1

OF/DE

3

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

32162

National Library
of CanadaBibliothèque nationale
du CanadaCANADIAN THESES
ON MICROFICHETHÈSES CANADIENNES
SUR MICROFICHE

NAME OF AUTHOR/NOM DE L'AUTEUR

DANIEL KREWSKI

TITLE OF THESIS/TITRE DE LA THÈSE

LINEARIZATION AND REPLICATION

METHODS IN FINITE POPULATION SAMPLING

UNIVERSITY/UNIVERSITÉ

CARLETON

DEGREE FOR WHICH THESIS WAS PRESENTED/
GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE

Ph. D. (MATHEMATICS)

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE DEGRÉ

1977

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE

PROFESSOR J. N. K. RAO

Permission is hereby granted to the NATIONAL LIBRARY OF
CANADA to microfilm this thesis and to lend or sell copies
of the film.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈ-
QUE NATIONALE DU CANADA de microfilmer cette thèse et
de prêter ou de vendre des exemplaires du film.

The author reserves other publication rights, and neither the
thesis nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.

L'auteur se réserve les autres droits de publication; ni la
thèse ni de longs extraits de celle-ci ne doivent être imprimés
ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE

10 MAY. 1977

SIGNED/SIGNÉ

D. Krewski

PERMANENT ADDRESS/RÉSIDENCE FIXE

13 LANSFIELD WAY

OTTAWA, ONT.

K2G 3V7

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

LINEARIZATION AND REPLICATION METHODS
IN
FINITE POPULATION SAMPLING

by

Daniel Richard Krewski, B.Sc., M.Sc.

A thesis submitted to the Faculty of
Graduate Studies in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Carleton University

Ottawa, Ontario

May, 1977

© Daniel Krewski 1977

The undersigned hereby recommend to the Faculty of
Graduate Studies and Research acceptance of this thesis,
submitted by Daniel Richard Krewski, B.Sc., M.Sc. in partial
fulfillment of the requirements for the degree of Doctor of
Philosophy.

D. A. Dawson
Chairman, Department of Mathematics

J. N. Krewski
Supervisor

Philip J. M. Carthy
External Examiner

May, 1977

ACKNOWLEDGEMENTS

Professor J.N.K. Rao supervised the research embodied in this thesis. I am deeply grateful for the advice and encouragement he has provided during the course of this work. It has been gratifying to work under his direction on a problem which is both mathematically challenging and of considerable practical importance.

I would like to acknowledge Professors P.R. Beesack, P.S.R.S. Rao and P.K. Sen and Mr. G. Brackstone of Statistics Canada for their helpful discussions on specific aspects of the present work. I would also like to thank Professors M. Csörgő and D.A. Dawson for serving as members of my advisory committee and Mrs. Jill Murray for her outstanding job of preparing the typescript.

This research was made possible through an educational leave of absence from Health & Welfare Canada. Additional support was provided by a Scholarship from the National Research Council of Canada.

Finally, special thanks go to my wife Penny who has seen me through the trials of such an endeavor.

ABSTRACT

Analytic properties of the linearization, jackknife and balanced repeated replication (BRR) methods are investigated for both large and small sample sizes. In Chapter 1, applications of these methods in the analysis of complex survey data are described and a review of the relevant literature is given.

In Chapter 2, some asymptotic distribution theory for the jackknife statistic in simple random sampling without replacement from a finite population is developed. In particular, the asymptotic normality of the finite population jackknife statistic and the consistency of the corresponding jackknife variance estimator are established. The results are applicable in cases where the unjackknifed statistic may be expressed as a function of several U-statistics.

The performance of a jackknife variance estimator applied to ratio estimation is investigated in small samples under a particular model in Chapter 3. The stability of this variance estimator is compared with that of the two commonly used linearization variance estimators under this model.

In Chapter 4, some asymptotic distribution theory for stratified samples is developed. Asymptotic results are obtained as $L \rightarrow \infty$ within the context of a sequence $\{\Pi_L\}$ of finite populations with L strata in Π_L . In particular, the asymptotic normality

of both linear and nonlinear statistics and the consistency of the variance estimators obtained using the linearization, jackknife and BRR methods are established. The results are valid for any stratified multi-stage design in which the primary sampling units are selected with replacement. /

7
In Chapter 5, small sample properties of the above three methods applied to ratio estimation in stratified samples are investigated under a particular model. Comparisons of the bias of four alternative ratio estimators are made as are comparisons of the bias of six alternative variance estimators. The stabilities of particular BRR and linearization variance estimators are also compared under this model.

Several replication variance estimators involving a "minimal" number of replicates are discussed in Chapter 6. The efficiency of these estimators is evaluated in the linear case.

In Chapter 7, a brief summary and suggestions for further research are given.

TABLE OF CONTENTS

Chapter	Page
1. INTRODUCTION	1
1.1. Motivation and Objectives	1
1.2. Review of Literature	5
1.2.1 The Linearization Method	5
1.2.2 The Jackknife Method	6
1.2.3 Balanced Repeated Replication (BRR)	17
1.3. Outline of Thesis	22
2. SIMPLE RANDOM SAMPLING: LARGE SAMPLE RESULTS	25
2.1. U-Statistics	25
2.2. Some Large Sample Theory for U-Statistics	28
2.3. Jackknifing U-Statistics	35
2.4. Summary	45
3. SIMPLE RANDOM SAMPLING: SMALL SAMPLE RESULTS	47
3.1. A Modified Jackknife Variance Estimator	47
3.2. Bias of the Modified Jackknife Variance Estimator	51
3.3. Stability of the Modified Jackknife Variance Estimator	58
3.4. Summary	78
4. STRATIFIED SAMPLING: LARGE SAMPLE RESULTS	80
4.1. Some Large Sample Theory for Independent Random Variables	81
4.2. Stratified Simple Random Sampling	86
4.3. Linearization Methods	99
4.4. Jackknife Methods	103
4.5. Balanced Repeated Replication (BRR)	122
4.6. Stratified Multi-Stage Sampling	133
4.7. Summary	136
5. STRATIFIED SAMPLING: SMALL SAMPLE RESULTS	138
5.1. Bias of Ratio Estimators	138
5.2. Bias of Variance Estimators	145
5.3. Stability of Variance Estimators	159
5.4. Summary	168

Chapter		Page
6.	STRATIFIED SAMPLING: SOME INEFFICIENT REPLICATION VARIANCE ESTIMATORS REQUIRING A REDUCED NUMBER OF REPLICATES	171
6.1.	Variance Estimation in the Linear Case	172
6.2.	Randomly Selected Subsamples.	174
6.3.	Balanced Half-Samples	181
6.4.	Partially Balanced Half-Samples	184
6.5.	Jackknife Methods	188
6.6.	Summary	190
7.	CONCLUDING REMARKS	192
7.1.	Brief Summary	192
7.2.	Further Research	194
	REFERENCES	195

LIST OF TABLES

Table	Page
3.1. Coefficients C_3 and C_4 in $B(v_3)$ and $B(v_4)$ with $g=n$. . .	56
3.2. Coefficients D_3 and D_4 in $B(v_3)$ and $B(v_4)$ with $g=n$. . .	57
3.3. Coefficients J_4 , K_4 and L_4 in $MSE(v_4)$ for $g=4,8$ and Coefficients J_3 , K_3 and L_3 in $MSE(v_3)$ for $g=2$ ($n=8$) . .	73
3.4. Coefficients J_4 , K_4 and L_4 in $MSE(v_4)$ for $g=3,4,6,12$ and Coefficients J_3 , K_3 and L_3 in $MSE(v_3)$ for $g=2$ ($n=12$) . . .	74
3.5. Coefficients J_1 , J_2 and J_4 in $MSE(v_1)$, $MSE(v_2)$ and $MSE(v_4)$ with $g=n$	75
3.6. Coefficients K_1 , K_2 and K_4 in $MSE(v_1)$, $MSE(v_2)$ and $MSE(v_4)$ with $g=n$	76
3.7. Coefficients L_1 , L_2 and L_4 in $MSE(v_1)$, $MSE(v_2)$ and $MSE(v_4)$ with $g=n$	77
5.1. Coefficients D_1 in $B(r_1)$, $i = 1,2,3,4$ for $p=2$	144
5.2. Coefficients F_1 in $B(v_1)$, $i = 1,2,3,4,5,6$ for $p=2$. . .	155
5.3. Coefficients H_1 in $B(v_1)$, $i = 1,2,3,4,5,6$ for $p=2$ and $t=0$	156
5.4. Coefficients H_1 in $B(v_1)$, $i = 1,2,3,4,5,6$ for $p=2$ and $t=1$	157

Table	Page
5.5. Coefficients H_i in $B(v_i)$, $i = 1, 2, 3, 4, 5, 6$ for $p=2$ and $t=2$	158
5.6. Coefficients J_i in $MSE(v_i)$, $i = 1, 2$, for $p=2$	165
5.7. Coefficients K_i in $MSE(v_i)$, $i = 1, 2$, for $p=2$	166
5.8. Coefficients L_i in $MSE(v_i)$, $i = 1, 2$, for $p=2$	167

CHAPTER 1

INTRODUCTION

1.1. Motivation and Objectives

The last decade has witnessed a dramatic increase in the use of sample surveys in many areas of research, including the biological and health sciences, and particularly the social sciences. In addition, while survey data was initially used for descriptive purposes only (with emphasis placed on the estimation of simple population parameters such as means, totals and percentages), the needs of modern survey researchers are more analytic in nature (Rao, 1975) (with emphasis on comparisons among subgroups in the population or on investigating interrelationships among a number of variables).

Sampling theory has advanced to the point where most large scale surveys now involve complex designs and methods of data analysis. Bean (1975, p.1) offers the following description:

"Most scientific sample surveys today are based on complex multistage probability samples, with design components often including unequal probabilities of selection for different elements in the population, stratification, and two or more stages of clustering. The estimation procedures may involve adjustment for nonresponse, use of concomitant variables, and poststratification."

The problem of variance estimation in such complex analytical surveys, however, has been described by McCarthy (1969b, p.239) as follows:

The extensive literature on sampling theory for sample surveys well documents the fact that commonly used sample designs and estimation procedures lead to approximate and extremely complicated methods of variance estimation. Furthermore, the survey data often do not satisfy the conditions required for the application of even quite elementary statistical techniques of analysis. In the absence of techniques for handling these problems, the ordinary researcher will ignore the complexities of design and treat the data as though they had been obtained by simple random sampling; the more sophisticated researcher will be frustrated in not being able to account for all the effects of design in his analysis."

In this connection, Kish & Frankel (1974, p. 20) candidly admit that "the failure to compute sampling errors is a widely known scandal among practitioners".

Three general methods of variance estimation for nonlinear statistics used in the analysis of complex surveys have been developed in response to these needs. Variance estimators obtained using the jackknife and balanced repeated replication (BRR) methods are both based on the variability amongst a number of replicate estimates of the parameter of interest computed from subsamples of the complete sample. Variance estimators obtained using the linearization method are obtained by estimating the variance of the linear term in the Taylor series expansion of the statistic of interest. While these methods have received considerable attention during the last ten years (Cochran, 1975), the behaviour of the variance estimators obtained using these methods is still not thoroughly understood.

The properties of these alternative variance estimators may be investigated in a variety of ways, including the following (McCarthy, 1969a):

- (i) exact analytic in which the functional form of a distribution or joint distribution is assumed,
- (ii) approximate analytic and asymptotic in which various bounds and large sample approximations are employed,
- (iii) empirical studies in which data from actual surveys are used as the basis for sampling experiments, and
- (iv) Monte Carlo sampling from synthetic populations.

While each of the above approaches has been utilized to some extent under the assumption of simple random sampling, much of the work concerned with the jackknife, BRR and linearization methods in stratified samples has been empirical in nature. One obvious limitation of the empirical approach is that conclusions based on the particular population or populations studied may not remain valid for other populations. In the absence of analytical results, however, Frankel (1971, p.6) suggests that empirical results may "provide some first insights into the development of much needed statistical theory for probability samples".

The primary objective of this research is to investigate properties of the jackknife, BRR and linearization methods using

both asymptotic and exact analytic techniques. With Kish & Frankel (1974, p.11) stating that "we need laws of large numbers and central limit theorems for samples in which the elements are not independent" and McCarthy (1969a, p.3) stating that "in the area of present concern, there appears to be little hope of obtaining exact analytic solutions", results along these lines would appear to be welcome.

A secondary objective of this research is to investigate (again using asymptotic and exact analytic techniques) properties of the actual estimates of the parameter of interest produced by the jackknife and BRR methods, which are of interest in themselves. Jackknifing, for example, often produces an estimator with less bias than the original statistic. However, with a number of studies (see Kish, Namboodiri & Pillai [1962], McCarthy [1969a], Frankel [1971] and Bean [1975] for example) finding bias to be negligible for the most part in many large scale surveys, estimators with reduced bias may be of limited interest to practitioners.

1.2. Review of the Literature

1.2.1. The Linearization Method

The most commonly used general procedure for troublesome variance estimation problems is the linearization method (also known as the delta method) in which an approximate variance estimate is obtained by estimating the variance of the linear term in the Taylor series expansion of the statistic of interest. While all sampling textbooks discuss its use in connection with ratio estimation (see Deming [1960], Cochran [1963] and Kish [1965] for example), the first detailed description of the method for more complex estimators seems to have been given by Tepping (1968).

To describe the method in general terms let $\underline{y} = (y_1, \dots, y_p)$ be a vector of statistics whose expected value is a vector of population parameters $\underline{V} = (V_1, \dots, V_p)$. If the function $\theta = f(\underline{V})$ is estimated by $\hat{\theta} = f(\underline{y})$, the linearization method may be used to obtain an approximate estimator of the variance of $\hat{\theta}$. The first order Taylor series approximation for $f(\underline{y})$ is

$$f(\underline{y}) \doteq f(\underline{V}) + \sum_{k=1}^p (y_k - V_k) \frac{\partial f(\underline{V})}{\partial V_k} \quad (1.1)$$

where the partial derivatives are evaluated at $\underline{y} = \underline{V}$. Since the terms $f(\underline{V})$ and $V_k (\partial f(\underline{V}) / \partial V_k)$ are constant, it follows from (1.1) that

$$V(\hat{\theta}) = V \left(\sum_{k=1}^p v_k \frac{\partial f(\underline{y})}{\partial v_k} \right) \\ = \sum_{k=1}^p \sum_{k'=1}^p \frac{\partial f(\underline{y})}{\partial v_k} \frac{\partial f(\underline{y})}{\partial v_{k'}} \text{Cov} (v_k, v_{k'}) . \quad (1.2)$$

The linearization estimator $v_L(\hat{\theta})$ of $V(\hat{\theta})$ is then obtained from (1.2) by estimating the covariance terms and the required partial derivatives from the sample at hand. (Woodruff [1971] has discussed means of reducing the computational effort involved in the use of this method.)

Properties of the linearization method applied to ratio estimation in comparison with the jackknife method are discussed in section 1.2.2. Results of two large scale empirical studies involving the linearization, jackknife and BRR methods are discussed in section 1.2.3.

1.2.2. The Jackknife Method

Quenouille (1949, 1956) originally proposed the jackknife method as a means of bias reduction while Tukey (1958) suggested that the technique may also be useful as a means of variance estimation and constructing approximate confidence intervals. To illustrate the method, suppose that $\hat{\theta}$ denotes an estimator of some parameter θ based on a simple random sample of size n

selected from an infinite population (or selected with replacement from a finite population). The essence of the jackknife technique is to randomly divide the n observations into g groups each of size $p = n/g$ and compute an estimate $\hat{\theta}^i$ of θ after deleting the i^{th} group ($i = 1, \dots, g$). With "pseudovalue" $J_1(\hat{\theta})$ defined by

$$J_1(\hat{\theta}) = g\hat{\theta} - (g-1)\hat{\theta}^i \quad (1.3)$$

($i = 1, \dots, g$) the jackknife estimator $J(\hat{\theta})$ of θ is

$$J(\hat{\theta}) = g^{-1} \sum J_1(\hat{\theta}) \quad (1.4)$$

The estimator $J(\hat{\theta})$ exactly eliminates the n^{-1} term from any bias in $\hat{\theta}$. More precisely, if

$$E(\hat{\theta}) = \theta + \frac{a}{pg} + \frac{b}{(pg)^2} + \dots \quad (1.5)$$

then

$$E(J(\hat{\theta})) = \theta - \frac{b}{p^2 g(g-1)} + \dots \quad (1.6)$$

Tukey proposed the conceptually simple jackknife variance estimator

$$v_J(J(\hat{\theta})) = (g-1)^{-1} \sum (J_1(\hat{\theta}) - J(\hat{\theta}))^2 \quad (1.7)$$

and conjectured that

$$\sqrt{g}(J(\hat{\theta}) - \theta) / \sqrt{v_J} \quad (1.8)$$

is approximately distributed as a t random variable with $g-1$ degrees of freedom in many cases and hence may be used as a pivotal statistic for robust interval estimation.

Several generalizations of the above procedure that have been proposed are indicated briefly. Quenouille (1956) also introduced a "second order" jackknife in which the first order jackknife $J(\hat{\theta})$ is again "jackknifed" in order to eliminate the n^{-2} term from the bias in $J(\hat{\theta})$. Sharot (1976b) introduced a new family of jackknives, consisting of linear combinations of the first and second order jackknives and $\hat{\theta}$, from which an estimator more precise than either order jackknife is derived. Shucany, Gray & Owen (1971) and Gray & Shucany (1972) generalized the jackknife method to handle more general forms of bias than that in (1.5).

When sampling without replacement from a finite population of size N , Jones (1964) points out that the asymptotic bias of $J(\hat{\theta})$ contains terms of order N^{-1} . Jones notes that replacing g by $g(1-(n-p)/N)$ in (1.3) eliminates terms of order n^{-1} and N^{-1} from the asymptotic bias of $J(\hat{\theta})$ in (1.4).

Research on the jackknife method for simple random sampling has been recently summarized in some detail by Miller (1974a). Despite the generalizations described above, much of this research deals with the jackknife as described in (1.3) - (1.8) and in particular with the case $g = n$. Miller notes that the latter case

is the most appealing since it eliminates the need for random grouping and is probably the best form of the jackknife to use in any problem, although for large sample sizes, taking $g = n$ may not be computationally feasible. (General results on the optimality of the choice $g = n$ in terms of bias reduction have recently been reported by Sharot [1976a].)

The review of the literature on jackknife methodology presented here is limited to that material which is relevant to sampling theory. Results on bias reduction, variance estimation and robust interval estimation for simple random samples are first surveyed following which extensions of the jackknife method to stratified samples are discussed. Practical aspects of the jackknife method are discussed by Mosteller & Tukey (1968) and Bissel & Ferguson (1975).

Research concerned with the jackknife as a means of bias reduction in sample surveys has centered on its use for ratios since (while the ratio method of estimation provides a powerful technique for obtaining increased precision through the use of concomitant information) the ordinary ratio estimator is biased. Suppose (y_i, x_i) ($i = 1, \dots, n$) denotes a simple random sample of n observations on the characters 'y' and 'x' and let $\bar{y}$ and $\bar{x}$ denote the sample means. The usual estimator of the population ratio $R = \bar{Y}/\bar{X} = E(y)/E(x)$ is $r = \bar{y}/\bar{x}$ with bias of order n^{-1} .

Durbin (1959) pioneered the application of the jackknife method to ratio estimation by studying the behaviour of $J(r)$ under the linear regression model

$$y = \alpha + \beta x + e \quad (1.9)$$

where e is the error term and x is normally distributed. Neglecting terms of order n^{-4} , Durbin showed that $J(r)$ with $g = 2$ has smaller variance than r . Neglecting terms of order n^{-3} , Rao (1965) showed that both the bias and variance of $J(r)$ are decreasing functions of g so that $g = n$ would be the optimum choice.

When x/n has a gamma distribution, Durbin showed that $J(r)$ with $g = 2$ has reduced bias, increased variance but smaller mean square error than r under (1.9). Under the same model, Rao & Webster (1966) showed that $J(r)$ with $g > 2$ has smaller variance than r and that both the bias and variance of $J(r)$ are again decreasing functions of g .

The jackknife is not the only means of reducing bias in ratio estimation. In fact, when $\bar{X}$ is known, wholly unbiased ratio estimators have been developed by Hartley & Ross (1954) and Mickey (1959). Papers by Tin (1965), Rao (1967), Rao & Beegle (1967), P.S.R.S. Rao (1969), J.N.K. Rao (1969), Rao & Rao (1971), Hutchison (1971), P.S.R.S. Rao (1974), Rao & Kuzik (1974) and Sharot (1976b) compare r , $J(r)$ and several unbiased and approximately unbiased ratio estimators with respect to bias, variance and mean

square error. While no one estimator is uniformly best in all situations, the jackknife is competitive in most cases.

Three estimators of the variance of r have been studied in some detail in the literature. When $\bar{X}$ is known the commonly used linearization estimator (section 1.2.1) for the variance of r is

$$v_1(r) = n^{-1} \bar{X}^{-2} (s_y^2 - 2rs_{xy} + r^2 s_x^2) \quad (1.10)$$

where $s_y^2 = (n-1)^{-1} \sum (y_i - \bar{y})^2$, $s_x^2 = (n-1)^{-1} \sum (x_i - \bar{x})^2$ and $s_{xy} = (n-1)^{-1} \sum (x_i - \bar{x})(y_i - \bar{y})$. Rao & Rao (1971) also consider the linearization variance estimator

$$v_2(r) = n^{-1} \bar{x}^{-2} (s_y^2 - 2rs_{xy} + r^2 s_x^2) \quad (1.11)$$

which does not require the knowledge of $\bar{X}$. A third variance estimator $v_3(r)$ may be obtained by taking the jackknife variance estimator in (1.7) as an estimator of $V(r)$ rather than $V(J(r))$.

