, Ottawa: Statistics Canada.
- Tukey, J.W. (1958). Bias and Confidence in Not-Quite so Large Samples (Abstract), *Annals of Mathematical Statistics*, **29**, 614.
- Valliant, R. (1993). Poststratification and Conditional Variance Estimation, *Journal of the American Statistical Association*, **88**, 89-96.

END

1 | 9 - 1 , 1 - 9 6

FIN