In a Monte Carlo study, Lauh & Williams (1963) compared the bias and variance of v_1 and v_3 (with $g = n$) in small samples under the model

$$y = 5(x + e) \quad (1.12)$$

where x is a chi-square random variable (with two degrees of freedom) plus a constant and the error e is normally distributed. They found v_1 to be a serious overestimate of $V(r)$ with v_3

8

exhibiting little bias as an estimator of $V(r)$. In addition, v_3 had a much lower variance than v_1 . With x normally distributed in (1.12), Lauh & Williams found v_1 comparable to v_3 with respect to variance. Under the same model, Rao & Beegle (1967) found that the coefficient of variation of v_3 decreases considerably as g increases and that the coefficients of variation of v_1 and v_3 with $g = n$ are essentially the same. In a second model not so favorable to ratio estimation (with the regression of y on x not through the origin and x having a large coefficient of variation) Rao & Beegle also found these same conclusions to hold.

In an empirical investigation involving sixteen natural populations, J.N.K. Rao (1969) observed that in small samples v_1 was a consistent underestimate of the mean square error of r while v_3 with $g = n$ tended to overestimate $MSE(r)$. While the absolute bias of v_3 was considerably smaller than the absolute bias of v_1 , the stability of v_3 was poor in comparison to that of v_1 .

Rao & Rao (1971) compared v_1 , v_2 and v_3 using analytical means under the model (1.9) with x having a gamma distribution and normally distributed errors with $V(e|x) = \delta x^t$. They found that v_1 and v_2 both tended to underestimate $MSE(r)$ while v_3 with $g = n$ was an overestimate for $t = 0$ and 1 . (Clearcut results concerning the bias of v_3 when $t = 2$ were not obtained. In this case

taking $g = n$ is no longer optimum.) In general v_2 appeared to have the smallest absolute bias. For $t = 0$ and 1 , v_1 had a smaller mean square error than v_2 . The stability of v_3 with $g = 2$ (results for $g > 2$ were not obtained) was poor in comparison with that of v_1 and v_2 .

Similar results were obtained in a semi-empirical study by Rao & Kuzik (1974) using the same model but with the x values chosen from a variety of natural and synthetic populations. The jackknife variance estimator v_3 with $g = n$ was generally less stable than either v_1 or v_2 although v_2 and v_3 were of comparable stability for $t = 0$ and 1 and $n \geq 20$, provided the regression of y on x passes through the origin.

The use of the jackknife method for interval estimation on the basis of Tukey's conjecture may be justified by showing that the statistic in (1.8) has asymptotically a standard normal distribution. Suppose $y_1, \dots, y_n$ denotes a random sample of n observations on the character 'y' from an infinite population (with $E(y) = \mu$ and $V(y) = \sigma^2 < \infty$) and let $\bar{y}$ denote the sample mean. Since $\hat{\theta} = f(\bar{y})$ is generally a biased estimator of $\theta = f(\mu)$, the jackknife estimator $J(\hat{\theta})$ may be considered as an estimator of θ on the basis of bias reduction. In this case, Miller (1964) proved that (for the case $g = n$)

$$\sqrt{n}(J(\hat{\theta}) - \theta) \xrightarrow{L} N(0, |f'(\mu)|^2 \sigma^2) \quad (1.13)$$

and

$$v_J \xrightarrow{P} |f'(\mu)|^2 \sigma^2 \quad (1.14)$$

as $n \rightarrow \infty$ provided f has a bounded second derivative in a neighborhood of μ . It follows from (1.13) and (1.14) that

$$\sqrt{n}(J(\hat{\theta}) - \theta) / \sqrt{v_J} \xrightarrow{L} N(0, 1) \quad (1.15)$$

providing an asymptotic justification of Tukey's conjecture. Since (see Wilks [1962] for example)

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N(0, |f'(\mu)|^2 \sigma^2) \quad (1.16)$$

it follows from (1.14) and (1.16) that

$$\sqrt{n}(\hat{\theta} - \theta) / \sqrt{v_J} \xrightarrow{L} N(0, 1) \quad (1.17)$$

so that the jackknife variance estimator v_J may be used in conjunction with $\hat{\theta}$ in constructing approximate confidence intervals for θ . (Recall that v_J was previously considered as an estimator of $V(\hat{\theta})$ in the case of ratio estimation.)

While the class of statistics covered by the above asymptotic distribution theory is quite limited, Miller (1968) obtained similar results for statistics such as $f(s^2)$ where s^2 is the sample variance. In particular, Miller showed that jackknifing $\log s^2$ provides a robust method of setting confidence intervals on σ^2 when the observations are not normally distributed.

Arvesen (1969) further showed that (1.13) and (1.14) hold for those statistics $\hat{\theta}$ which may be expressed as functions of several U-statistics (see section 2.1). This class of statistics is quite broad and includes most of the commonly encountered nonlinear statistics such as ratios and correlation coefficients. With restrictions on the expectations of the kernels of the U-statistics, Arvesen's results were further extended by Arvesen & Layard (1975) to the case of nonidentically distributed random variables. Miller (1974b) has also shown that (1.13) and (1.14) hold when θ is a function of the regression coefficients in the general linear model.

While the above results constitute an asymptotic justification of Tukey's conjecture, Miller (1974a) notes several instances in which the limiting distribution of (1.8) is not the standard normal. For the jackknife to work, the estimator has to have a locally linear quality. In this case a Taylor series expansion is used where the linear term in the expansion has the correct behaviour with higher order terms being asymptotically negligible.

It also appears that asymptotic normality can only be preserved and not created by jackknifing. Conditions under which the jackknifed and unjackknifed statistic have the same asymptotic distribution have been given recently by Thorburn (1976).

The extension of the jackknife method to stratified samples is not immediate and several different versions have been proposed. Suppose

that $\hat{\theta}$ is an estimator of some parameter θ based on a stratified simple random sample with n_h observations in stratum $h = 1, \dots, L$. For $i = 1, \dots, n_h$ and $h = 1, \dots, L$ let $\hat{\theta}^{hi}$ denote the estimator of θ computed from the sample after deleting the i^{th} unit in the h^{th} stratum and let $\hat{\theta}^h = n_h^{-1} \sum_{i=1}^{n_h} \hat{\theta}^{hi}$. Then Jones (1974) has proposed the jackknife estimator (omitting for simplicity the finite population correction factors included by Jones)

$$J(\hat{\theta}) = [1 + \sum (n_h - 1)] \hat{\theta} - \sum (n_h - 1) \hat{\theta}^h \quad (1.18)$$

and the jackknife variance estimator

$$v(J(\hat{\theta})) = \sum_{h=1}^L \frac{(n_h - 1)}{n_h} \sum_{i=1}^{n_h} (\hat{\theta}^{hi} - \hat{\theta}^h)^2. \quad (1.19)$$

When $\hat{\theta}$ and θ are functions of the sample and population stratum means respectively, Jones shows that in large samples $J(\hat{\theta})$ is unbiased for θ to second order moments and $v(J(\hat{\theta}))$ is unbiased, to second order moments, for either $V(J(\hat{\theta}))$ or $V(\hat{\theta})$. Other variations of (1.18) and (1.19) have been proposed by McCarthy (1966), Frankel (1971) and Lee (1973a) and are described in detail in section 4.4.

Research concerned with the jackknife method for stratified samples is considerably less extensive than that for simple random sampling and has been (with the exception of Jones' results on bias

reduction noted above) essentially empirical in nature. The results of large scale empirical studies by Frankel (1971) and Bean (1975) will be discussed in the next section in conjunction with comparative results for the BRR method.

1.2.3. Balanced Repeated Replication (BRR)

The discussion of balanced repeated replication presented here will largely be confined to the case of stratified sampling with two units selected from each stratum since the literature is almost exclusively devoted to this situation. (Note that as the efficiency of an estimator generally increases with the number of strata for the same sample size, the sample design considered here is of considerable importance and has been frequently utilized in practice.) Extensions of the BRR method to the case of more than two units per stratum are discussed briefly at the end of this section.

With two units selected in each stratum, there are only two independent replicates available for variance estimation using the replication methods of Mahalanobis (1944) and Deming (1956) yielding a highly unstable variance estimator. (Note that for most nonlinear statistics, it is not possible to estimate the variance first within each stratum and then pool across strata.) The BRR method is designed to provide stable variance estimates for nonlinear statistics through the use of a larger number of replicates which are, however,

not independent.

To illustrate the method, suppose $\hat{\theta}$ is an estimator of some parameter θ based on a stratified sample with two units selected in each of L strata. Suppose that a replicate or half-sample is then created by randomly selecting one of the two units in the sample in each stratum and let $\hat{\theta}^{(1)}$ denote the estimator of θ based on the half-sample. Then

$$v_1(\hat{\theta}) = (\hat{\theta}^{(1)} - \hat{\theta})^2 \quad (1.20)$$

is an unbiased estimator of the variance of $\hat{\theta}$ in the linear case. The estimator $v_1(\hat{\theta})$ has also been proposed (although without formal analytical support) as an approximate estimator of the variance of $\hat{\theta}$ in the nonlinear case.

Since the estimator $v_1(\hat{\theta})$ is based on a single degree of freedom and is highly unstable, k repeated replications may be used to obtain an estimator with greater precision. Letting $\hat{\theta}^{(i)}$ denote the estimator of θ based on the i^{th} replicate ($i = 1, \dots, k$) the variance estimator is then

$$v_k(\hat{\theta}) = k^{-1} \sum (\hat{\theta}^{(i)} - \hat{\theta})^2 \quad (1.21)$$

With L strata, there are 2^L possible half-sample replicates. In the linear case, McCarthy (1966) showed that the use of a "balanced" set of replicates produces the same estimate

that would be obtained using all possible 2^L replicates.

(Selecting a set of balanced replicates causes the between stratum component of variance introduced by each individual replication to cancel out over the full set of replicates.) A balanced set of replicates is obtained by constructing a $k \times L$ matrix whose entries are either +1 or -1 with orthogonal columns. With the labels +1 and -1 arbitrarily assigned to the two units in each stratum, the rows of the matrix identify half-samples while the columns refer to strata. With $L = 7$, for example, the following matrix defines a set of $k = 8$ balanced half-samples.

		Stratum						
		1	2	3	4	5	6	7
Half Sample Replication	1	+1	-1	-1	+1	-1	+1	+1
	2	+1	+1	-1	-1	+1	-1	+1
	3	+1	+1	+1	-1	-1	+1	-1
	4	-1	+1	+1	+1	-1	-1	+1
	5	+1	-1	+1	+1	+1	-1	-1
	6	-1	+1	-1	+1	+1	+1	-1
	7	-1	-1	+1	-1	+1	+1	+1
	8	-1	-1	-1	-1	-1	-1	-1

McCarthy noted that a balanced set of at most $L+3$ half-samples may always be obtained using the Plackett & Burman (1946) orthogonal main effects designs for factorial experiments. (Other variations of (1.21) based on a balanced set of replicates are described in

section 4.5.)

Kish & Frankel (1970) used the BRR method in order to estimate the design effect (the ratio of the actual variance of a statistic to that obtained assuming simple random sampling) for ratios and regression and correlation coefficients in a variety of studies involving social and physiological data. Their analysis revealed that while standard errors of analytical statistics based on simple random sampling assumptions are not gross underestimates, the design effects are of an appreciable magnitude.

Frankel (1971) carried out a large scale empirical study involving the three methods of variance estimation (linearization, jackknife and BRR) using data from the Current Population Survey conducted by the U.S. Bureau of the Census. Three designs involving 6, 12 and 30 strata were used. Means, multiple regression coefficients and simple, multiple and partial correlation coefficients were computed using variables related to income, age, education and household size.

Frankel concluded that all three of the methods produce approximately unbiased variance estimators. Differences in the relative mean square errors of the three variance estimators were small, although the variability was consistently lowest using the linearization method and highest for BRR. While no one method was

clearly and uniformly superior to the others on the basis of bias or stability, Frankel found the BRR method best according to a third and perhaps more important criterion: the ratio of the sample estimate minus its expected value to the square root of the variance estimate most closely followed a t distribution (with degrees of freedom equal to the number of strata) using the BRR method. Similar conclusions were obtained in a second large scale empirical study conducted by Bean (1975) but involving only the linearization and BRR methods.

It may be noted that the parameters being estimated in the above empirical studies are all defined in terms of the parameters of the finite population with no reference to an underlying model. In the case of regression coefficients for example, the parameters to be estimated are the coefficients in a regression equation obtained by including all units in the population. Brewer & Mellor (1974) have discussed the propriety of this approach in comparison with the alternative framework provided by an underlying superpopulation model. (Some asymptotic theory for regression coefficients in the latter situation has been developed by Fuller [1975].)

Extensions of the BRR method to the case of more than two units per stratum have been made by Borack (1971), Lee (1971) and Gurney & Jewett (1975). These methods are appropriate when the number of units selected in each stratum is constant and equal to a prime (or possibly a power of a prime).

1.3. Outline of Thesis

Background to the problem of variance estimation in complex surveys was provided in section 1.1 and the relevant literature was reviewed in section 1.2. While the objectives of this thesis were stated in general terms in section 1.1, specific problems considered in subsequent chapters are described in this section. While Chapters 2 and 3 are concerned with simple random samples, the main results of this research are for stratified samples and are presented in Chapters 4, 5 and 6.

* In Chapter 2, asymptotic distribution theory for the jackknife statistic assuming sampling without replacement from a finite population is developed. In particular, the asymptotic normality of the finite population jackknife statistic (Jones, 1964) and the consistency of the corresponding jackknife variance estimator are established. The results are applicable in cases where the unjackknifed statistic may be expressed as a function of several U-statistics (Hoeffding, 1948). An application involving the 1976 Canadian Census is indicated.

In Chapter 3, the stability of a modified jackknife variance estimator applied to ratio estimation is investigated in small samples under the model (1.9) with x having a gamma distribution and normally distributed errors e_i : the modification involves replacing $J(\hat{\theta})$ by $\hat{\theta}$ in (1.7) as proposed by Kish & Frankel

(1970). The stability of the modified jackknife variance estimator is evaluated for arbitrary g in (1.3) and compared with that of the linearization variance estimators in (1.10) and (1.11). (Note that Rao & Rao [1971] were able to evaluate the stability of the usual jackknife variance estimator in (1.7) under this model only for $g = 2$.)

In Chapter 4, asymptotic distribution theory for stratified samples is developed. Asymptotic results are obtained as $L \rightarrow \infty$ within the context of a sequence $\{\pi_L\}$ of finite populations with L strata in π_L . (The results are thus applicable to surveys involving the selection of a small number of primary sampling units within each of a large number of strata.) In particular, the asymptotic normality of both linear and nonlinear statistics and the consistency of the variance estimators obtained using the linearization, jackknife and BRR methods are established.

In Chapter 5, properties of the linearization, jackknife and BRR methods applied to ratio estimation are investigated in small samples under an extension of the linear regression model in (1.9) to stratified samples (Rao & Ramachandran, 1974). Comparisons of the bias of alternative ratio estimators are made under the model as are comparisons of the bias of the alternative variance estimators obtained using the above three methods. The stabilities of the BRR and linearization variance estimators are also compared under the model.

Since the application of both the BRR and jackknife methods of variance estimation can be both cumbersome and expensive when the number of replicates involved is large, several replication methods of variance estimation requiring a reduced number of replicates are discussed in Chapter 6. In particular, the loss in efficiency suffered by these estimators is evaluated in the linear case.

In Chapter 7, a brief summary of the results obtained and their potential impact is provided. Suggestions for further research are also indicated.

CHAPTER 2

SIMPLE RANDOM SAMPLING: LARGE SAMPLE RESULTS

In this chapter, some asymptotic distribution theory for the jackknife statistic is developed for simple random sampling without replacement from a finite population. The results are applicable in cases where the unjackknifed statistic may be expressed as a function of several U-statistics.

The class of U-statistics is introduced in section 2.1. Asymptotic results for U-statistics due to Nandi & Sen (1963) are presented in section 2.2 and then extended to functions of several U-statistics. The asymptotic normality of the jackknife statistic and the consistency of the jackknife variance estimator are established in section 2.3. (An application involving the 1976 Canadian Census is given in this section.) A summary of the results obtained is provided in section 2.4.

2.1. U-Statistics

Let $\Pi = \{Y_1, \dots, Y_N\}$ be a p-variate finite population of size N and let $y_1, \dots, y_n$ be a simple random sample of size $n = \lambda N$ ($0 < \lambda < 1$) selected without replacement from Π . (Note that both Y_1 and y_1 are p-dimensional vectors.) Let $f(y_1, \dots, y_m)$ be a real-valued statistic, symmetric in its arguments,

which is an unbiased estimate of some parameter η where m is the smallest number of observations needed to estimate η . The U-statistic for the parameter η is then given by

$$U(y_1, \dots, y_n) = \binom{n}{m}^{-1} \sum_{C_m} f(y_{\alpha_1}, \dots, y_{\alpha_m}) \quad (2.1)$$

where C_m indicates that the summation extends over all possible $1 \leq \alpha_1 < \dots < \alpha_m \leq n$. The statistic $f(y_1, \dots, y_m)$ is called the kernel of $U(y_1, \dots, y_n)$ and m is called the degree of f .

Many of the commonly encountered statistics are U-statistics or functions of several U-statistics. For example let $y_1 = (z_1, x_1), \dots, y_n = (z_n, x_n)$ be a sample from the bivariate population $Y_1 = (Z_1, X_1), \dots, Y_N = (Z_N, X_N)$ and let

$$\begin{aligned} f^{(1)}((z_1, x_1)) &= z_1, \quad f^{(2)}((z_1, x_1)) = x_1, \\ f^{(3)}((z_1, x_1), (z_2, x_2)) &= (z_1 - z_2)^2/2, \\ f^{(4)}((z_1, x_1), (z_2, x_2)) &= (x_1 - x_2)^2/2 \text{ and} \\ f^{(5)}((z_1, x_1), (z_2, x_2)) &= (z_1 - z_2)(x_1 - x_2)/2. \end{aligned} \quad (2.2)$$

Then the U-statistics corresponding to $f^{(1)}, \dots, f^{(5)}$ are

$$\begin{aligned} U_{(1)} &= n^{-1} \sum z_i = \bar{z}, \quad U_{(2)} = n^{-1} \sum x_i = \bar{x}, \\ U_{(3)} &= (n-1)^{-1} \sum (z_i - \bar{z})^2 = s_z^2, \end{aligned}$$

$$U_{(4)} = (n-1)^{-1} \sum (x_i - \bar{x})^2 = s_x^2 \quad \text{and}$$

$$U_{(5)} = (n-1)^{-1} \sum (z_i - \bar{z})(x_i - \bar{x}) = s_{zx} \quad (2.3)$$

with expectations

$$\eta^{(1)} = N^{-1} \sum Z_i = \bar{Z}, \quad \eta^{(2)} = N^{-1} \sum X_i = \bar{X},$$

$$\eta^{(3)} = (N-1)^{-1} \sum (Z_i - \bar{Z})^2 = S_z^2,$$

$$\eta^{(4)} = (N-1)^{-1} \sum (X_i - \bar{X})^2 = S_x^2 \quad \text{and}$$

$$\eta^{(5)} = (N-1)^{-1} \sum (Z_i - \bar{Z})(X_i - \bar{X}) = S_{zx}. \quad (2.4)$$

The ratio estimator $\bar{z}/\bar{x}$ and the sample correlation coefficient $s_{zx}/(s_z s_x)$ are both functions of several U-statistics. Further examples include regression coefficients, multiple and partial correlation coefficients, higher order moments and Fisher's k-statistics (see Collins & Graham [1973]).

2.2. Some Large Sample Theory for U-Statistics

A framework for finite population asymptotics is provided by the concept of a sequence of finite populations $\{\pi_r\} = \{Y_{r1}, \dots, Y_{rN_r}\}$ of sizes $\{N_r\}$. This approach was introduced by Madow (1948) and subsequently utilized by Erdős & Rényi (1959), Hajék (1960), Nandi & Sen (1963), Barton & David (1963), Sen (1972) and others.

Suffixing the population index r in the notation of the preceding section, consider a collection of q U-statistics $U_{r(j)}$ with kernels $f^{(j)}$ of degree m_j ($j = 1, \dots, q$). For $j = 1, \dots, q$ and $c = 1, \dots, m_j$ define

$$\begin{aligned} f_{rc}^{(j)}(Y_{r1}, \dots, Y_{rc}) &= E\{f^{(j)}(y_{r1}, \dots, y_{rm_j}) | y_{r1} = Y_{r1}, \dots, y_{rc} = Y_{rc}\}, \\ \phi_{rc}^{(j)}(Y_{r1}, \dots, Y_{rc}) &= f_{rc}^{(j)}(Y_{r1}, \dots, Y_{rc}) - \eta_r^{(j)}, \end{aligned} \quad (2.5)$$

and for $j, k = 1, \dots, q$ and $c = 1, \dots, \min\{m_j, m_k\}$ define

$$\begin{aligned} \bar{\zeta}_{rc}^{(j,k)} &= E \phi_{rc}^{(j)}(y_{r1}, \dots, y_{rc}) \phi_{rc}^{(k)}(y_{r1}, \dots, y_{rc}) \text{ and} \\ \zeta_{rc}^{(j,k)} &= \text{Cov} \{f^{(j)}(y_{r1}, \dots, y_{rm_j}), f^{(k)}(y_{r1}, \dots, y_{rc}, y_{r, m_j+1}, \dots, y_{r, m_j+m_k-c})\} \end{aligned} \quad (2.6)$$

Then both consistency and asymptotic normality of U-statistics in sampling without replacement from a finite population have been established by Nandi & Sen (1963).

Theorem 2.1: If $n_r \rightarrow \infty$ as $r \rightarrow \infty$ and $\sup_r E |\phi_{rm_j}^{(j)}|^{1+\delta} < \infty$ for some $\delta > 0$ then $U_{r(j)} - \eta_r^{(j)} \xrightarrow{p} 0$.

Theorem 2.2: If $n_r \rightarrow \infty$, $\lambda_r \rightarrow \lambda < 1$, $\bar{\zeta}_{r1}^{(j,k)} \rightarrow \bar{\zeta}_1^{(j,k)}$ ($j, k = 1, \dots, q$) as $r \rightarrow \infty$ with $((\bar{\zeta}_1^{(j,k)}))$ positive definite and $\sup_r E |\phi_{rm_j}^{(j)}|^{2+\delta} < \infty$ for some $\delta > 0$ and $j = 1, \dots, q$ then the joint asymptotic distribution of $\sqrt{n_r}(U_{r(j)} - \eta_r^{(j)})$, $j = 1, \dots, q$, is q -variate normal with mean vector zero and covariance matrix $((m_j m_k (1-\lambda) \bar{\zeta}_1^{(j,k)}))$:

For $i = 1, \dots, n_r$ and $j = 1, \dots, q$ let

$$v_{r(j)}^i = \left(\frac{n_r - 1}{m_j - 1} \right)^{-1} \sum_{D_{n_r-1}^i} f^{(j)}(y_{r1}, y_{ra_1^i}, \dots, y_{ra_{m_j-1}^i}) \quad (2.7)$$

where $D_{n_r-1}^i$ indicates that the summation extends over all possible $1 \leq \alpha_1^i < \dots < \alpha_{m_j-1}^i \leq n_r$ with $\alpha_k^i \neq i$ for $k = 1, \dots, m_j-1$ and define for $j, k = 1, \dots, q$

$$\hat{\zeta}_{r1}^{(j,k)} = (n_r - 1)^{-1} \sum_{i=1}^{n_r} (v_{r(j)}^i - U_{r(j)})(v_{r(k)}^i - U_{r(k)}) \quad (2.8)$$

Example 2.1: Examples of the quantities defined in (2.5)-(2.8) are now given for the case of means. With the notation of (2.3) and (2.4), $\phi_{r1}^{(1)}(z_{r1}) = (z_{r1} - \bar{z}_r)$, $\bar{\zeta}_{r1}^{(1,1)} = N_r^{-1} E (z_{r1} - \bar{z}_r)^2$, $v_{r(1)}^i = z_{r1}$ and $\hat{\zeta}_{r1}^{(1,1)} = (n_r - 1)^{-1} E (z_{r1} - \bar{z}_r)^2 = s_{rz}^2$.

The following theorem shows $\hat{\zeta}_{r1}^{(j,k)}$ is a consistent estimator of $\bar{\zeta}_{r1}^{(j,k)}$.

Theorem 2.3: If $n_r \rightarrow \infty$ as $r \rightarrow \infty$ and $\sup_r E |\hat{\phi}_{rmj}^{(j)}|^{2+\delta} < \infty$, $\sup_r E |\hat{\phi}_{rmk}^{(k)}|^{2+\delta} < \infty$ for some $\delta > 0$ then $\hat{\zeta}_{r1}^{(j,k)} - \bar{\zeta}_{r1}^{(j,k)} \xrightarrow{P} 0$.

Proof: The case $j = k$ has been established by Nandi & Sen (1963, p.145). For $j \neq k$ write

$$\begin{aligned} & (n_r - 1)^{-1} \sum_{i=1}^{n_r} (v_{r(j)}^i - U_{r(j)})(v_{r(k)}^i - U_{r(k)}) \\ & = (n_r - 1)^{-1} \left\{ \sum_{i=1}^{n_r} (v_{r(j)}^i - \eta_r^{(j)})(v_{r(k)}^i - \eta_r^{(k)}) - n_r (U_{r(j)} - \eta_r^{(j)})(U_{r(k)} - \eta_r^{(k)}) \right\} \end{aligned} \quad (2.9)$$

and note that $(U_{r(j)} - \eta_r^{(j)})(U_{r(k)} - \eta_r^{(k)}) \xrightarrow{P} 0$ by Theorem 2.1.

Recalling that $E \hat{\phi}_{r1}^{(j)}(y_{r1}) \hat{\phi}_{r1}^{(k)}(y_{r1}) = \bar{\zeta}_{r1}^{(j,k)}$ it follows from Theorem 2.1 that

$$n_r^{-1} \sum_{i=1}^{n_r} \hat{\phi}_{r1}^{(j)}(y_{r1}) \hat{\phi}_{r1}^{(k)}(y_{r1}) - \bar{\zeta}_{r1}^{(j,k)} \xrightarrow{P} 0 \quad (2.10)$$

since by the Cauchy-Schwarz and Jensen inequalities (in that order)

$$\begin{aligned} \sup_r E |\hat{\phi}_{r1}^{(j)} \hat{\phi}_{r1}^{(k)}|^{1+\delta/2} & \leq \left\{ \sup_r E |\hat{\phi}_{r1}^{(j)}|^{2+\delta} \sup_r E |\hat{\phi}_{r1}^{(k)}|^{2+\delta} \right\}^{\frac{1}{2}} \\ & \leq \left\{ \sup_r E |\hat{\phi}_{rmj}^{(j)}|^{2+\delta} \sup_r E |\hat{\phi}_{rmk}^{(k)}|^{2+\delta} \right\}^{\frac{1}{2}} < \infty. \end{aligned} \quad (2.11)$$

Next note that by the triangle and Cauchy-Schwarz inequalities (in that order)

$$\begin{aligned}
 & \left| n_r^{-1} \sum_{i=1}^{n_r} (v_{r(j)}^i - \eta_r^{(j)}) (v_{r(k)}^i - \eta_r^{(k)}) - n_r^{-1} \sum_{i=1}^{n_r} \phi_{r1}^{(j)}(y_{r1}) \phi_{r1}^{(k)}(y_{r1}) \right| \\
 & \leq \left| n_r^{-1} \sum_{i=1}^{n_r} (v_{r(j)}^i - \eta_r^{(j)}) (v_{r(k)}^i - f_{r1}^{(k)}(y_{r1})) \right| \\
 & \quad + \left| n_r^{-1} \sum_{i=1}^{n_r} (f_{r1}^{(k)}(y_{r1}) - \eta_r^{(k)}) (v_{r(j)}^i - f_{r1}^{(j)}(y_{r1})) \right| \\
 & \leq \left\{ n_r^{-1} \sum_{i=1}^{n_r} |v_{r(j)}^i - \eta_r^{(j)}|^2 \cdot n_r^{-1} \sum_{i=1}^{n_r} |v_{r(k)}^i - f_{r1}^{(k)}(y_{r1})|^2 \right\}^{\frac{1}{2}} \\
 & \quad + \left\{ n_r^{-1} \sum_{i=1}^{n_r} |f_{r1}^{(k)}(y_{r1}) - \eta_r^{(k)}|^2 \cdot n_r^{-1} \sum_{i=1}^{n_r} |v_{r(j)}^i - f_{r1}^{(j)}(y_{r1})|^2 \right\}^{\frac{1}{2}}. \quad (2.12)
 \end{aligned}$$

Now $n_r^{-1} \sum_{i=1}^{n_r} |v_{r(j)}^i - \eta_r^{(j)}|^2 = \bar{\zeta}_{r1}^{(j,j)} \xrightarrow{P} 0$ (Nandi & Sen) and

$n_r^{-1} \sum_{i=1}^{n_r} |f_{r1}^{(k)}(y_{r1}) - \eta_r^{(k)}|^2 = \bar{\zeta}_{r1}^{(k,k)} \xrightarrow{P} 0$ (Theorem 2.1) with

$\sup_r \bar{\zeta}_{r1}^{(j,j)} < \infty$ and $\sup_r \bar{\zeta}_{r1}^{(k,k)} < \infty$ as in (2.11). Finally by

Chebyshev's inequality, for any $\epsilon > 0$,

$$\begin{aligned}
 & P\left\{ \left| n_r^{-1} \sum_{i=1}^{n_r} (v_{r(k)}^i - f_{r1}^{(k)}(y_{r1}))^2 \right| > \epsilon \right\} \\
 & \leq \epsilon^{-1} n_r^{-1} \sum_{i=1}^{n_r} E |v_{r(k)}^i - f_{r1}^{(k)}(y_{r1})|^2 \\
 & = o(n_r^{-1}) \quad (2.13)
 \end{aligned}$$

since Nandi & Sen (1963, p.144) show $E|v_{r(k)}^1 - f_{r1}^{(k)}(y_{r1})|^2 = O(n_r^{-1})$ uniformly in $i = 1, \dots, n_r$. Thus from (2.12) and (2.13)

$$\begin{aligned} n_r^{-1} \sum_{i=1}^{n_r} (v_{r(j)}^1 - \eta_r^{(j)})(v_{r(k)}^1 - \eta_r^{(k)}) &= n_r^{-1} \sum_{i=1}^{n_r} \phi_{r1}^{(j)}(y_{r1}) \phi_{r1}^{(k)}(y_{r1}) \\ &\xrightarrow{P} 0 \end{aligned} \quad (2.14)$$

and the result now follows from (2.9), (2.10) and (2.14). ///

Example 2.2 (Estimation of means): With the notation of (2.3) and (2.4) it follows from Theorems 2.2 and 2.3 that

$$\sqrt{n_r} (\bar{x}_r - \bar{X}_r) / \sqrt{(1-\lambda_r)s_{rx}^2} \xrightarrow{L} N(0,1) \text{ under the conditions}$$

$$n_r \rightarrow \infty, \lambda_r \rightarrow \lambda < 1, s_{rx}^2 \rightarrow s_x^2 > 0 \text{ as } r \rightarrow \infty \text{ and}$$

$$\sup_r n_r^{-1} \sum |X_{ri} - \bar{X}_r|^{2+\delta} < \infty.$$

As noted in the preceding section many nonlinear statistics may be expressed as functions of several U-statistics. The asymptotic normality of such statistics and the consistency of the corresponding "linearization" variance estimator are shown in the following two theorems:

Theorem 2.4: Let g be a real-valued function defined on R^q with continuous first partial derivatives in a neighborhood of $(\eta^{(1)}, \dots, \eta^{(q)})$. If $n_r \rightarrow \infty$, $\lambda_r \rightarrow \lambda < 1$, $\eta_r^{(j)} \rightarrow \eta^{(j)}$, $\bar{\zeta}_{r1}^{(j,k)} \rightarrow \bar{\zeta}_1^{(j,k)}$ ($j, k = 1, \dots, q$) as $r \rightarrow \infty$ with $((\bar{\zeta}_1^{(j,k)}))$ positive definite and $\sup_r E|\phi_{rmj}^{(j)}|^{2+\delta} < \infty$ for some $\delta > 0$ and $j = 1, \dots, q$ then

$$\sqrt{n_r}(\hat{\theta}_r - \theta_r) \xrightarrow{L} N(0, (1-\lambda) \sum_{j=1}^q \sum_{k=1}^q m_j m_k g_j g_k \bar{c}_1^{(j,k)})$$

where $\theta_r = g(\eta_r^{(1)}, \dots, \eta_r^{(q)})$, $\hat{\theta}_r = g(U_{r(1)}, \dots, U_{r(q)})$ and
 $g_k = \partial g(t^{(1)}, \dots, t^{(q)}) / \partial t^{(k)} \Big|_{(\eta_r^{(1)}, \dots, \eta_r^{(q)})}$ for $k = 1, \dots, q$.

Proof: Let $I_j = (\eta_r^{(j)} - \Delta, \eta_r^{(j)} + \Delta)$, $\Delta > 0$, be any interval about $\eta_r^{(j)}$ in which $g_j(\cdot)$ is continuous ($j = 1, \dots, q$). Then by Theorem 2.1 (and noting that $\eta_r^{(j)} \rightarrow \eta^{(j)}$)

$$g_j(\eta_r^{(j)}), U_{r(j)} \in I_j, \quad j = 1, \dots, q \rightarrow 1 \quad (2.15)$$

as $r \rightarrow \infty$. Now for two sequences of events $\{A_r\}$ and $\{B_r\}$ in which $\lim_r P\{B_r\} = 1$, $\lim_r P\{A_r\} = \lim_r P\{A_r \cap B_r\}$ so that it may be tacitly assumed in what follows that the events described in (2.15) hold.

$$\text{For } \eta_r^{(j)}, U_{r(j)} \in I_j \quad (j = 1, \dots, q)$$

$$\begin{aligned} g(U_{r(1)}, \dots, U_{r(q)}) &= g(\eta_r^{(1)}, \dots, \eta_r^{(q)}) \\ &+ \sum_{j=1}^q (U_{r(j)} - \eta_r^{(j)}) g_j(\xi_r) \end{aligned} \quad (2.16)$$

where ξ_r is on the line segment joining $(U_{r(1)}, \dots, U_{r(q)})$ and $(\eta_r^{(1)}, \dots, \eta_r^{(q)})$. Rewriting (2.16) as

$$\sqrt{n_r}(\hat{\theta}_r - \theta_r) = \sqrt{n_r} \sum_{j=1}^q (U_{r(j)} - \eta_r^{(j)}) g_j(\xi_r) \quad (2.17)$$

and noting that by continuity $g_j(\xi_r) \xrightarrow{P} g_j$ (with ξ_r defined arbitrarily when the events in (2.15) do not hold) the result now follows from Theorem 2.2 and Slutsky's theorem. ///

Theorem 2.5: Under the conditions of the preceding theorem

$$v_{rL} \xrightarrow{P} \sum_{j=1}^q \sum_{k=1}^q m_j m_k g_j g_k \bar{c}_1^{(j,k)} \quad \text{where}$$

$$v_{rL} = \sum_{j=1}^q \sum_{k=1}^q m_j m_k g_j(U_{r(1)}, \dots, U_{r(q)}) g_k(U_{r(1)}, \dots, U_{r(q)}) \hat{c}_{r1}^{(j,k)}$$

is the "linearization" variance estimator.

Proof: By continuity $g_j(U_{r(1)}, \dots, U_{r(q)}) \xrightarrow{P} g_j(\eta^{(1)}, \dots, \eta^{(q)})$.

By Theorem 2.3 $\hat{c}_{r1}^{(j,k)} \xrightarrow{P} \bar{c}_1^{(j,k)}$. ///

Example 2.3 (Ratio Estimation): With the notation of (2.3) and (2.4)

the classical estimator of the population ratio $\theta_r = g(\bar{Z}_r, \bar{X}_r) = \bar{Z}_r / \bar{X}_r$ is $\hat{\theta}_r = g(\bar{z}_r, \bar{x}_r) = \bar{z}_r / \bar{x}_r$. The linearization variance estimator is $v_{rL} = \bar{x}_r^{-2} (s_{rz}^2 + \hat{\theta}_r^2 s_{rx}^2 - 2\hat{\theta}_r s_{rzx})$. Under the

conditions of Theorems 2.4 and 2.5 $\sqrt{n_r}(\hat{\theta}_r - \theta_r) / \sqrt{(1-\lambda_r)v_{rL}} \xrightarrow{L} N(0,1)$.

2.3. Jackknifing U-Statistics

While $U_{r(j)}$ is an unbiased estimator of $\eta_r^{(j)}$ ($j = 1, \dots, q$), $\hat{\theta}_r = g(U_{r(1)}, \dots, U_{r(q)})$ is not in general an unbiased estimator of $\theta_r = (\eta_r^{(1)}, \dots, \eta_r^{(q)})$ due to the nonlinearity of the transformation. The jackknife statistic and its generalization (see Gray & Shucany [1972, p.138]) may then be considered as an estimator of θ_r on the basis of bias reduction.

The essence of the jackknife method is to randomly divide the n_r observations into g_r groups of size $p_r = n_r/g_r$ and compute an estimate $\hat{\theta}_r^i$ of θ_r after deleting the i th group ($i = 1, \dots, g_r$). Since results by Rao & Webster (1966), Rao & Rao (1971), Sharot (1976a) and others indicate that $g_r = n_r$ is the "optimal" choice in many situations only this case will be treated explicitly here.

For $j = 1, \dots, q$ and $i = 1, \dots, n_r$ let

$$U_{r(j)}^i = \binom{n_r - 1}{m_j}^{-1} \sum_{C_{n_r-1}^i} f^{(j)}(y_{r\alpha_1^i}, \dots, y_{r\alpha_{m_j}^i}) \quad (2.18)$$

where $C_{n_r-1}^i$ indicates that the summation extends over all possible $1 \leq \alpha_1^i < \dots < \alpha_{m_j}^i \leq n_r$ with $\alpha_k^i \neq i$ for $k = 1, \dots, m_j$. With $\hat{\theta}_r^i = g(U_{r(1)}^i, \dots, U_{r(q)}^i)$ define $G_i(\hat{\theta}_r) = (1 - R_r)^{-1} (\hat{\theta}_r - R_r \hat{\theta}_r^i)$ for $i = 1, \dots, n_r$ where $R_r \neq 1$ is a known constant. The generalized jackknife statistic is then defined by

$$G(\hat{\theta}_r) = n_r^{-1} \sum_{i=1}^{n_r} G_i(\hat{\theta}_r) \quad (2.19)$$

and the jackknife variance estimator is

$$v_{rG} = (n_r - 1)^{-1} \sum_{i=1}^{n_r} (G_i(\hat{\theta}_r) - G(\hat{\theta}_r))^2 \quad (2.20)$$

When g is a function of means only Jones (1964, 1974) has given a jackknife statistic for sampling without replacement from a finite population which is a special case of (2.19) with $R_r = n_r^{-1}(N_r - n_r + 1)^{-1}(N_r - n_r)(n_r - 1)$. Jones has shown that this choice of R_r eliminates bias to second order moments in large samples while J.N.K. Rao (1969) provides empirical evidence of bias reduction in small samples.

With the notation $\omega_r = (n_r - 1)^{-1}(1 - R_r)^{-1}R_r$ the asymptotic normality of $G(\hat{\theta}_r)$ and the consistency of v_{rG} are shown in the following two theorems.

Theorem 2.6: Let g be a real-valued function defined on R^q with bounded second derivatives in a neighborhood of $(\eta^{(1)}, \dots, \eta^{(q)})$. If $n_r \rightarrow \infty$, $\lambda_r \rightarrow \lambda < 1$, $\omega_r \rightarrow \omega \neq 0$, $\eta_r^{(j)} \rightarrow \eta^{(j)}$, $\bar{\zeta}_{r1}^{(j,k)} \rightarrow \bar{\zeta}_1^{(j,k)}$ ($j, k = 1, \dots, q$) as $r \rightarrow \infty$ with $((\bar{\zeta}_1^{(j,k)}))$ positive definite and $\sup_r E|\phi_{rm_j}^{(j)}|^{2+\delta} < \infty$ for some $\delta > 0$ and $j = 1, \dots, q$ then

$$\sqrt{n_r} \omega_r (G(\hat{\theta}_r) - \theta_r) \xrightarrow{L} N(0, \omega^2(1-\lambda) \sum_{j=1}^q \sum_{k=1}^q n_j n_k g_j g_k \bar{\zeta}_1^{(j,k)})$$

where $g_k = \partial g(t^{(1)}, \dots, t^{(q)}) / \partial t^{(k)} \Big|_{(\eta^{(1)}, \dots, \eta^{(q)})}$ for $k = 1, \dots, q$.

Proof: The method of proof is that introduced by Miller (1964) and followed by Arvesen (1969). For simplicity only the case $q=1$ is proved in detail; the case $q > 1$ is essentially the same.

Suppressing the index j let $I = (\eta - 2\Delta, \eta + 2\Delta)$, $\Delta > 0$, be any neighborhood of η in which $g''(\cdot)$ is bounded. By Theorem 2.1 $U_r - \eta_r \xrightarrow{P} 0$ so that

$$P\{U_r \in (\eta - \Delta, \eta + \Delta)\} \rightarrow 1 \quad (2.21)$$

as $r \rightarrow \infty$. Nandi & Sen (1963, p.143) show that

$$\text{Var}(V_r^1) = \binom{n_r - 1}{m - 1}^{-1} \sum_{c=1}^m \binom{m - 1}{c - 1} \binom{n_r - m}{m - c} \zeta_{rc} \quad (2.22)$$

from which $\text{Var}(V_r^1) - \zeta_{r1} \rightarrow 0$ since $\sup_r |\zeta_{rc}| \leq \sup_r |\bar{\zeta}_{rm}| < \infty$ for $c = 1, \dots, m$ and since

$$\binom{n_r - 1}{m - 1}^{-1} \binom{n_r - m}{m - c} \rightarrow \begin{cases} 1 & \text{if } c = 1 \\ 0 & \text{if } c > 1 \end{cases} \quad (2.23)$$

as $r \rightarrow \infty$. Nandi & Sen (1963, p.132) also show

$$\bar{\zeta}_{r1} = \binom{n_r - 1}{m - 1}^{-1} \sum_{c=1}^m \binom{m - 1}{c - 1} \binom{n_r - m}{m - c} \zeta_{rc} \quad (2.24)$$

from which $\bar{\zeta}_{r1} - \zeta_{r1} \rightarrow 0$ (as in (2.23)) so that $\text{Var}(V_r^1) \rightarrow \bar{\zeta}_1$.

Then using the fact that the V_r^i ($i=1, \dots, n_r$) are identically

distributed random variables and the Chebycheff inequality

$$\begin{aligned}
 & \lim_r P \left\{ \max_{1 \leq i \leq n_r} \frac{m}{n_r} |V_r^i - \eta| \geq \Delta \right\} \\
 & \leq \lim_r P \left\{ \max_{1 \leq i \leq n_r} \frac{m}{n_r} |V_r^i - \eta_r| + \frac{m}{n_r} |\eta_r - \eta| \geq \Delta \right\} \\
 & \leq \lim_r n_r P \left\{ \frac{m}{n_r} |V_r^i - \eta_r| \geq \frac{\Delta}{2} \right\} \\
 & \leq \lim_r 4m^2 (\zeta_1 + 1) / (n_r \Delta^2) = 0 .
 \end{aligned} \tag{2.25}$$

Since $U_r^i = (n_r U_r - m V_r^i) / (n_r - m)$,

$$|U_r^i - \eta| \leq \frac{n_r}{(n_r - m)} (|U_r - \eta| + \frac{m}{n_r} |V_r^i - \eta|) . \tag{2.26}$$

From (2.21), (2.25) and (2.26)

$$\begin{aligned}
 & \lim_r P \left\{ \max_{1 \leq i \leq n_r} |U_r^i - \eta| < 2\Delta \right\} \\
 & \geq \lim_r P \left\{ \frac{n_r}{n_r - m} (|U_r - \eta| + \max_{1 \leq i \leq n_r} \frac{m}{n_r} |V_r^i - \eta|) < 2\Delta \right\} \\
 & = \lim_r P \left\{ |U_r - \eta| + \max_{1 \leq i \leq n_r} \frac{m}{n_r} |V_r^i - \eta| < 2\Delta \right\} \\
 & \geq \lim_r P \left\{ |U_r - \eta| < \Delta, \max_{1 \leq i \leq n_r} \frac{m}{n_r} |V_r^i - \eta| < \Delta \right\} \\
 & = 1 .
 \end{aligned} \tag{2.27}$$

Combining (2.21) and (2.27)

$$P\{U_r, U_r^1 \in I; i = 1, \dots, n_r\} \rightarrow 1. \quad (2.28)$$

By the argument following (2.15), it may be tacitly assumed in what follows that the events described in (2.28) hold.

For $U_r, U_r^1 \in I$ ($i = 1, \dots, n_r$)

$$g(U_r^1) = g(U_r) + (U_r^1 - U_r) g'(U_r) + (U_r^1 - U_r)^2 g''(\xi_r^1)/2 \quad (2.29)$$

where ξ_r^1 lies between U_r^1 and U_r . Substituting (2.29) in (2.19) and noting that $n_r^{-1} \sum U_r^1 = U_r$ leads to

$$\begin{aligned} \sqrt{n_r} \omega_r(G(\hat{\theta}_r) - \theta_r) &= \sqrt{n_r} \omega_r(\hat{\theta}_r - \theta_r) \\ &= \omega_r^2 n_r^{-1/2} (n_r - 1) \sum_{i=1}^{n_r} (U_r^1 - U_r)^2 g''(\xi_r^1)/2. \end{aligned} \quad (2.30)$$

Since the first term on the right of (2.30) is asymptotically normally distributed with mean zero and variance $\omega^2(1-\lambda)|g'(\eta)|^2 m^2 \bar{\zeta}_1$ by Theorem 2.4 the result will follow from Slutsky's theorem if it can be shown that the second term on the right $\xrightarrow{p} 0$.

$$\text{Since } \sum (U_r^1 - U_r)^2 = m^2 (n_r - 1) \hat{\zeta}_{r1} / (n_r - m)^2 \text{ and } \hat{\zeta}_{r1} - \bar{\zeta}_{r1} \xrightarrow{p} 0$$

by Theorem 2.3 it follows that

$$(n_r - 1) \sum_{i=1}^{n_r} (U_r^1 - U_r)^2 \xrightarrow{p} m^2 \bar{\zeta}_1. \quad (2.31)$$

Then since $|g''(\xi_r^1)| < M < \infty$ ($i = 1, \dots, n_r$) when the events in (2.28) hold, the second term on the right side of (2.30) $\xrightarrow{P} 0$ due to the $\sqrt{n_r}$ term in the denominator (with ξ_r^1 defined arbitrarily when the events in (2.28) do not hold). ///

It may be noted that Arvesen's (1969) proof of (2.31) in the case of independent identically distributed y_i 's is based on a combinatorial argument and could be simplified in the manner of the present proof by appealing to Proposition 1 of Sen (1960).

Theorem 2.7: Let g be a real-valued function defined on R^q with continuous first partial derivatives in a neighborhood of $(\eta^{(1)}, \dots, \eta^{(q)})$. If $n_r \rightarrow \infty$, $\lambda_r \rightarrow \lambda < 1$, $\omega_r \rightarrow \omega \neq 0$, $\eta_r^{(j)} \rightarrow \eta^{(j)}$, $\bar{\zeta}_{r1}^{(j,k)} \rightarrow \bar{\zeta}_1^{(j,k)}$ ($j, k = 1, \dots, q$) as $r \rightarrow \infty$ with $((\bar{\zeta}_1^{(j,k)}))$ positive definite and $\sup_r E|\phi_{rmj}^{(j)}|^{2+\delta} < \infty$ for some $\delta > 0$ and $j = 1, \dots, q$ then

$$v_{rG} \xrightarrow{P} \omega^2 \sum_{j=1}^q \sum_{k=1}^q m_j m_k g_j g_k \bar{\zeta}_1^{(j,k)}$$

where $g_k = \partial g(t^{(1)}, \dots, t^{(q)}) / \partial t^{(k)} \Big|_{(\eta^{(1)}, \dots, \eta^{(q)})}$ for $k = 1, \dots, q$.

Proof: Details of the proof need again only be given for $q = 1$.

As in the preceding theorem

$$P\{U_r, U_r^1 \in I, i = 1, \dots, n_r\} \rightarrow 1 \quad (2.32)$$

as $r \rightarrow \infty$ where $I = (\eta - \Delta, \eta + \Delta)$, $\Delta > 0$, is any neighborhood of η in which $g'(\cdot)$ is continuous. For $U_r, U_r^i \in I$ ($i=1, \dots, n_r$)

$$g(U_r^i) = g(U_r) + (U_r^i - U_r) g'(\xi_r^i) \quad (2.33)$$

where ξ_r^i lies between U_r^i and U_r . Since $g'(\cdot)$ is continuous at η , for any $\epsilon > 0$ there exists $\Delta_\epsilon > 0$ such that

$|g'(\xi_r^i) - g'(\eta)| < \epsilon$ whenever $|\xi_r^i - \eta| < \Delta_\epsilon$. Thus using (2.32) (with ξ_r^i defined arbitrarily when the events in (2.32) do not hold)

$$\begin{aligned} \lim_r P\{ \max_{1 \leq i \leq n_r} |g'(\xi_r^i) - g'(\eta)| < \epsilon \} \\ \geq \lim_r P\{ \max_{1 \leq i \leq n_r} |\xi_r^i - \eta| < \Delta_\epsilon \} = 1. \end{aligned} \quad (2.34)$$

Substituting (2.33) into (2.20) leads to

$$\begin{aligned} v_{rG} &= \omega_r^2 (n_r - 1) \sum_{i=1}^{n_r} \{ (U_r^i - U_r) g'(\eta) + (U_r^i - U_r) (g'(\xi_r^i) - g'(\eta)) \} \\ &\quad - n_r^{-1} \sum_{j=1}^{n_r} (U_r^j - U_r) (g'(\xi_r^j) - g'(\eta)) \}^2 \\ &= \omega_r^2 |g'(\eta)|^2 (n_r - 1) \sum_{i=1}^{n_r} (U_r^i - U_r)^2 \\ &\quad + \omega_r^2 (n_r - 1) \sum_{i=1}^{n_r} \{ (U_r^i - U_r) (g'(\xi_r^i) - g'(\eta)) \} \\ &\quad - n_r^{-1} \sum_{j=1}^{n_r} (U_r^j - U_r) (g'(\xi_r^j) - g'(\eta)) \}^2 \\ &\quad + \text{cross-product term.} \end{aligned} \quad (2.35)$$

By (2.31) the first term on the right side of (2.35) $\xrightarrow{P}$
 $\omega^2 |g'(\eta)|^2 m^2 \zeta_1$. Since

$$\begin{aligned} & (n_r - 1) \sum_{i=1}^{n_r} (U_r^i - U_r)^2 (g'(\xi_r^i) - g'(\eta))^2 \\ & \leq \max_{1 \leq i \leq n_r} |g'(\xi_r^i) - g'(\eta)|^2 (n_r - 1) \sum_{i=1}^{n_r} (U_r^i - U_r)^2 \\ & \xrightarrow{P} 0 \end{aligned} \quad (2.36)$$

by (2.34) and (2.31), the second term on the right side of (2.35)
 $\xrightarrow{P} 0$. The result now follows as the cross-product term $\xrightarrow{P} 0$
 by the Cauchy-Schwarz inequality. ///

Replacing $G(\hat{\theta}_r)$ by $\hat{\theta}_r$ in (2.20) yields the modified
 jackknife variance estimator

$$v_{rG'} = (n_r - 1)^{-1} \sum_{i=1}^{n_r} (G_i(\hat{\theta}_r) - \hat{\theta}_r)^2 \quad (2.37)$$

proposed by Kish & Frankel (1970). Consistency of $v_{rG'}$ may be
 established using precisely the same argument as in Theorem 2.7.

Remark 2.1: Since

$$v_{rG'} = v_{rG} + \frac{n_r}{(n_r - 1)} (G(\hat{\theta}_r) - \hat{\theta}_r)^2 \quad (2.38)$$

the difference between the two jackknife variance estimators is
 expected to be small.

Taking $R_r = n_r^{-1}(N_r - n_r + 1)^{-1}(N_r - n_r)(n_r - 1)$ and combining Theorems 2.6 and 2.7 yields

$$\sqrt{n_r}(1 - \lambda_r)(G(\hat{\theta}_r) - \theta_r)/\sqrt{(1 - \lambda_r)v_{rG}} \xrightarrow{L} N(0,1) . \quad (2.39)$$

While the jackknife may be useful as a method of bias reduction in small samples, bias in large samples will usually be negligible. In this case the primary appeal of the jackknife method arises from the computationally simple variance estimator v_{rG} (partial derivatives need not be calculated as for the linearization variance estimator v_{rL}). By Theorems 2.4 and 2.7 the jackknife variance estimator v_{rG} may be used in conjunction with the estimator $\hat{\theta}_r$ in that

$$\sqrt{n_r}(1 - \lambda_r)(\hat{\theta}_r - \theta_r)/\sqrt{(1 - \lambda_r)v_{rG}} \xrightarrow{L} N(0,1) . \quad (2.40)$$

A potential application of (2.40) in the analysis of data from the Canadian Census is described below.

Example 2.4: The 1976 Canadian Census involved the distribution of a detailed questionnaire (2B) to every third household in the country ($\lambda = 1/3$) with a shorter version (2A) distributed to the remaining households. A comparison (within selected enumeration areas) of the distribution of household sizes in the sample (2B) with the distribution in the population (2A and 2B) provides a means of assessing possible enumerator bias in the selection of households. One of the comparisons under consideration is based on dispersion as measured by the

sample variance s^2 and the population mean square S^2 . Applying a logarithmic transformation, as recommended by Miller (1968), the test statistic $\sqrt{n(1-\lambda)}(\log s^2 - \log S^2)/\sqrt{(1-\lambda)v_G}$ has approximately a standard normal distribution for large sample sizes n (as in (2.40)) under the null hypothesis of no interviewer bias. Miller's results also suggest that this test should be robust against a nonnormal distribution of household sizes in the population.

2.4. Summary

The asymptotic normality of the jackknife statistic and the consistency of the jackknife variance estimator in the case of independent identically distributed random variables have been established by Arvesen (1969) when the unjackknifed statistic may be expressed as a function of several U-statistics. In this chapter similar results are obtained in the case of simple random sampling without replacement from a finite population where the observations are no longer independent. The large sample theory is developed for the generalized jackknife statistic defined in (2.19), of which Jones' (1964, 1974) finite population jackknife statistic is a special case.

The class of statistics which may be expressed as functions of several U-statistics is quite broad and includes many of the commonly encountered statistics such as ratios and correlation coefficients. Included in this class are frequently used transformations of statistics such as $\log s^2$ and $\tanh^{-1} r$, where s^2 and r denote the sample variance and product-moment correlation coefficient respectively. (The robustness of the jackknife method in constructing tests of hypotheses and confidence intervals in these two cases has been demonstrated by Miller [1968] and Duncan & Layard [1973] respectively.) The application involving $\log s^2$ described in Example 2.4 illustrates the need for large sample theory for the

jackknife in sampling without replacement from a finite population since, while the population size is large, the sampling fraction λ is appreciable ($\lambda = 1/3$) .

CHAPTER 3

SIMPLE RANDOM SAMPLING: SMALL SAMPLE RESULTS

In this chapter, the stability of a jackknife variance estimator applied to ratio estimation is evaluated in small samples under a particular model. The stability of this estimator is then compared with that of the two commonly used linearization variance estimators.

Details of the model employed and the variance estimators considered are provided in section 3.1. Results on the bias of the jackknife variance estimator are presented in section 3.2 while stability comparisons involving the jackknife and linearization variance estimators are presented in section 3.3. A summary of the results obtained is provided in section 3.4.

3.1. A Modified Jackknife Variance Estimator

Suppose (y_i, x_i) ($i = 1, \dots, n$) denotes a simple random sample of n observations on the characters 'y' and 'x' and let $\bar{y}$ and $\bar{x}$ denote the sample means. The classical estimator of the population ratio $R = \bar{Y}/\bar{X} = E(y)/E(x)$ is $r = \bar{y}/\bar{x}$.

The two commonly used linearization estimators for the variance of r are

$$v_1 = n^{-1} \bar{X}^{-2} (s_y^2 - 2r s_{xy} + r^2 s_x^2) \quad (3.1)$$

and

$$v_2 = n^{-1} \bar{x}^{-2} (s_y^2 - 2rs_{xy} + r^2 s_x^2) \quad (3.2)$$

where $s_y^2 = (n-1)^{-1} \sum (y_i - \bar{y})^2$, $s_{xy} = (n-1)^{-1} \sum (x_i - \bar{x})(y_i - \bar{y})$
and $s_x^2 = (n-1)^{-1} \sum (x_i - \bar{x})^2$. (Note that v_1 may only be used if $\bar{X}$ is known.)

If the sample is randomly divided into g groups each of size $p = n/g$ then the jackknife variance estimator is

$$v_3 = g^{-1}(g-1) \sum_{j=1}^g (r'_j - \bar{r}_g)^2 \quad (3.3)$$

where $\bar{r}_g = \sum r'_j / g$ and r'_j is the classical ratio estimator computed from the sample after omitting the j^{th} group.

P.S.R.S. Rao & J.N.K. Rao (1971) compared the biases of v_1 , v_2 and v_3 under the model

$$y_i = \alpha + \beta x_i + e_i, \quad (3.4)$$

$$E(e_i | x_i) = 0,$$

$$E(e_i^2 | x_i) = \delta x_i^t,$$

$$E(e_i e_j | x_i x_j) = 0$$

($i \neq j = 1, \dots, n$) where x has a gamma distribution with mean h . They also compared the mean square errors of v_1 , v_2 and v_3 (with $g = 2$) assuming in addition a normal distribution for the errors e_1 , conditional on the x_1 . (Since v_1 may only be used when $\bar{X}$ is known, comparisons of either v_2 or v_3 with v_1 are made under the assumption that $\bar{X}$ is known.)

Under the model (3.4) Rao & Rao were able to evaluate the mean square error of v_3 only for $g = 2$. In sections 3.2 and 3.3 the bias and mean square error of the modified jackknife variance estimator

$$v_4 = g^{-1}(g-1) \sum_{j=1}^g (r'_j - r)^2 \quad (3.5)$$

(proposed by Kish & Frankel [1970]) are evaluated under the above model for $g > 2$. Since

$$v_4 = v_3 + (g-1)(\bar{r}'_g - r)^2 \quad (3.6)$$

the difference between v_4 and v_3 in terms of bias and mean square error should be small. (The results in section 3.2 confirm this with respect to bias.)

For a variety of natural and artificial populations Rao & Kuzik (1974) found the coefficient of variation of the x population to lie between 0.4 and 1.0. Since the coefficient of

variation of x under the model (3.4) is $h^{-1/2}$, small values of h are thus of most interest. In practice t has often been found to lie between zero and two.

For simplicity the population size is assumed to be infinite. The results are exact for any sample size n .

3.2. Bias of the Modified Jackknife Variance Estimator

Under the model (3.4) the mean square error of r is

$$MSE(r) = \frac{n^2(m+2)\alpha^2}{m^2(m-1)(m-2)} + \frac{nf(t)\delta}{(m+t-1)(m+t-2)} \quad (3.7)$$

where $m = nh > 2$ and $f(t) = \Gamma(h+t)/\Gamma(h)$. To obtain the bias and mean square error of v_4 as an estimator of $MSE(r)$ the following four results are required. (The first result is well known but is given here for future reference.)

Lemma 3.1: If X is a gamma variate with mean a and $b < a$ then

$$E(X^{-b}) = \Gamma(a-b)/\Gamma(a) .$$

Proof:

$$\begin{aligned} E(X^{-b}) &= \frac{1}{\Gamma(a)} \int_0^{\infty} X^{-b} e^{-X} X^{a-1} dX \\ &= \frac{\Gamma(a-b)}{\Gamma(a)} \int_0^{\infty} \frac{e^{-X} X^{a-b-1}}{\Gamma(a-b)} dX \\ &= \frac{\Gamma(a-b)}{\Gamma(a)} . \quad /// \end{aligned} \quad (3.8)$$

The second result is an extension of Theorem 3 of Rao & Webster (1966). $\mathcal{P}$

Lemma 3.2: Let $X_1, \dots, X_n$ be independent gamma variates with means $a_1, \dots, a_n$ respectively. Then for $b_i \geq 0$ ($i=1, \dots, n$)

with $\sum_{i=1}^n a_i + \sum_{i=1}^{n-1} b_i > b_n$

$$E \frac{\prod_{i=1}^{n-1} X_i^{b_i}}{(\sum_{i=1}^n X_i)^{b_n}} = \prod_{i=1}^{n-1} \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)} \cdot \frac{\Gamma(\sum_{i=1}^n a_i + \sum_{i=1}^{n-1} b_i - b_n)}{\Gamma(\sum_{i=1}^n a_i + \sum_{i=1}^{n-1} b_i)}$$

Proof: For simplicity of notation the proof is given only for $n = 2$. (The proof is identical for $n > 2$.)

$$\begin{aligned} E \frac{X_1^{b_1}}{(X_1 + X_2)^{b_2}} &= \int_0^\infty \int_0^\infty \frac{X_1^{b_1}}{(X_1 + X_2)^{b_2}} \frac{\exp(-X_1 - X_2) X_1^{a_1-1} X_2^{a_2-1}}{\Gamma(a_1) \Gamma(a_2)} dX_1 dX_2 \\ &= \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)} \int_0^\infty \int_0^\infty \frac{1}{(X_1 + X_2)^{b_2}} \frac{\exp(-X_1 - X_2) X_1^{a_1+b_1-1} X_2^{a_2-1}}{\Gamma(a_1 + b_1) \Gamma(a_2)} dX_1 dX_2 \\ &= \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)} E(X^{-b_2}) \end{aligned} \quad (3.9)$$

where X is a gamma variate with mean $a_1 + a_2 + b_1$. The result now follows from Lemma 3.1. ///

The third result is an extension of Theorem 1 of Rao & Webster (1966).

Lemma 3.3: Let X_1 and X_2 be independent gamma variates with means a_1 and a_2 respectively. Then for $b_1, b_2 \geq 0$ with $a_1 > b_1$ and $a_1 + a_2 > b_1 + b_2$

$$E \frac{1}{X_1^{b_1} (X_1 + X_2)^{b_2}} = \frac{\Gamma(a_1 - b_1)}{\Gamma(a_1)} \frac{\Gamma(a_1 + a_2 - b_1 - b_2)}{\Gamma(a_1 + a_2 - b_1)}.$$

Proof:

$$\begin{aligned} E \frac{1}{X_1^{b_1} (X_1 + X_2)^{b_2}} &= \frac{\Gamma(a_1 - b_1)}{\Gamma(a_1)} \int_0^\infty \int_0^\infty \frac{1}{(X_1 + X_2)^{b_2}} \frac{\exp(-X_1 - X_2) X_1^{a_1 - b_1 - 1} X_2^{a_2 - 1}}{\Gamma(a_1 - b_1) \Gamma(a_2)} dX_1 dX_2 \\ &= \frac{\Gamma(a_1 - b_1)}{\Gamma(a_1)} E(X^{a_1 + a_2 - b_1 - b_2}) \end{aligned} \quad (3.10)$$

where X is a gamma variate with mean $a_1 + a_2 - b_1$. The result now follows from Lemma 3.1. ///

Lemma 3.4: Let X_1, X_2, X_3, X_4 be independent gamma variates with means a_1, a_2, a_3, a_4 respectively with $b_i \geq 0$ ($i = 1, 2, 3, 4$).

Then for $S_k = X_1 + \dots + X_k$ ($k = 2, 3, 4$)

$$E \frac{X_1^{b_1}}{S_2^{b_2} S_3^{b_3}} = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)} \frac{\Gamma(a_2' + b_1 - b_2)}{\Gamma(a_2' + b_1)} \frac{\Gamma(a_3' + b_1 - b_2 - b_3)}{\Gamma(a_3' + b_1 - b_2)}$$

$$E \frac{X_1^{b_1} X_2^{b_2}}{S_3^{b_3} S_4^{b_4}} = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)} \frac{\Gamma(a_2 + b_2)}{\Gamma(a_2)} \frac{\Gamma(a_3' + b_2' - b_3)}{\Gamma(a_3' + b_2')} \frac{\Gamma(a_4' + b_2' - b_3 - b_4)}{\Gamma(a_4' + b_2' - b_3)}$$

and

$$E \frac{x_1^{b_1} x_3^{b_2}}{s_2^{b_3} s_4^{b_4}} = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)} \frac{\Gamma(a_3 + b_2)}{\Gamma(a_3)} \frac{\Gamma(a_2' + b_1 - b_3)}{\Gamma(a_2' + b_1)} \frac{\Gamma(a_4' + b_2 - b_3 - b_4)}{\Gamma(a_4' + b_2 - b_3)}$$

where $a_k' = a_1 + \dots + a_k$, $b_k' = b_1 + \dots + b_k$ ($k = 2, 3, 4$) provided the above expectations exist (e.g., provided $a_2' + b_1 > b_2$ and $a_3' + b_1 > b_2 + b_3$ in the first case).

Proof: Only the proof for the first case is given (the proofs for the remaining two cases are similar):

$$E \frac{x_1^{b_1}}{(x_1 + x_2)^{b_2} (x_1 + x_2 + x_3)^{b_3}} = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)} E \frac{1}{y_1^{b_2} (y_1 + y_2)^{b_3}} \quad (3.11)$$

where Y_1 and Y_2 are independent gamma variates with means $a_2' + b_1$ and a_3 respectively. The result now follows from Lemma 3.3. ///

Noting that all the r_j' have the same expectation the bias of v_4 is

$$B(v_4) = E(v_4) - MSE(r)$$

$$= (g-1) E(r_g' - r)^2 - MSE(r)$$

$$= (g-1) E \left\{ \frac{(n-p)\alpha}{\sum_{i=1}^{n-p} x_i} - \frac{n\alpha}{\sum_{i=1}^n x_i} + \frac{\sum_{i=1}^{n-p} e_i}{\sum_{i=1}^{n-p} x_i} - \frac{\sum_{i=1}^n e_i}{\sum_{i=1}^n x_i} \right\}^2 - MSE(r)$$

$$\begin{aligned}
 &= \left\{ \frac{m^2 [m(g-1)-g]-(m-2)}{h^2 (m-1)^2 (m-2)} - \frac{n^2 (g-1)^2 [m(g-1)-2(g+1)]}{[(m(g-1)-g)[m(g-1)-2g](m-2)} \right\} \alpha^2 \\
 &+ \left\{ \frac{f(t)n(g-2)}{(m+t-1)(m+t-2)} - \frac{f(t)n(g-1)^2 [g(m+t-2)-2m]}{[m(g-1)+(t-1)g][m(g-1)+(t-2)g](m+t-2)} \right\} \delta \\
 &= C_4 \alpha^2 + D_4 \delta \quad (3.12)
 \end{aligned}$$

provided $m > 2g/(g-1)$ using Lemmas 3.1-3.4.

The biases of v_1 , v_2 and v_3 were expressed as $B(v_i) = C_i \alpha^2 + D_i \delta$ ($i = 1, 2, 3$) by Rao & Rao (1971) facilitating comparison of the biases of these three variance estimators in terms of the coefficients C_i and D_i ($i = 1, 2, 3$). Since the expressions for C_3 and D_3 are not in closed form, $B(v_3)$ and $B(v_4)$ are compared for particular values of t , h and n in Tables 3.1 and 3.2 with $g = n$. As v_3 overestimates $MSE(r)$, v_4 is a further overestimate in view of (3.6); however, the ratios C_4/C_3 and D_4/D_3 tend to one as n or h increases.

Rao & Rao note that the minimal values of C_3 and D_3 , for $t = 0$ and 1 , occur at $g = n$ while the minimal value of D_3 for $t = 2$ occurs at a value of g other than n . C_4 and D_4 also follow the same pattern.

Table 3.1. Coefficients C_3 and C_4 in $B(v_3)$ and $B(v_4)$ with $g=n$
(original values multiplied by 1000)

n	h = 1		h = 2		h = 3		h = 4	
	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
4	1935	2500	26	33	3.5	4.4	.9	1.1
6	230	267	6.3	7.3	1.0	1.1	.27	.31
8	76	84	2.6	2.9	.43	.48	.12	.14
10	36	39	1.4	1.5	.24	.26	.070	.076
12	20	22	.86	.92	.15	.16	.044	.048
20	4.8	5.0	.24	.25	.044	.046	.013	.014

Table 3.2. Coefficients D_3 and D_4 in $B(v_3)$ and $B(v_4)$ with $g=n$

(original values multiplied by 1000)

n	t	h = 1		h = 2		h = 3		h = 4	
		D_3	D_4	D_3	D_4	D_3	D_4	D_3	D_4
4	0	1111	1333	36	40	7.7	8.4	2.8	3.0
	1	139	167	25	29	10	11	5.4	6.1
	2	29	50	16	24	11	15	8	11
6	0	148	158	9.4	9.8	2.3	2.4	.88	.91
	1	46	50	9.6	10	4.0	4.2	2.2	2.3
	2	27	32	12	14	8.0	8.8	5.8	6.3
8	0	52	54	4.1	4.2	1.1	1.1	.42	.43
	1	23	24	5.0	5.1	2.1	2.2	1.2	1.2
	2	19	21	8.3	8.8	5.2	5.5	3.8	3.9
10	0	26	26	2.3	2.3	.60	.61	.24	.24
	1	14	14	3.0	3.1	1.3	1.3	.72	.73
	2	14	15	5.8	6.0	3.6	3.7	2.6	2.6
12	0	15	15	1.4	1.4	.39	.39	.16	.16
	1	8.9	9.1	2.0	2.1	.88	.89	.49	.49
	2	10	11	4.3	4.3	2.6	2.6	1.9	1.9
20	0	4.0	4.0	.42	.42	.12	.12	.05	.05
	1	2.9	2.9	.69	.69	.30	.30	.17	.17
	2	4.2	4.3	1.7	1.7	1.0	1.0	.72	.72

3.3. Stability of the Modified Jackknife Variance Estimator

Rao & Rao (1971) derived the mean square errors of v_1 and v_2 under the model (3.4) with normally distributed errors and expressed their results in the form $MSE(v_i) = J_i \alpha^4 + K_i \alpha^2 \delta + L_i \delta^2$ ($i = 1, 2$). While the evaluation of $MSE(v_3)$ appears formidable for arbitrary g , Rao & Rao were able to evaluate $MSE(v_3)$ for $g = 2$. To obtain the mean square error of v_4 for general g (> 2) note that

$$MSE(v_4) = E(v_4^2) - 2MSE(r) B(v_4) - MSE^2(r) \quad (3.13)$$

where $MSE(r)$ and $B(v_4)$ are given in (3.7) and (3.12) respectively. Letting $x' = \sum x_j$, $a' = \sum a_j$ and x_j' and a_j' denote the sample sums in the j^{th} group, define

$$A_j = \left(\frac{n-p}{x_j'} - \frac{n}{x'} \right)$$

$$B_j = \left(\frac{a_j'}{x_j'} - \frac{a'}{x'} \right) \quad (3.14)$$

With this notation and noting that the errors e are normally distributed

$$\frac{g^2}{(g-1)^2} E(v_4^2) = E \left[\sum_{j=1}^g (\alpha A_j + B_j)^2 \right]^2$$

$$\begin{aligned}
 &= E\{\alpha^4 (\sum_j A_j^4 + \sum_{j \neq i} A_j^2 A_i^2) \\
 &\quad + 6\alpha^2 \sum_j A_j^2 B_j^2 + 4\alpha^2 \sum_{j \neq i} A_j B_j A_i B_i + 2\alpha^2 \sum_{j \neq i} A_j^2 B_i^2 \\
 &\quad + \sum_j B_j^4 + \sum_{j \neq i} B_j^2 B_i^2 \} \\
 &= J'_4 \alpha^4 + K'_4 \alpha^2 \delta + L'_4 \delta^2 \quad (\text{say}). \quad (3.15)
 \end{aligned}$$

After considerable algebra the coefficients J'_4 , K'_4 and L'_4 in (3.15) may be expressed using Lemmas 3.1-3.4 as follows:

$$\begin{aligned}
 J'_4 &= \frac{g(n-p)}{((n-p)h-1)[4]} - \frac{4gn(n-p)^3}{((n-p)h-1)[3](m-4)} \\
 &\quad + \frac{2g(g+2)n^2(n-p)^2}{((n-p)h-1)[2](m-3)[2]} - \frac{4g^2n^3(n-p)}{((n-p)h-1)(m-2)[3]} + \frac{g^2n^4}{(m-1)[4]} \\
 &\quad + g(g-1)(n-p)^4 E_{220}(a,a,b) - 4g(g-1)n(n-p)^3 E_{211}(a,a,b) \\
 &\quad + 4g(g-1)n^2(n-p)^2 E_{112}(a,a,b), \quad (3.16)
 \end{aligned}$$

$$\begin{aligned}
 K'_4 &= \frac{6g(n-p)^3 f(t)}{((n-p)h+t-1)[4]} - \frac{12g(n-p)^2 (2n-p)f(t)}{((n-p)h+t-1)[3](m+t-4)} \\
 &\quad + \frac{2g(g+2)(6n^2-6np+p^2)(n-p)f(t)}{((n-p)h+t-1)[2](m+t-3)[2]}
 \end{aligned}$$

$$+ \frac{2g(g+2) p(n-p)^2 f(t)}{((n-p)h-1)^{[2]} (m+t-3)^{[2]}} - \frac{12g^2 n(n-p)(2n-p) f(t)}{((n-p)h+t-1)(m+t-2)^{[3]}}$$

$$- \frac{12g^2 np(n-p) f(t)}{((n-p)h-1)(m+t-2)^{[3]}} + \frac{6g^2 n^3 f(t)}{(m+t-1)^{[4]}}$$

$$+ 6g(g-1)(n-p)^2(n-2p)f(t) E_{220}(a, a, b+t)$$

$$+ 2g(g-1)p(n-p)^2 f(t) E_{220}(a, a+t, b)$$

$$- 4g(g-1)p(3n-2p)(n-p) f(t) E_{211}(a+t, a, b)$$

$$- 4g(g-1)p(n-p)^2 f(t) E_{211}(a, a+t, b)$$

$$- 12g(g-1)(n-p)(n-2p)(2n-p) f(t) E_{211}(a, a, b+t)$$

$$+ 8g(g-1)p(n-p)(3n-p) f(t) E_{112}(a, a+t, b)$$

$$+ 4g(g-1)(n-2p)(6n^2 - 6np + p^2) f(t) E_{112}(a, a, b+t), \quad (3.17)$$

$$L'_4 = \frac{3g(n-p)[f(2t) + (n-p-1)f^2(t)]}{((n-p)h + 2t - 1)^{[4]}}$$

$$- \frac{12g(n-p)[f(2t) + (n-p-1)f^2(t)]}{((n-p)h + 2t - 1)^{[3]}(m + 2t - 4)}$$

$$+ \frac{6g(g+2)(n-p)[f(2t) + (n-p-1)f^2(t)]}{((n-p)h + 2t - 1)^{[2]}(m + 2t - 3)^{[2]}}$$

$$+ \frac{2g(g+2)p(n-p)f^2(t)}{((n-p)h + t - 1)^{[2]}(m + 2t - 3)^{[2]}}$$

$$- \frac{12g^2(n-p)[f(2t) + (n-p-1)f^2(t)]}{((n-p)h + 2t - 1)(m + 2t - 2)^{[3]}}$$

$$- \frac{12g^2p(n-p)f^2(t)}{((n-p)h + t - 1)(m + 2t - 2)^{[3]}}$$

$$+ \frac{3g^2n[f(2t) + (n-1)f^2(t)]}{(m + 2t - 1)^{[4]}}$$

$$+ g(g-1)p^2f^2(t) E_{220}(a+t, a+t, b)$$

$$+ 2g(g-1)p(n-2p)f^2(t) E_{220}(a+t, a, b+t)$$

$$+ 3g(g-1)(n-2p)[f(2t) + (n-2p-1)f^2(t)] E_{220}(a, a, b+2t)$$

$$- 4g(g-1)p^2f^2(t) E_{211}(a+t, a+t, b)$$

$$- 4g(g-1)p(n-2p)f^2(t) E_{211}(a, a+t, b+t)$$

$$- 12g(g-1)p(n-2p)f^2(t) E_{211}(a+t, a, b+t)$$

$$- 12g(g-1)(n-2p)[f(2t) + (n-2p-1)f^2(t)] E_{211}(a, a, b+2t)$$

$$\begin{aligned}
 & + 8g(g-1) p^2 f^2(t) E_{112}(a+t, a+t, b) \\
 & + 24g(g-1) p(n-2p) f^2(t) E_{112}(a, a+t, b+t) \\
 & + 12g(g-1)(n-2p) [f(2t) + (n-2p-1)f^2(t)] E_{112}(a, a, b+2t), \quad (3.18)
 \end{aligned}$$

provided $m > 4 + ph$, where $z^{[k]} = z(z-1) \dots (z-k+1)$, $a = ph$,
 $b = (n-2p)h$ and

$$E_{b_1 b_2 b_3}(a_1, a_2, a_3) = E[(X_1 + X_3)^{-b_1} (X_2 + X_3)^{-b_2} (X_1 + X_2 + X_3)^{-b_3}] \quad (3.19)$$

where X_1 , X_2 and X_3 are independent gamma variates with means a_1 , a_2 and a_3 respectively.

In order to express the mean square error of v_4 in the form

$$MSE(v_4) = J_4 \alpha^4 + K_4 \alpha^2 \delta + L_4 \delta^2 \quad (3.20)$$

it remains only to evaluate the expectations E_{220} , E_{211} and E_{112} in (3.16), (3.17) and (3.18). Explicit expressions for E_{110} , E_{210} and E_{220} are provided in Theorems 3.5-3.7 below. Expressions for E_{112} and E_{211} are then easily obtained from the expressions for E_{110} and E_{210} respectively using Theorem 3.8.

The first theorem generalizes Theorem 2 of Rao & Webster (1966). While this generalization has already been made by P.S.R.S.

Rao (1974, p. 86) the form given below may be preferred as it involves fewer terms.

Theorem 3.5: Let X_1, X_2, X_3 be independent gamma variates with means a, b, c respectively. Then provided a, b, c are positive integers with $a \leq b$

$$E_{110}(a, b, c) = \frac{\Gamma(a+b+c-2)}{\Gamma(a)\Gamma(b)\Gamma(c)} (I_1 + I_2 + I_3 + I_4)$$

where

$$I_1 = \sum_{k=1}^{a-1} (-1)^{k+1} \frac{\Gamma(a-k) \Gamma(b-k) \Gamma(c+k-1)}{\Gamma(a+b+c-k-1)},$$

$$I_2 = (-1)^a \Gamma(c+a-1) \sum_{k=1}^{b-a} \frac{\Gamma(b-a-k+1)}{\Gamma(c+b-k)} \frac{1}{(c+b-k-1)},$$

$$I_3 = 2(-1)^{a+1} \sum_{k=1}^{c+a-2} \frac{(-1)^{k+1}}{(c+a-k-1)^2},$$

$$I_4 = (-1)^{c+1} \frac{\pi^2}{6}.$$

Proof: The following notation will be useful in the proof:

$$C_{b_1 b_2}(a, b, c) = \int_0^1 \int_{1-s_1}^1 \frac{(1-s_2)^{a-1} (1-s_1)^{b-1} (s_1+s_2-1)^{c-1}}{s_1^{b_1} s_1^{b_2}} ds_2 ds_1, \quad (3.21)$$

Consider the transformation $s_1 = W_1/W_3$, $s_2 = W_2/W_3$ and

$W_3 = X_1 + X_2 + X_3$ where $W_1 = X_1 + X_3$ and $W_2 = X_2 + X_3$. Since the Jacobian of the transformation is W_3^2 , integrating over W_3 yields

$$E_{110}(a, b, c) = \frac{\Gamma(a+b+c-2)}{\Gamma(a) \Gamma(b) \Gamma(c)} C_{11}(a, b, c). \quad (3.22)$$

For $a \geq 2$,

$$\begin{aligned} C_{11}(a, b, c) &= C_{00}(a-1, b-1, c) - C_{11}(a-1, b-1, c+1) \\ &= \sum_{k=1}^{a-1} (-1)^{k+1} C_{00}(a-k, b-k, c+k-1) \\ &\quad + (-1)^{a+1} C_{11}(1, b-a+1, c+a-1), \end{aligned} \quad (3.23)$$

where

$$C_{00}(a, b, c) = \frac{\Gamma(a) \Gamma(b) \Gamma(c)}{\Gamma(a+b+c)}. \quad (3.24)$$

For any integer $b' \geq 2$,

$$\begin{aligned} C_{11}(1, b', c+a-1) &= C_{11}(1, b'-1, c+a-1) - C_{01}(1, b'-1, c+a-1) \\ &= C_{11}(1, 1, c+a-1) - \sum_{k=1}^{b'-1} C_{01}(1, b'-k, c+a-1), \end{aligned} \quad (3.25)$$

where

$$C_{01}(1, b, c) = \frac{\Gamma(b) \Gamma(c)}{\Gamma(b+c)} \frac{1}{(b+c-1)}. \quad (3.26)$$

For any integer $c' \geq 2$,

$$\begin{aligned}
 c_{11}(1,1,c') &= \frac{2}{(c'-1)^2} - c_{11}(1,1,c'-1) \\
 &= 2 \sum_{k=1}^{c'-1} \frac{(-1)^{k+1}}{(c'-k)^2} + (-1)^{c'+1} \frac{\pi^2}{6} \quad (3.27)
 \end{aligned}$$

since $c_{11}(1,1,1) = \pi^2/6$. The result now follows from (3.22) - (3.27). ///

Theorem 3.6: Let X_1, X_2, X_3 be independent gamma variates with means a, b, c respectively and let $d = \min \{a, b\}$. Then provided a, b, c are positive integers with $c \geq 2$

$$E_{210}(a,b,c) = \frac{\Gamma(a+b+c-3)}{\Gamma(a)\Gamma(b)\Gamma(c)} (I_1 + I_2 + I_3 + I_4)$$

where

$$I_1 = \frac{1}{(a+c-2)} \sum_{k=1}^{d-1} (-1)^{k+1} \frac{\Gamma(a-k) \Gamma(b-k) \Gamma(c+k-1)}{\Gamma(a+b+c-k-2)},$$

$$I_2 = (-1)^{d+1} \sum_{k=1}^{c+d-3} (-1)^{k+1} \left[\frac{2k}{(c+d-k-2)^2} + \frac{1}{(c+d-k-1)(c+d-k-2)} \right],$$

$$I_3 = (-1)^c (c+d-2) \frac{\pi^2}{6} + (-1)^{c+1},$$

$$I_4 = \begin{cases} 2(-1)^a \sum_{k=1}^{c+a-3} \frac{(-1)^{k+1} (b-a)}{(c+a-k-1)^2} + (-1)^{a+1} \sum_{k=1}^{b-a-1} \frac{k \Gamma(b-a-k) \Gamma(c+a-1)}{\Gamma(b+c-k-1) (b+c-k-2)} \\ \quad + (-1)^c (b-a) \frac{\pi^2}{6} + 2(-1)^{c+1} (b-a) , & \text{if } a \leq b \\ (-1)^b \Gamma(c+b-1) \sum_{k=1}^{a-b} \frac{\Gamma(a-b-k+1)}{\Gamma(a+c-k)} \frac{1}{(a+c-k-2)} , & \text{if } b \leq a . \end{cases}$$

Proof: Consider first the case $a \leq b$. Applying the same transformation as in Theorem 3.5 and integrating over W_3 yields

$$E_{210}(a, b, c) = \frac{\Gamma(a+b+c-3)}{\Gamma(a)\Gamma(b)\Gamma(c)} C_{21}(a, b, c) . \quad (3.28)$$

For $a \geq 2$,

$$\begin{aligned} C_{21}(a, b, c) &= C_{10}(a-1, b-1, c) - C_{21}(a-1, b-1, c+1) \\ &= \sum_{k=1}^{a-1} (-1)^{k+1} C_{10}(a-k, b-k, c+k-1) \\ &\quad + (-1)^{a+1} C_{21}(1, b-a+1, c+a-1) , \end{aligned} \quad (3.29)$$

where

$$C_{10}(a, b, c) = \frac{\Gamma(a) \Gamma(b) \Gamma(c)}{\Gamma(a+b+c-1)} \frac{1}{(a+c-1)} . \quad (3.30)$$

For any integer $b' \geq 2$,

$$\begin{aligned}
 C_{21}(1, b', c+a-1) &= C_{21}(1, b'-1, c+a-1) - C_{11}(1, b'-1, c+a-1) \\
 &= C_{21}(1, 1, c+a-1) - \sum_{k=1}^{b'-1} C_{11}(1, b'-k, c+a-1). \quad (3.31)
 \end{aligned}$$

For any integer $c' \geq 3$,

$$\begin{aligned}
 C_{21}(1, 1, c') &= C_{11}(1, 1, c'-1) + \frac{1}{(c'-1)(c'-2)} - C_{21}(1, 1, c'-1) \\
 &= \sum_{k=1}^{c'-2} (-1)^{k+1} \left[C_{11}(1, 1, c'-k) + \frac{1}{(c'-k)(c'-k-1)} \right] \\
 &\quad + (-1)^{c'} \left(\frac{\pi^2}{6} - 1 \right) \quad (3.32)
 \end{aligned}$$

since $C_{21}(1, 1, 2) = (\pi^2/6) - 1$. The case $a \leq b$ now follows by combining (3.28) - (3.32) and substituting the expressions given in Theorem 3.5 for the C_{11} terms.

Consider now the case $b \leq a$. For $b \geq 2$,

$$C_{21}(a, b, c) = \sum_{k=1}^{b-1} (-1)^{k+1} C_{10}(a-k, b-k, c+k-1) + (-1)^{b+1} C_{21}(a-b+1, 1, c+b-1) \quad (3.33)$$

where C_{10} is evaluated explicitly in (3.30). For any integer $a' \geq 2$,

$$\begin{aligned}
 C_{21}(a', 1, c+b-1) &= C_{21}(a'-1, 1, c+b-1) - C_{20}(a'-1, 1, c+b-1) \\
 &= C_{21}(1, 1, c+b-1) - \sum_{k=1}^{a'-1} C_{20}(a'-k, 1, c+b-1) \quad (3.34)
 \end{aligned}$$

where

$$C_{20}(a, 1, c) = \frac{\Gamma(a) \Gamma(c)}{\Gamma(a+c-1)} \frac{1}{(a+c-1)(a+c-2)} \quad (3.35)$$

The case $b \leq a$ now follows from (3.28), (3.33) - (3.35) and (3.32) and substituting the expressions given in Theorem 3.5 for the C_{11} terms. ///

The next theorem generalizes Theorem 3.2.1 of Chakrabarty (1968).

Theorem 3.7: Let X_1, X_2, X_3 be independent gamma variates with means a, b, c respectively. Then provided a, b, c are positive integers with $a \leq b$, $c \geq 2$ and $c+b \geq 4$

$$K_{220}(a, b, c) = \frac{\Gamma(a+b+c-4)}{\Gamma(a)\Gamma(b)\Gamma(c)} (I_1 + I_2 + I_3 + I_4 + I_5)$$

where

$$I_1 = \sum_{k=1}^{a-2} (-1)^{k+1} \frac{k\Gamma(a-k-1) \Gamma(b-k-1) \Gamma(c+k-1)}{\Gamma(a+b+c-k-3)},$$

$$I_2 = (a-1)(-1)^{a+1} \Gamma(c+a-2) \sum_{k=1}^{b-a} \frac{\Gamma(b-a-k+1)}{\Gamma(c+b-k-1)} \frac{1}{(c+b-k-2)} ,$$

$$I_3 = (-1)^{a+1} \Gamma(c+a-1) \sum_{k=1}^{b-a-1} \frac{k \Gamma(b-a-k)}{\Gamma(c+b-k-1)(c+b-k-3)} ,$$

$$I_4 = (-1)^{a+1} \sum_{k=1}^{c+a-3} (-1)^k \left[\frac{2(a-1)+2k(k-1)+2k(b-a)}{(c+a-k-2)^2} + \frac{2(k-1)+(b-a)}{(c+a-k-1)(c+a-k-2)} \right]$$

$$I_5 = (-1)^{c+1} [(c+a-3)(c+b-4) + (2a+b+2c-7)] \frac{\pi^2}{6} + (a+b+2c-5)(-1)^c .$$

Proof: Applying the same transformation as in Theorem 3.5 and integrating over W_3 yields

$$E_{220}(a,b,c) = \frac{\Gamma(a+c-4)}{\Gamma(a) \Gamma(b) \Gamma(c)} C_{22}(a, b, c) . \quad (3.36)$$

For $a \geq 2$,

$$\begin{aligned} C_{22}(a,b,c) &= C_{11}(a-1,b-1,c) - C_{22}(a-1,b-1,c+1) \\ &= \sum_{k=1}^{a-1} (-1)^{k+1} C_{11}(a-k,b-k,c+k-1) \\ &\quad + (-1)^{a+1} C_{22}(1, b-a+1, c+a-1) . \end{aligned} \quad (3.37)$$

For any integer $b' \geq 2$,

$$\begin{aligned} C_{22}(1,b',c+a-1) &= C_{22}(1,b'-1,c+a-1) - C_{12}(1,b'-1,c+a-1) \\ &= C_{22}(1,1,c+a-1) - \sum_{k=1}^{b'-1} C_{12}(1,b'-k,c+a-1) . \end{aligned} \quad (3.38)$$

For any integer $b'' \geq 2$,

$$\begin{aligned} C_{12}(1, b'', c+a-1) &= C_{12}(1, b''-1, c+a-1) - C_{02}(1, b''-1, c+a-1) \\ &= C_{12}(1, 1, c+a-1) - \sum_{k=1}^{b''-1} C_{02}(1, b''-k, c+a-1) \end{aligned} \quad (3.39)$$

where $C_{12}(1, 1, c+a-1) = C_{21}(1, 1, c+a-1)$ is evaluated in (3.32) and

$$C_{02}(1, b, c) = \frac{\Gamma(b) \Gamma(c)}{\Gamma(b+c-1)} \frac{1}{(b+c-1)(b+c-2)}. \quad (3.40)$$

Finally for any integer $c' \geq 3$,

$$\begin{aligned} C_{22}(1, 1, c') &= 2C_{21}(1, 1, c'-1) - C_{22}(1, 1, c'-1) \\ &= 2 \sum_{k=1}^{c'-2} (-1)^{k+1} C_{21}(1, 1, c'-k) + (-1)^{c'}. \end{aligned} \quad (3.41)$$

since $C_{22}(1, 1, 2) = 1$ with C_{21} evaluated in (3.32). The result now follows by combining (3.36) - (3.41) and substituting the expressions given in Theorem 3.5 for the C_{11} terms. ///

Theorem 3.8: Let X_1, X_2, X_3 be independent gamma variates with means a, b, c respectively. Then for $b_1, b_2, b_3 \geq 0$ with $a + b + c > b_1 + b_2 + b_3$

$$E_{b_1 b_2 b_3}(a, b, c) = \frac{\Gamma(a+b+c-b_1-b_2-b_3)}{\Gamma(a+b+c-b_1-b_2)} E_{b_1 b_2 0}(a, b, c) .$$

Proof: Applying the same transformation as in Theorem 3.5 and integrating over W_3 yields

$$\begin{aligned} E_{b_1 b_2 b_3}(a, b, c) &= \frac{\Gamma(a+b+c-b_1-b_2-b_3)}{\Gamma(a) \Gamma(b) \Gamma(c)} C_{b_1 b_2}(a, b, c) \\ &= \frac{\Gamma(a+b+c-b_1-b_2-b_3)}{\Gamma(a+b+c-b_1-b_2)} E_{b_1 b_2 0}(a, b, c) . \quad /// \quad (3.42) \end{aligned}$$

The coefficients J_4 , K_4 and L_4 in (3.20) were found to decrease significantly with g for $3 \leq g \leq n$ as illustrated in Tables 3.3 and 3.4 for $n = 8$ and $n = 12$ respectively. For completeness the coefficients J_3 , K_3 and L_3 are also included in these tables. (Derivation of the expressions for J_4 , K_4 and L_4 for the special case $g = 2$ has been omitted as the details involved are tedious while the values of these coefficients are expected to be close to those of J_3 , K_3 and L_3 with $g = 2$. In this case Rao & Rao (1971) found the stability of v_3 to be poor in comparison to v_1 and v_2 .)

The coefficients in the mean square errors of v_1 , v_2 and v_4 with $g = n$ are given in Tables 3.5-3.7. These coefficients decrease as n or h increase so that the mean square errors of the variance estimators decrease as the sample size increases or

as the coefficient of variation of the x population decreases.

Tables 3.5 and 3.6 indicate that $J_1 < J_2 < J_4$ and $K_1 < K_2 < K_4$ for all values of t, h and n . From Table 3.7, $L_1 < L_2 < L_4$ for $t = 0$ and 1 in which case v_1 is more stable than v_2 which in turn is more stable than v_4 . Note however that for $t = 0$ and moderately large values of m , L_4/L_2 is close to one so that under these conditions v_2 and v_4 are of comparable stability when $\alpha = 0$. When $t = 1$, L_2/L_1 is close to one for moderately large values of m in which case v_1 and v_2 are of comparable stability when $\alpha = 0$ (Rao & Rao [1971]).

For $t = 2$, $L_2 < L_1 < L_4$ except for large m where $L_2 < L_4 < L_1$. For moderately large m however, L_4/L_1 is close to one. Thus for $t = 2$ and $\alpha = 0$, v_2 is the most stable while v_1 and v_4 are of comparable stability for moderate m . The comparison between v_1, v_2 and v_4 for $t = 2$ depends on α^2/δ when $\alpha \neq 0$.

Table 3.3. Coefficients J_4 , K_4 and L_4 in MSE (v_4) for $g = 4, 8$ and

Coefficients J_3 , K_3 and L_3 in MSE (v_3) for $g = 2$ ($n = 8$)

(original values multiplied by 10^6)

g	$J_4(J_3)$		$K_4(K_3)$		$L_4(L_3)$	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$
0	+	4783	180	+	16651	983
	2149993	2057	79	2984801	6061	370
	776548	966	38	1034910	2849	180
1	+	4783	180	+	12642	1625
	2149993	2057	79	535273	6465	755
	776548	966	38	265638	3317	377
2	+	4783	180	+	18419	4125
	2149993	2057	79	430844	14684	2638
	776548	966	38	310525	9599	1603
	+	8367	928	+	11502	4446
	436732	2904	331	48151	5619	1940
	167400	1462	166	28965	3065	999
	+	57167	48136	+	57167	48136
	92025	41093	28395	92025	41093	28395
	79583	29855	18611	79583	29855	18611

+ Not defined for $nh < 8$.

Table 3.4. Coefficients J_4 , K_4 and L_4 in $MSE(v_4)$ for $g = 3, 4, 6, 12$ and
Coefficients J_3 , K_3 and L_3 in $MSE(v_3)$ for $g = 2$ ($n = 12$)

(original values multiplied by 10^6)

t	g	$J_4(J_3)$			$K_4(K_3)$			$L_4(L_3)$		
		h = 1	h = 2	h = 3	h = 1	h = 2	h = 3	h = 1	h = 2	h = 3
0	2	377484	911	48	627117	3316	271	125424	2089	298
	3	239092	614	30	344794	1941	153	63750	1116	157
	4	138128	398	20	195682	1250	101	38622	745	105
	6	81034	249	13	113486	783	63	23875	478	67
	12	46330	143	7	65140	457	36	14630	281	37
1	2	377484	911	48	146461	3655	555	23833	4446	1812
	3	239092	614	30	117359	2487	346	17571	2745	1047
	4	138128	398	20	78157	1698	236	12847	1940	728
	6	81034	249	13	50914	1096	150	8984	1282	468
	12	46330	143	7	31310	632	82	5811	741	255
2	2	377484	911	48	104241	6961	1657	43063	27283	22466
	3	239092	614	30	123535	5856	1203	41845	20394	14984
	4	138128	398	20	99075	4452	886	36701	16219	11367
	6	81034	249	13	77612	3243	619	31511	12456	8230
	12	46330	143	7	58883	2210	394	26480	9103	5512

Table 3.5. Coefficients J_1 , J_2 and J_4 in $MSE(v_1)$, $MSE(v_2)$ and $MSE(v_4)$ with g^n

(original values multiplied by 10^6)

n	h = 1			h = 2			h = 3			h = 4		
	J_1	J_2	J_4	J_1	J_2	J_4	J_1	J_2	J_4	J_1	J_2	J_4
6	74286	262857	18161413	311	777	5065	19	42	150	3.1	5.8	15.5
8	20230	56279	776548	115	262	966	8	15	38	1.3	2.2	4.4
10	8094	20806	144559	55	119	323	3.9	7.2	14.6	.6	1.1	1.8
12	4011	9921	46330	30	64	143	2.2	4.0	7.0	.36	.59	.92
20	657	1522	3553	6	12	19	.47	.78	1.08	.08	.12	.15

Table 3.6. Coefficients K_1 , K_2 and K_4 in $MSE(v_1)$, $MSE(v_2)$ and $MSE(v_4)$ with $g = n$

(original values multiplied by 10^6)

n	t	h = 1			h = 2			h = 3			h = 4		
		K_1	K_2	K_4	K_1	K_2	K_4	K_1	K_2	K_4	K_1	K_2	K_4
6	0	97143	840000	25184615	1283	4227	14186	135	320	674	29	57	98
	1	45714	125714	1925781	1369	3156	12728	282	473	1240	87	131	274
	2	64515	76369	1174294	4847	5534	29105	1245	1372	4567	486	525	1351
8	0	30289	159732	1034910	490	1331	2849	53	111	180	12	21	29
	1	16272	39398	265638	660	1206	3317	120	187	377	37	52	90
	2	29797	34799	310525	2373	2631	9599	590	635	1603	224	237	490
10	0	13274	55274	196496	237	577	996	26	51	72	6	10	13
	1	7770	17448	76899	341	586	1297	62	92	161	19	26	40
	2	16966	19461	121583	1346	1464	4228	326	345	734	121	127	228
12	0	6989	25240	65140	132	300	457	15	27	36	3	5	6
	1	4359	9278	31310	199	328	632	36	52	82	11	15	21
	2	10786	12173	58883	839	900	2210	199	209	394	73	76	124
20	0	1272	3545	5596	26	53	65	3	5	6	.7	1.0	1.1
	1	922	1739	3535	44	67	98	8	11	14	2.4	3.1	3.8
	2	2979	3225	8860	213	222	388	48	49	73	17	18	24

Table 3.7. Coefficients L_1 , L_2 and L_4 in $MSE(v_1)$, $MSE(v_2)$ and $MSE(v_4)$ with $g = n$
(original values multiplied by 10^6)

n	t	h = 1			h = 2			h = 3			h = 4		
		L_1	L_2	L_4	L_1	L_2	L_4	L_1	L_2	L_4	L_1	L_2	L_4
6	0	28095	450000	3256913	928	4024	5969	161	445	541	48	105	120
	1	21746	31429	105979	4040	4803	8984	1607	1800	2759	852	927	1282
	2	125278	45916	164240	56922	28796	67604	39019	23080	43608	31026	20168	33461
8	0	8696	79144	167400	347	1212	1462	62	150	166	19	37	40
	1	9117	11833	28965	1692	1919	3065	668	726	999	352	374	477
	2	61039	25373	79583	26233	14495	29855	17443	11006	18611	13604	9297	14039
10	0	3750	26022	38599	165	512	571	30	67	72	9	17	18
	1	4679	5727	11689	865	955	1388	339	362	468	178	187	227
	2	34717	15768	43822	14298	8373	15584	9301	6117	9540	7159	5049	7133
12	0	1944	11465	14630	91	261	281	17	36	37	5	9	10
	1	2720	3208	5811	500	543	741	195	206	255	102	107	125
	2	21765	10562	26480	8666	5288	9103	5546	3754	5512	4227	3047	4100
20	0	340	1494	1605	18	44	45	3	7	7	1	2	2
	1	597	657	939	108	114	137	42	43	49	22	22	25
	2	5689	3220	6142	2079	1387	1986	1281	921	1180	955	720	870

3.4. Summary

The stability of the jackknife variance estimator v_4 (as an estimator of the mean square error of the classical ratio estimator) is evaluated in small samples under the linear regression model in (3.4) and compared with that of the two commonly used linearization variance estimators v_1 and v_2 .

Evaluation of the mean square error of the jackknife variance estimator v_3 has so far proven intractable, except for the special case $g = 2$ (Chakrabarty [1968], Rao & Rao [1971]). In view of the simple relationship between v_4 and v_3 noted in (3.6), however, the difference between these two jackknife variance estimators in terms of bias and mean square error is not expected to be great. A comparison of $B(v_4)$ and $B(v_3)$ with $g = n$ confirms this expectation in the case of bias (Tables 3.1 and 3.2).

The mean square error of v_4 decreases significantly with g so that $g = n$ would be the optimal choice (Tables 3.3 and 3.4). (This conclusion was also reached by Rao & Beagle [1967] on the basis of Monte Carlo results.) The mean square errors of v_1 , v_2 and v_4 (with $g = n$) decrease as the sample size n increases or as the coefficient of variation C_x of the x population decreases (Tables 3.5-3.7).

For $t = 0$ or 1 , v_1 is more stable than v_2 which

in turn is more stable than v_4 with $g = n$. (Recall that since $V(e_1 | x_1) = \delta x_1^t$ in (3.4), the parameter t reflects the degree of heteroskedasticity of the error variance.) When the regression of y on x passes through the origin, v_4 and v_2 are of comparable stability for $t = 0$, provided n is moderately large or C_x is relatively small (Table 3.7).

For $t = 2$, results concerning the mean square errors of v_1 , v_2 and v_4 are not clearcut. When the regression of y on x passes through the origin, however, v_2 is the most stable while v_4 and v_1 are of comparable stability for moderate n or relatively small C_x .

The above conclusions are in close agreement with those drawn by Rao & Kuzik (1974) in a semi-empirical study using the same model but with the x values chosen from a variety of natural and synthetic populations.

CHAPTER 4

STRATIFIED SAMPLING: LARGE SAMPLE RESULTS

Many large scale surveys involve large numbers of strata with relatively few primary sampling units (p.s.u.'s) selected within each stratum. For example, the Nutrition Canada (1973) survey involved 110 strata with between two and eleven p.s.u.'s selected in each stratum. The asymptotic results for stratified samples presented in this chapter were developed primarily with such situations in mind. Results are obtained as $L \rightarrow \infty$ within the context of a sequence $\{\pi_L\}$ of finite populations with L strata in π_L .

In section 4.1 some general large sample theory for independent nonidentically distributed random variables is outlined. The large sample theory for stratified samples developed in subsequent sections of this chapter is based on the results of this section.

In section 4.2, the fundamental large sample theory for stratified simple random sampling is developed in the linear case. Regularity conditions are detailed under which the consistency and asymptotic normality of the usual estimator of the population mean are established. Consistency of the usual variance estimator is also shown.

The results of section 4.2 are extended to nonlinear statistics in section 4.3. The variance estimator obtained by the linearization method is shown to be consistent. Specific examples involving ratios and correlation coefficients are provided.

In sections 4.4 and 4.5 applications of the jackknife and balanced repeated replication (BRR) methods in the nonlinear case are described with the asymptotic properties of these methods investigated as in sections 4.2 and 4.3.

Extensions of the results of sections 4.2 - 4.5 to stratified multi-stage samples are indicated in section 4.6. (The results remain valid for any sample design in which the p.s.u.'s are selected with replacement.)

A summary of the results obtained is provided in section 4.7.

4.1. Some Large Sample Theory for Independent Random Variables

Consider a double sequence of p -dimensional random variables $(X_{Lh} = (X_{Lh1}, \dots, X_{Lhp}) ; h = 1, \dots, H_L ; L = 1, 2, \dots)$ such that $H_L \rightarrow \infty$ as $L \rightarrow \infty$ and where $X_{L1}, \dots, X_{LH_L}$ are independent for each L (rowwise independence). For $h = 1, \dots, H_L$ and $L = 1, 2, \dots$ let $A_{Lh} = E(X_{Lh}) = (A_{Lh1}, \dots, A_{Lhp})$ and let

$\Gamma_{Lh} = E(X_{Lh} - A_{Lh})(X_{Lh} - A_{Lh})'$. For $L = 1, 2, \dots$ let

$$\bar{A}_L = H_L^{-1} \sum_{h=1}^{H_L} A_{Lh} = (\bar{A}_{L1}, \dots, \bar{A}_{Lp}) \quad \text{and let} \quad \bar{X}_L = H_L^{-1} \sum_{h=1}^{H_L} X_{Lh} = (\bar{X}_{L1}, \dots, \bar{X}_{Lp}).$$

Then the following result will be used in establishing consistency of estimators to be defined in subsequent sections.

Theorem 4.1: If $H_L^{-1} \sum_{h=1}^{H_L} E|X_{Lhk}|^{1+\delta} = O(1)$ as $L \rightarrow \infty$ for some

$\delta > 0$, then, for any $\epsilon > 0$,

$$P(|\bar{X}_{Lk} - \bar{A}_{Lk}| \geq \epsilon) = O(H_L^{-r})$$

for $k = 1, \dots, p$ where $r = \delta$ for $\delta \leq 1$ and $r = (1+\delta)/2$ for $\delta > 1$.

Proof: See Sen (1970, p. 2141).

A central limit theorem for double sequences of rowwise independent p -dimensional random variables is now given.

Theorem 4.2: If

$$(1) \quad \lim_{L \rightarrow \infty} H_L^{-1} \sum_{h=1}^{H_L} \Gamma_{Lh} = \bar{\Gamma} \quad \text{with } \bar{\Gamma} \text{ positive definite and}$$

$$(11) \quad \lim_{L \rightarrow \infty} H_L^{-(1+\delta/2)} \sum_{h=1}^{H_L} E|\lambda(X_{Lh} - A_{Lh})'|^{2+\delta} = 0 \quad \text{for all}$$

$$\lambda = (\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$$

then $\sqrt{H_L} (\bar{X}_L - \bar{A}_L) \xrightarrow{L} N(0, \bar{\Gamma})$ as $L \rightarrow \infty$.

Proof: Hoadley's (1971, p.1990) proof for single sequences of independent random variables using Liapounov's central limit theorem also holds for double sequences of rowwise independent random variables. (See Gnedenko [1962, p.310] and Chung [1974, p.205] for Liapounov's theorem in the case of double sequences.) ///

A slightly weaker version of Theorem 4.2 which leads to more readily interpreted regularity conditions in the central limit theorems for stratified samples is also given (Theorem 4.5). The following two well-known inequalities (Lemmas 4.3 and 4.4) are required.

Lemma 4.3 (Jensen's inequality): Let $f(\cdot)$ be a real-valued convex function on (a, b) where the cases $a = -\infty$ and $b = +\infty$ are not excluded. Then for $W_i \geq 0$ ($i = 1, \dots, n$) with $\sum W_i = 1$

$$f(\sum W_i x_i) \leq \sum W_i f(x_i).$$

Proof: See Rudin (1966, p. 61).

Lemma 4.4: Let $X_1, \dots, X_n$ be arbitrary random variables. Then

$$\text{for } r > 1, \quad E \left| \sum_{i=1}^n X_i \right|^r \leq n^{r-1} \sum_{i=1}^n E |X_i|^r.$$

Proof: Since $f(x) = |x|^r$ is convex for, $r > 1$, Jensen's inequality yields

$$|n^{-1} \sum X_i|^r \leq n^{-1} \sum |X_i|^r. \quad (4.1)$$

Taking expectations in (4.1) yields the result. ///

Theorem 4.5: If

$$(i) \lim_{L \rightarrow \infty} H_L^{-1} \sum_{h=1}^{H_L} \Gamma_{Lh} = \bar{\Gamma} \text{ with } \bar{\Gamma} \text{ positive definite and}$$

$$(ii) H_L^{-1} \sum_{h=1}^{H_L} E|X_{Lhk} - A_{Lhk}|^{2+\delta} = o(1) \text{ as } L \rightarrow \infty$$

for some $\delta > 0$ and $k = 1, \dots, p$

then $\sqrt{H_L}(\bar{X}_L - \bar{A}_L) \xrightarrow{L} N(0, \bar{\Gamma})$ as $L \rightarrow \infty$.

Proof: Since (using Lemma 4.4)

$$\begin{aligned} & \frac{1}{H_L^{1+\delta/2}} \sum_{h=1}^{H_L} E|\lambda(X_{Lh} - A_{Lh})|^{2+\delta} \\ & \leq \frac{p^{1+\delta}}{H_L^{\delta/2}} \sum_{k=1}^p |\lambda_k|^{2+\delta} \sum_{h=1}^{H_L} \frac{E|X_{Lhk} - A_{Lhk}|^{2+\delta}}{H_L} = o(1) \end{aligned} \quad (4.2)$$

by assumption (ii), the result follows from Theorem 4.2. ///

The following result will be used in connection with condition (ii) of Theorem 4.5 in the context of stratified sampling.

Lemma 4.6: 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 $\sum W_i = 1$

$$\sum W_i E|X_i|^r \leq (\sum W_i E|X_i|^s)^{r/s}.$$

Proof: By Holder's inequality (C.R. Rao, 1973, p. 149)

$$E|X_i|^r \leq (E|X_i|^s)^{r/s} \quad (4.3)$$

for $i = 1, \dots, n$ so that

$$\sum W_i E|X_i|^r \leq \sum W_i (E|X_i|^s)^{r/s} \quad (4.4)$$

Since $f(x) = -x^{r/s}$ is convex on $(0, \infty)$ it follows from Jensen's inequality that

$$\sum W_i (E|X_i|^s)^{r/s} \leq (\sum W_i E|X_i|^s)^{r/s} \quad (4.5)$$

The result now follows from (4.4) and (4.5). ///

2 3

OF/DE

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

4.2. Stratified Simple Random Sampling

The framework for the asymptotic theory presented here is provided by a sequence of finite populations $\{\pi_L\}$ with L strata in π_L and the asymptotic results obtained as $L \rightarrow \infty$. The case of stratified simple random sampling with replacement will be discussed in detail with extensions to more complex sampling schemes indicated in section 4.6.

Suppose $y_{Lhi} = (y_{Lhi1}, \dots, y_{Lhip})$ denotes the (p-dimensional) value of the i^{th} unit in the h^{th} stratum in the population π_L ($i = 1, \dots, N_{Lh}$; $h = 1, \dots, L$) and let $\bar{y}_{Lh} = (\bar{y}_{Lh1}, \dots, \bar{y}_{Lhp})$ denote the mean of all N_{Lh} units in the h^{th} stratum ($h = 1, \dots, L$). The population mean is then $\bar{y}_L = \sum_h W_{Lh} \bar{y}_{Lh} = (\bar{y}_{L1}, \dots, \bar{y}_{Lp})$ where $W_{Lh} = N_{Lh}/N_L$ ($h = 1, \dots, L$) and $N_L = \sum_h N_{Lh}$.

Similarly, suppose $y_{Lhi} = (y_{Lhi1}, \dots, y_{Lhip})$ denotes the (p-dimensional) value of the i^{th} unit in the sample in the h^{th} stratum in π_L ($i = 1, \dots, n_{Lh}$; $h = 1, \dots, L$) and let $\bar{y}_{Lh} = (\bar{y}_{Lh1}, \dots, \bar{y}_{Lhp})$ denote the mean of the n_{Lh} units in the sample in the h^{th} stratum ($h = 1, \dots, L$). The usual estimator of $\bar{y}_L$ is the "stratified sample mean" $\bar{y}_{L,st} = \sum_h W_{Lh} \bar{y}_{Lh} = (\bar{y}_{L1,st}, \dots, \bar{y}_{Lp,st})$.

The dispersion matrix $\Gamma_{Lh} = ((\gamma_{Lhkk}'))$ of y_{Lhi} is defined by

$$\gamma_{Lhkk'} = N_{Lh}^{-1} \sum_{i=1}^{N_{Lh}} (y_{Lhik} - \bar{y}_{Lhk})(y_{Lhik'} - \bar{y}_{Lhk'}) \quad (4.6)$$

with $\gamma_{Lhkk'}$ estimated by (provided $n_{Lh} \geq 2$)

$$\hat{\gamma}_{Lhkk'} = (n_{Lh} - 1)^{-1} \sum_{i=1}^{n_{Lh}} (y_{Lhik} - \bar{y}_{Lhk})(y_{Lhik'} - \bar{y}_{Lhk'}) \quad (4.7)$$

for $k, k' = 1, \dots, p$ and $h = 1, \dots, L$. The dispersion matrix of $\bar{y}_{L,st}$ is then

$$D(\bar{y}_{L,st}) = \sum_h \frac{w_{Lh}^2}{n_{Lh}} \Gamma_{Lh} \quad (4.8)$$

and is estimated by

$$\hat{D}(\bar{y}_{L,st}) = \sum_h \frac{w_{Lh}^2}{n_{Lh}} \hat{\Gamma}_{Lh} \quad (4.9)$$

where $\hat{\Gamma}_{Lh} = ((\hat{\gamma}_{Lhkk'}))$.

Regularity Conditions

The regularity conditions involved in the asymptotic theory are now stated. In order to avoid unnecessary repetition all limits are understood to be taken as $L \rightarrow \infty$ in what follows. Letting

$n_L = \sum_{h=1}^L n_{Lh}$ denote the total sample size, consider the following conditions:

$$(C1) \quad \sum_{h=1}^L W_{Lh} E |y_{Lhik} - \bar{Y}_{Lhk}|^{2+\delta} = O(1)$$

for some $\delta > 0$ and $k = 1, \dots, p$.

$$(C1') \quad \sum_{h=1}^L W_{Lh} E |y_{Lhik} - \bar{Y}_{Lhk}|^{1+\delta} = O(1)$$

for some $\delta > 0$ and $k = 1, \dots, p$.

$$(C2) \quad \frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{N_{Lh}}{n_{Lh}} = O(1).$$

$$(C2') \quad \max_{1 \leq h \leq L} W_{Lh} = O(L^{-1}).$$

$$(C3) \quad n_L \sum_{h=1}^L \frac{W_{Lh}^2}{n_{Lh}} \Gamma_{Lh} + \bar{\Gamma} = ((\gamma_{kk}'))$$

where $\bar{\Gamma}$ is positive definite.

Condition C1 states that the (weighted) average of the within stratum $2+\delta$ absolute moments is bounded uniformly in L .

A sufficient condition for C1 to hold is that the sequence of finite populations $\{\pi_L\}$ be uniformly bounded (i.e.,

$$\max_{\substack{1 \leq i \leq N_h \\ 1 \leq h \leq L}} |Y_{Lhik}| = O(1) \text{ for } k = 1, \dots, p). \text{ By Lemma 4.6 condition}$$

C1 implies C1'.

While condition C2 is difficult to interpret in general, its implications are more readily understood in three important special cases. Under bounded allocation ($\max_{1 \leq h \leq L} n_{Lh} = O(1)$) C2 reduces to C2'. Under proportional allocation ($n_{Lh}/n_L = w_{Lh}$) C2 is automatically satisfied since

$$\frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{N_{Lh}}{n_{Lh}} \equiv 1. \quad (4.10)$$

When $p = 1$ and optimal allocation is employed ($n_{Lh}/n_L = (w_{Lh} \gamma_{Lh11}^{1/2}) / (\sum_h w_{Lh} \gamma_{Lh11}^{1/2})$) C2 reduces to

$$\left(\sum_{h=1}^L w_{Lh} \gamma_{Lh11}^{1/2} \right) \max_{1 \leq h \leq L} \gamma_{Lh11}^{-1/2} = O(1). \quad (4.11)$$

Since C3 reduces to $\sum_h w_{Lh} \gamma_{Lh11}^{1/2} + \gamma_{11}^{1/2} > 0$ (see (4.16) below)

(4.11) will hold when $\{ \min_{1 \leq h \leq L} \gamma_{Lh11} \}$ is bounded away from zero.

Specific examples of cases in which C2 and C2' hold (or do not hold) are now given. While C2 implies C2' in the case of bounded allocation, neither condition implies the other in general (Examples 4.2 and 4.3).

Example 4.1. (C2' holds): Taking $N_{Lh} \leq M$ for $h = 1, \dots, L$ and $L = 1, 2, \dots$ yields $\max_{1 \leq h \leq L} w_{Lh} \leq M/L = O(L^{-1})$. Taking $N_{Lh} = h$

yields $W_{Lh} = L^{-1}(L+1)^{-1}2h$ so that $\max_{1 \leq h \leq L} W_{Lh} = 2/(L+1) = O(L^{-1})$.

Example 4.2 ($C2'$ holds but $C2$ does not): When $N_{Lh} = M$ $C2'$ holds by Example 4.1. Taking $n_{Lh} = h$ leads to

$$\frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{N_{Lh}}{n_{Lh}} = \frac{(L+1)}{2} = O(L). \quad (4.12)$$

Example 4.3 ($C2$ holds but $C2'$ does not): For $L = 2, 3, \dots$, taking $N_{L1} = (L-1)M$ and $N_{Lh} = M$ for $h > 1$ yields $W_{L1} = 1/2$ and $W_{Lh} = 1/(2(L-1))$ so that $\max_{1 \leq h \leq L} W_{Lh} = 1/2$. Taking $n_{L1} = 1$ and $n_{Lh} = L$ for $h > 1$ leads to

$$\frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{N_{Lh}}{n_{Lh}} = \frac{(2L-1)}{(2L-2)} = O(1). \quad (4.13)$$

Condition $C3$ requires that the dispersion matrix of $\bar{y}_{L,st}$ (multiplied by the normalizing factor n_L) tend to a limit as $L \rightarrow \infty$. Similar assumptions in situations involving independent nonidentically distributed random variables have been made by Hoadley (1971) and Philippou (1973). (Application of the results in section 4.1 will exploit the fact that y_{Lhi} and $y_{Lh'i}$ are independent but not necessarily identically distributed random variables.) Condition $C3$ may not be unreasonable under bounded allocation since (for $p=1$ and $\gamma_{Lh11} = 1$)

$$n_L \sum_{h=1}^L \frac{W_{Lh}^2}{n_{Lh}} = O(L) \cdot O(L^{-1}) \quad (4.14)$$

as C2' implies $\sum_{h=1}^L W_{Lh}^2 = O(L^{-1})$. Under proportional allocation C3 reduces to

$$\sum_{h=1}^L W_{Lh} \Gamma_{Lh} \rightarrow \bar{\Gamma} \quad (\text{pos. def.}). \quad (4.15)$$

When $p = 1$ and optimal allocation is used C3 reduces to

$$\sum_{h=1}^L W_{Lh} \gamma_{Lh11} \rightarrow \gamma_{11} > 0. \quad (4.16)$$

Consistency and Asymptotic Normality of $\bar{y}_{L,st}$

The consistency and asymptotic normality of $\bar{y}_{L,st}$ are now established in Theorems 4.7 and 4.8 respectively.

Theorem 4.7: If C1' and C2 hold then $\bar{y}_{L,st} - \bar{y}_L \xrightarrow{P} 0$.

Proof: To apply Theorem 4.1 define for $k = 1, \dots, p$

$$X_{Lbik} = n_L \frac{W_{Lh}}{n_{Lh}} (y_{Lhik} - \bar{y}_{Lhk}). \quad (4.17)$$

Then under conditions C1' and C2

$$\begin{aligned} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} \frac{E|X_{Lhik}|^{1+\delta}}{n_L} &= n_L^\delta \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} \left(\frac{W_{Lh}}{n_{Lh}} \right)^{1+\delta} E|y_{Lhik} - \bar{y}_{Lhk}|^{1+\delta} \\ &\leq \left(\frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{n_{Lh}}{n_L} \right)^\delta \sum_{h=1}^L W_{Lh} E|y_{Lhik} - \bar{y}_{Lhk}|^{1+\delta} \\ &= O(1) \end{aligned} \quad (4.18)$$

so that

$$\frac{1}{n_L} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} x_{Lhik} = \bar{y}_{Lk, st} - \bar{y}_{Lk} \xrightarrow{P} 0 \quad (4.19)$$

by Theorem 4.1 (with $H_L = n_L$) . ///

Theorem 4.8: Under $C1 \wedge C3$, $\sqrt{n_L} (\bar{y}_{L, st} - \bar{y}_L) \xrightarrow{L} N(0, \bar{\Gamma})$.

Proof: To apply Theorem 4.5, let $x_{Lhi} = (x_{Lhi1}, \dots, x_{Lhip})$

where x_{Lhik} is defined in (4.17). Under $C3$

$$\frac{1}{n_L} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} D(x_{Lhi}) = n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \Gamma_{Lh} \rightarrow \bar{\Gamma} \quad (4.20)$$

so that condition (i) of Theorem 4.5 is satisfied. Under $C1$ and $C2$

$$\begin{aligned} \frac{1}{n_L} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} |x_{Lhik}|^{2+\delta} &= n_L^{1+\delta} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} \left(\frac{w_{Lh}}{n_{Lh}} \right)^{2+\delta} E |y_{Lhik} - \bar{y}_{Lhk}|^{2+\delta} \\ &\leq \left(\frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{N_{Lh}}{n_{Lh}} \right)^{1+\delta} \sum_{h=1}^L w_{Lh} E |y_{Lhik} - \bar{y}_{Lhk}|^{2+\delta} \\ &= O(1) \end{aligned} \quad (4.21)$$

so that condition (ii) of Theorem 4.5 is satisfied. Thus

$$\frac{\sqrt{n_L}}{n_L} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} x_{Lhi} = \sqrt{n_L} (\bar{y}_{L, st} - \bar{y}_L) \xrightarrow{L} N(0, \bar{\Gamma}) \quad (4.22)$$

by Theorem 4.5. ///

Consistency of $\hat{D}(\bar{y}_{L,st})$

Theorem 4.9 establishes the consistency of $\hat{D}(\bar{y}_{L,st})$, suitably normalized.

Theorem 4.9: Under C1 and C2, $n_L \{ \hat{D}(\bar{y}_{L,st}) - D(\bar{y}_{L,st}) \} \xrightarrow{P} 0$.

Proof: The result is first shown for the variance terms. Note that for $k = 1, \dots, p$

$$\begin{aligned} n_L \text{var}(\bar{y}_{Lk,st}) &= n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhik} - \bar{y}_{Lhk})^2}{(n_{Lh} - 1)} \\ &= n_L \sum_{h=1}^L w_{Lh}^2 \frac{(\bar{y}_{Lhk} - \bar{y}_{Lhk})^2}{(n_{Lh} - 1)}. \end{aligned} \quad (4.23)$$

To apply Theorem 4.1 let

$$x_{Lhik} = n_L \frac{w_{Lh} (y_{Lhik} - \bar{y}_{Lhk})}{[n_{Lh} (n_{Lh} - 1)]^{1/2}} \quad (4.24)$$

for $k = 1, \dots, p$. Then under C1 and C2 (noting that $n_{Lh} \geq 2$ implies $(n_{Lh} - 1)^{-1} \leq 2/n_{Lh}$)

$$\begin{aligned} \frac{1}{n_L} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} E |x_{Lhik}^2|^{1+\delta/2} &= n_L^{1+\delta} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} \frac{w_{Lh}^{2+\delta} E |y_{Lhik} - \bar{y}_{Lhk}|^{2+\delta}}{[n_{Lh} (n_{Lh} - 1)]^{1+\delta/2}} \\ &\leq 2^{1+\delta/2} \left(\frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{N_{Lh}}{n_{Lh}} \right)^{1+\delta} \sum_{h=1}^L w_{Lh} E |y_{Lhik} - \bar{y}_{Lhk}|^{2+\delta} \\ &= O(1) \end{aligned} \quad (4.25)$$

so that by Theorem 4.1

$$\begin{aligned} & \frac{1}{n_L} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} \{x_{Lhi}^2 - E(x_{Lhi}^2)\} \\ &= n_L \left\{ \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} - \sum_{h=1}^L \frac{w_{Lh}^2}{(n_{Lh} - 1)} \right\} \\ &\xrightarrow{P} 0. \end{aligned} \quad (4.26)$$

Similarly let

$$x_{Lhk} = (Ln_L)^{\frac{1}{2}} w_{Lh} \frac{(\bar{y}_{Lhk} - \bar{y}_{Lh})}{(n_{Lh} - 1)^{\frac{1}{2}}} \quad (4.27)$$

for $k = 1, \dots, p$ and for convenience take $0 < \delta' \leq \min \left\{ \frac{\delta}{2}, \frac{1}{2} \right\}$.

Then under C2 (noting that $L < n_L$ and $(n_{Lh} - 1)^{-1} \leq 2/n_{Lh}$)

$$\begin{aligned} \frac{1}{L} \sum_{h=1}^L E |x_{Lhk}^2|^{1+\delta'/2} &= L^{\delta'/2} n_L^{1+\delta'/2} \sum_{h=1}^L w_{Lh}^{2+\delta'} \frac{E |\bar{y}_{Lhk} - \bar{y}_{Lh}|^{2+\delta'}}{(n_{Lh} - 1)^{1+\delta'/2}} \\ &\leq 2^{1+\delta'/2} \frac{L^{\delta'/2}}{n_L^{\delta'/2}} \left(\frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{n_{Lh}}{n_L} \right)^{1+\delta'} \\ &\quad \cdot \sum_{h=1}^L w_{Lh} n_L^{\delta'/2} E |\bar{y}_{Lhk} - \bar{y}_{Lh}|^{2+\delta'} \\ &= O(1) \end{aligned} \quad (4.28)$$

provided it can be shown that

$$\sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E|\bar{y}_{Lhk} - \bar{Y}_{Lhk}|^{2+\delta'} = O(1). \quad (4.29)$$

To establish (4.29) note that (applying first the Cauchy-Schwarz inequality and then the inequality of Lemma 4.4)

$$\begin{aligned} E|\bar{y}_{Lhk} - \bar{Y}_{Lhk}|^{2+\delta'} &\leq \{E|\bar{y}_{Lhk} - \bar{Y}_{Lhk}|^2 E|\bar{y}_{Lhk} - \bar{Y}_{Lhk}|^{2+2\delta'}\}^{1/2} \\ &\leq \left(\frac{\gamma_{Lhkk}}{n_{Lh}}\right)^{1/2} \left[\frac{n_{Lh}^{1+2\delta'}}{n_{Lh}^{2+2\delta'}} \sum_{i=1}^{n_{Lh}} E|y_{Lhik} - \bar{Y}_{Lhk}|^{2+2\delta'} \right]^{1/2} \\ &= \left(\frac{\gamma_{Lhkk}}{n_{Lh}}\right)^{1/2} [E|y_{Lhik} - \bar{Y}_{Lhk}|^{2+2\delta'}]^{1/2}. \quad (4.30) \end{aligned}$$

Substituting (4.30) in (4.29) yields (using first the fact that $\delta'/2 \leq 1/2$ and then the Cauchy inequality)

$$\begin{aligned} &\sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E|\bar{y}_{Lhk} - \bar{Y}_{Lhk}|^{2+\delta'} \\ &\leq \sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} \frac{\gamma_{Lhkk}^{1/2}}{n_{Lh}^{1/2}} \{E|y_{Lhik} - \bar{Y}_{Lhk}|^{2+2\delta'}\}^{1/2} \\ &\leq \sum_{h=1}^L w_{Lh} \gamma_{Lhkk}^{1/2} \{E|y_{Lhik} - \bar{Y}_{Lhk}|^{2+2\delta'}\}^{1/2} \\ &\leq \left\{ \left(\sum_{h=1}^L w_{Lh} E|y_{Lhik} - \bar{Y}_{Lhk}|^2 \right) \left(\sum_{h=1}^L w_{Lh} E|y_{Lhik} - \bar{Y}_{Lhk}|^{2+2\delta'} \right) \right\}^{1/2} \\ &= O(1) \quad (4.31) \end{aligned}$$

under C1 using Lemma 4.6 (note that $2 < 2 + 2\delta' \leq 2 + \delta$) so that (4.29) holds. It now follows from (4.28) (invoking Theorem 4.1) that

$$\begin{aligned} & \frac{1}{L} \sum_{h=1}^L \{X_{Lhk}^2 - E(X_{Lhk}^2)\} \\ &= n_L \left\{ \sum_{h=1}^L w_{Lh}^2 \frac{(\bar{y}_{Lhk} - \bar{y}_{Lhk})^2}{(n_{Lh} - 1)} - \sum_{h=1}^L \frac{w_{Lh}^2 \gamma_{Lhkk}}{n_{Lh}(n_{Lh}-1)} \right\} \\ &\xrightarrow{P} 0. \end{aligned} \quad (4.32)$$

Combining (4.23), (4.26) and (4.32) yields

$$n_L \{ \text{var}(\bar{y}_{Lk,st}) - \text{Var}(\bar{y}_{Lk,st}) \} \xrightarrow{P} 0. \quad (4.33)$$

To obtain the corresponding result for the covariance terms write (for $k \neq k' = 1, \dots, p$)

$$\begin{aligned} & n_L \text{cov}(\bar{y}_{Lk,st}, \bar{y}_{Lk',st}) \\ &= n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhik} - \bar{y}_{Lhk})(y_{Lhik'} - \bar{y}_{Lhk'})}{(n_{Lh} - 1)} \\ &= n_L \sum_{h=1}^L w_{Lh}^2 \frac{(\bar{y}_{Lhk} - \bar{y}_{Lhk})(\bar{y}_{Lhk'} - \bar{y}_{Lhk'})}{(n_{Lh} - 1)}. \end{aligned} \quad (4.34)$$

Letting $X_{Lhikk'} = X_{Lhik} X_{Lhik'}$ (where X_{Lhik} is defined in (4.24)) and using the Cauchy-Schwarz inequality,

$$E|X_{Lhkk'}|^{1+\delta/2} \leq \frac{n_L^{2+\delta} w_{Lh}^{2+\delta}}{[n_{Lh}(n_{Lh}-1)]^{1+\delta/2}} \{E|y_{Lhik} - \bar{y}_{Lhk}|^{2+\delta} E|y_{Lhik'} - \bar{y}_{Lhk'}|^{2+\delta}\}^{\frac{1}{2}}, \quad (4.35)$$

so that it follows as in (4.25)-(4.26) that

$$n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhik} - \bar{y}_{Lhk})(y_{Lhik'} - \bar{y}_{Lhk'})}{(n_{Lh} - 1)} - n_L \sum_{h=1}^L \frac{w_{Lh}^2 \gamma_{Lhkk'}}{(n_{Lh} - 1)} \xrightarrow{P} 0 \quad (4.36)$$

under C1 and C2. Now let $X_{Lhkk'} = X_{Lhk} X_{Lhk'}$ (where X_{Lhk} is defined in (4.27)) and choose $0 < \delta' \leq \min\{\frac{\delta}{2}, \frac{1}{2}\}$. Since, by the Cauchy-Schwarz inequality,

$$E|X_{Lhkk'}|^{1+\delta'/2} \leq \frac{L^{1+\delta'/2} n_L^{1+\delta'/2} w_{Lh}^{2+\delta'}}{(n_{Lh} - 1)^{1+\delta'/2}} \{E|\bar{y}_{Lhk} - \bar{y}_{Lhk}|^{2+\delta'} E|\bar{y}_{Lhk'} - \bar{y}_{Lhk'}|^{2+\delta'}\} \quad (4.37)$$

it follows as in (4.28)-(4.32) that

$$n_L \sum_{h=1}^L w_{Lh}^2 (\bar{y}_{Lhk} - \bar{y}_{Lhk})(\bar{y}_{Lhk'} - \bar{y}_{Lhk'}) - n_L \sum_{h=1}^L \frac{w_{Lh}^2 \gamma_{Lhkk'}}{n_{Lh}(n_{Lh} - 1)} \xrightarrow{P} 0 \quad (4.38)$$

under C1 and C2 . Finally, combining (4.34), (4.36) and (4.38),

$$n_L \{ \text{cov}(\bar{y}_{Lk,st}, \bar{y}_{Lk',st}) - \text{Cov}(\bar{y}_{Lk,st}, \bar{y}_{Lk',st}) \}$$

$$\xrightarrow{p} 0 . \quad ///$$

(4.39)

4. Linearization Methods

Many nonlinear population parameters θ_L may be expressed as (real-valued) functions of the population means $\bar{Y}_{L1}, \dots, \bar{Y}_{Lp}$ (say $\theta_L = g(\bar{Y}_L)$) and may be estimated by $\hat{\theta}_L = g(\bar{y}_{L,st})$. (Explicit examples involving ratios and correlation coefficients are provided at the end of this section.) The asymptotic normality of such statistics $\hat{\theta}_L$ and the consistency of the corresponding "linearization" variance estimator (Tepping, 1968) are shown in Theorems 4.10 and 4.11.

Two additional regularity conditions are required:

(C4) $\bar{Y}_{Lk} \rightarrow \mu_k$ for $k = 1, \dots, p$ and

(C5) $g : R^p \rightarrow R$ has continuous first partial derivatives

$g_k(t_1, \dots, t_p) = \partial g(t_1, \dots, t_p) / \partial t_k$ ($k = 1, \dots, p$) in a neighborhood of $(\mu_1, \dots, \mu_p)$.

Condition C4 is required since the variance of $\hat{\theta}_L$ involves $\bar{Y}_{L1}, \dots, \bar{Y}_{Lp}$. (Recall the remarks made concerning C3.) Condition C5 is standard and is satisfied in nearly all cases of interest (Miller, 1964, p.1600).

The notation $g_k = g_k(\mu_1, \dots, \mu_p)$ ($k = 1, \dots, p$) will be useful in what follows.

Theorem 4.10: Under C1 - C5

$$\sqrt{n_L}(\hat{\theta}_L - \theta_L) \xrightarrow{L} N(0, \sum_{k=1}^p \sum_{k'=1}^p g_k g_{k'} \gamma_{kk'}).$$

Proof: Let $I_k = (\mu_k - \Delta, \mu_k + \Delta)$, $\Delta > 0$, be any interval about μ_k in which $g_k(\cdot)$ is continuous ($k = 1, \dots, p$). Under C4 it follows from Theorem 4.7 that

$$P(\bar{y}_{Lk, st}, \bar{y}_{Lk} \in I_k, k = 1, \dots, p) \rightarrow 1. \quad (4.40)$$

Now for two sequences of events $\{A_L\}$ and $\{B_L\}$ in which $\lim P(B_L) = 1$, $\lim P(A_L) = \lim P(A_L B_L)$ so that it may be tacitly assumed in what follows that the events described in (4.40) hold.

For $\bar{y}_{Lk, st}, \bar{y}_{Lk} \in I_k$ ($k = 1, \dots, p$)

$$g(\bar{y}_{L, st}) = g(\bar{y}_L) + \sum_{k=1}^p (\bar{y}_{Lk, st} - \bar{y}_{Lk}) g_k(\xi_L) \quad (4.41)$$

where ξ_L is on the line segment joining $\bar{y}_{L, st}$ and $\bar{y}_L$.

Rewriting (4.41) as

$$\sqrt{n_L}(\hat{\theta}_L - \theta_L) = \sqrt{n_L} \sum_{k=1}^p (\bar{y}_{Lk, st} - \bar{y}_{Lk}) g_k(\xi_L) \quad (4.42)$$

and noting that under C5 $g_k(\xi_L) \xrightarrow{P} g_k$ (with ξ_L defined arbitrarily when the events in (4.40) do not hold) the result now follows from Theorem 4.8 and Slutsky's theorem. ///

Theorem 4.11: Under C1 - C5

$$n_L v_L \xrightarrow{P} \sum_{k=1}^P \sum_{k'=1}^P g_k g_{k'} \gamma_{kk'}$$

where

$$v_L = \sum_{k=1}^P \sum_{k'=1}^P g_k(\bar{y}_{L,st}) g_{k'}(\bar{y}_{L,st}) \text{cov}(\bar{y}_{Lk,st}, \bar{y}_{Lk',st})$$

is the "linearization" variance estimator.

Proof: Under C3, $n_L \text{cov}(\bar{y}_{Lk,st}, \bar{y}_{Lk',st}) \xrightarrow{P} \gamma_{kk'}$ by Theorem 4.9.

By continuity (condition C5) and Theorem 4.7, $g_k(\bar{y}_{L,st}) \xrightarrow{P} g_k$ under C4. ///

Applications of Theorems 4.10 and 4.11 are now discussed.

Example 4.4 (Ratio estimation): The combined ratio estimator

$r_L = \bar{y}_{L1,st} / \bar{y}_{L2,st}$ is often used as an estimator of $R_L = \bar{Y}_{L1} / \bar{Y}_{L2}$.

In this case the linearization variance estimator is

$$v_L = \bar{y}_{L2,st}^{-2} \{ \text{var}(\bar{y}_{L1,st}) + r_L^2 \text{var}(\bar{y}_{L2,st}) - 2r_L \text{cov}(\bar{y}_{L1,st}, \bar{y}_{L2,st}) \}.$$

Under C1 - C4 (with C5 satisfied whenever $\lim \bar{Y}_{L2} = \mu_2 \neq 0$)

$\sqrt{n_L}(r_L - R_L) / \sqrt{n_L v_L} = (r_L - R_L) / \sqrt{v_L} \xrightarrow{L} N(0,1)$ by Theorems 4.10 and 4.11.

Example 4.5 (Correlation coefficients): By defining the new

variables $Y_{Lh13} = Y_{Lh1}^2$, $Y_{Lh14} = Y_{Lh2}^2$ and $Y_{Lh15} = Y_{Lh1} Y_{Lh2}$,

the population correlation coefficient for variables 1 and 2 may be

expressed as (with the summations extending over $i = 1, \dots, N_{Lh}$

and $h = 1, \dots, L$)

$$\begin{aligned} \rho_L &= \frac{\sum Y_{Lh11} Y_{Lh12} - N_L^{-1} (\sum Y_{Lh11}) (\sum Y_{Lh12})}{\{[\sum Y_{Lh11}^2 - N_L^{-1} (\sum Y_{Lh11})^2][\sum Y_{Lh12}^2 - N_L^{-1} (\sum Y_{Lh12})^2]\}^{1/2}} \\ &= \frac{\bar{Y}_{L5} - \bar{Y}_{L1} \bar{Y}_{L2}}{\{(\bar{Y}_{L3} - \bar{Y}_{L1}^2)(\bar{Y}_{L4} - \bar{Y}_{L2}^2)\}^{1/2}} \\ \hat{\rho}_L &= g(\bar{Y}_{L1}, \dots, \bar{Y}_{L5}) \end{aligned} \quad (4.43)$$

and estimated by $\hat{\rho}_L = g(\bar{y}_{L1, st}, \dots, \bar{y}_{L5, st})$ as proposed by Frankel (1971, p.14). The partial derivatives required for the linearization variance estimator v_L in this case have been given by Frankel (1971, p.24). Under C1-C4 (with C5 clearly satisfied in this case) $(\hat{\rho}_L - \rho_L)/\sqrt{v_L} \xrightarrow{L} N(0,1)$ by Theorems 4.10 and 4.11.

A central limit theorem for regression coefficients may also be obtained by defining new variables corresponding to squares and cross products of the original variables as in Example 4.5. Frankel (1971, p.24) indicates how the partial derivatives required for the linearization variance estimator are obtained in this case.

The partial derivatives required for the linearization variance estimators for multiple and partial correlation coefficients are somewhat intractable (Frankel, 1971, p.55) although the numerical method given by Woodruff & Causey (1976) provides useful approximations to the required partial derivatives in such cases.

4.4. Jackknife Methods

While $\bar{y}_{L,st}$ is an unbiased estimator of $\bar{Y}_L$, $\hat{\theta}_L = g(\bar{y}_{L,st})$ is not in general an unbiased estimator of $\theta_L = g(\bar{Y}_L)$ due to the nonlinearity of the transformation. The jackknife statistic may then be considered as an estimator of θ_L on the basis of bias reduction (Jones, 1974). Perhaps more importantly (with Kish, Namboodiri & Pillai [1962], McCarthy [1969a], Frankel [1971] and Bean [1975] all discounting bias as being negligible for the most part in large scale surveys) the jackknife method also provides a conceptually simple variance estimator which does not require the evaluation of partial derivatives as in the case of the linearization variance estimator.

The extension of the jackknife method to stratified samples is not immediate and several different versions have been proposed. For $i = 1, \dots, n_{Lh}$ and $h = 1, \dots, L$ let $\bar{y}_{L,st}^{hi}$ denote the stratified sample mean computed from the sample after deleting the i th observation in the h th stratum and let $\hat{\theta}_L^{hi} = g(\bar{y}_{L,st}^{hi})$ and $\hat{\theta}_L^h = n_{Lh}^{-1} \sum_{i=1}^{n_{Lh}} \hat{\theta}_L^{hi}$. Then Jones (1974) has shown that the jackknife estimator

$$J_1(\hat{\theta}_L) = [1 + \sum_{h=1}^L (n_{Lh} - 1)] \hat{\theta}_L - \sum_{h=1}^L (n_{Lh} - 1) \hat{\theta}_L^h \quad (4.44)$$

reduces bias to second order moments in large samples. Jones' jackknife variance estimator is

$$v_{LJ}^{(1)} = \sum_{h=1}^L \frac{(n_{Lh} - 1)}{n_{Lh}} \sum_{i=1}^{n_{Lh}} (\hat{\theta}_L^{hi} - \hat{\theta}_L^h)^2. \quad (4.45)$$

McCarthy (1966) has also proposed a jackknife method for stratified samples when $n_{Lh} = 2$ ($h = 1, \dots, L$). This jackknife method may be extended in a natural way to the case of arbitrary $\{n_{Lh}\}$. With "pseudovalue" defined by

$$J_{Lhi} = n_{Lh} \hat{\theta}_L - (n_{Lh} - 1) \hat{\theta}_L^{hi} \quad (4.46)$$

for $i = 1, \dots, n_{Lh}$ and $h = 1, \dots, L$, McCarthy's jackknife estimator of θ_L is

$$\begin{aligned} J_2(\hat{\theta}_L) &= n_L^{-1} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} J_{Lhi} \\ &= \hat{\theta}_L \sum_{h=1}^L \frac{n_{Lh}^2}{n_L} - \sum_{h=1}^L \frac{n_{Lh}(n_{Lh} - 1)}{n_L} \hat{\theta}_L^h \end{aligned} \quad (4.47)$$

with the corresponding jackknife variance estimator

$$v_{LJ}^{(2)} = \sum_{h=1}^L \frac{1}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(J_{Lhi} - J_2(\hat{\theta}_L))^2}{(n_{Lh} - 1)}. \quad (4.48)$$

It may be noted that replacing $J_2(\hat{\theta}_L)$ by $J_{Lh} = n_{Lh}^{-1} \sum_{i=1}^{n_{Lh}} J_{Lhi}$ in (4.48) yields $v_{LJ}^{(1)}$ since $|J_{Lhi} - J_{Lh}| = (n_{Lh} - 1) |\hat{\theta}_L^{hi} - \hat{\theta}_L^h|$.

Frankel (1971, p.34) has also proposed several jackknife variance estimators for stratified samples when $n_{Lh} = 2$ ($h = 1, \dots, L$). His estimators may also be extended in a natural way to the case of arbitrary $\{n_{Lh}\}$. Frankel's JRR-D estimator is then identical to

$v_{LJ}^{(1)}$. For arbitrary $\{n_{Lh}\}$ Frankel's JRR-S estimator (also proposed by Lee [1973a] when all $n_{Lh} = 2$) is

$$v_{LJ}^{(3)} = \sum_{h=1}^L \frac{(n_{Lh} - 1)}{n_{Lh}} \sum_{i=1}^{n_{Lh}} (\hat{\theta}_L^{hi} - \hat{\theta}_L)^2 \quad (4.49)$$

while his JRR-H estimator is

$$v_{LJ}^{(4)} = \sum_{h=1}^L \frac{(n_{Lh} - 1)}{m_{Lh}} \sum_{i=1}^{m_{Lh}} (\hat{\theta}_L^{hi} - \hat{\theta}_L)^2 \quad (4.50)$$

where the $\hat{\theta}_L^{hi}$ ($i = 1, \dots, m_{Lh}$) are based on a random subsample of size m_{Lh} from the n_{Lh} units in stratum $h = 1, \dots, L$. The motivation for introducing $v_{LJ}^{(4)}$ in addition to $v_{LJ}^{(3)}$ is that fewer values of $\hat{\theta}_L^{hi}$ need be computed with $v_{LJ}^{(4)}$.

Remark 4.1: Frankel actually computes $\hat{\theta}_L^{hi}$ (when $n_{Lh} = 2$ for $h = 1, \dots, L$) by deleting the i th unit in the h th stratum and including the other unit twice. Since $\hat{\theta}_L^{hi}$ is assumed here to be of the form $\hat{\theta}_L^{hi} = g(y_{L,st}^{hi})$ the two methods are equivalent since $y_{L,st}^{hi}$ has the same value even if the other unit in the h th stratum is included twice.

One further jackknife variance estimator which is suggested by (4.45) and (4.49) is

$$v_{LJ}^{(5)} = \sum_{h=1}^L \frac{(n_{Lh} - 1)}{n_{Lh}} \sum_{i=1}^{n_{Lh}} (\hat{\theta}_L^{hi} - \bar{\theta}_L)^2 \quad (4.51)$$

where $\bar{\hat{\theta}}_L = n_L^{-1} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} \hat{\theta}_L^{hi}$. Under equal allocation ($n_{Lh} = p$ for $h = 1, \dots, L$) $v_{LJ}^{(5)}$ is the same as $v_{LJ}^{(2)}$ since
 $|J_{Lhi} - J_2(\hat{\theta}_L)| = (p-1) |\hat{\theta}_L^{hi} - \bar{\hat{\theta}}_L|$.

Remark 4.2: In the linear (univariate) case $g(\bar{y}_{L,st}) = \bar{y}_{L,st}$ both $J_1(\hat{\theta}_L)$ and $J_2(\hat{\theta}_L)$ reduce to $\bar{y}_{L,st}$. In this case all the jackknife variance estimators (with the exception of $v_{LJ}^{(4)}$ when some $n_{Lh} > 2$) reduce to $\text{var}(\bar{y}_{L,st})$.

The asymptotic normality of the jackknife statistics $J_1(\hat{\theta}_L)$ and $J_2(\hat{\theta}_L)$ is shown in Theorem 4.12 and the consistency of the jackknife variance estimators $v_{LJ}^{(i)}$ ($i = 1, \dots, 5$) is established in Theorems 4.13 - 4.16 below.

For the asymptotic normality of $J_1(\hat{\theta}_L)$ and $J_2(\hat{\theta}_L)$ a slightly stronger condition on g is required than for the asymptotic normality of $\hat{\theta}_L$ (see C5):

(C5') $g : R^p \rightarrow R$ has bounded second partial derivatives

$$g_{kk'}(t_1, \dots, t_p) = \partial^2 g(t_1, \dots, t_p) / \partial t_k \partial t_{k'}, \quad (k, k' = 1, \dots, p)$$

in a neighborhood of $(\mu_1, \dots, \mu_p)$.

This condition is standard in the case of the jackknife and is again satisfied in nearly all cases of interest (Miller, 1964, p.1600).

Theorem 4.12: Under C1-C4 and C5' for $i = 1, 2$

$$\sqrt{n_L} (J_1(\hat{\theta}_L) - \theta_L) \xrightarrow{L} N(0, \sum_{k=1}^P \sum_{k'=1}^P g_k g_{k'} \gamma_{kk'}) .$$

Proof: For simplicity only the case $p = 1$ is proved in detail; the extension to the case $p > 1$ is relatively straightforward.

Consider the estimator $J_1(\hat{\theta}_L)$ first. Suppressing the subscript k , let $I = (\mu - 2\Delta, \mu + 2\Delta)$, $\Delta > 0$, be any neighborhood of μ in which $g''(\cdot)$ is bounded. Under C4 it follows from Theorem 4.7 that

$$P(\bar{y}_{L,st} \in (\mu - \Delta, \mu + \Delta)) \rightarrow 1 . \quad (4.52)$$

Letting $\bar{y}_{Lh}^i$ denote the mean in the h^{th} stratum omitting the i^{th} unit

$$\begin{aligned} |\bar{y}_{L,st}^{hi} - \bar{y}_{L,st}| &= |w_{Lh}(\bar{y}_{Lh}^i - \bar{y}_{Lh})| \\ &= \left| w_{Lh} \frac{(y_{Lhi} - \bar{y}_{Lh})}{(n_{Lh} - 1)} \right| \\ &\leq \frac{w_{Lh}}{(n_{Lh} - 1)} \{ |y_{Lhi} - \bar{y}_{Lh}| + |\bar{y}_{Lh} - \bar{y}_{Lh}| \} \quad (4.53) \end{aligned}$$

so that

$$\begin{aligned} P(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} |\bar{y}_{L,st}^{hi} - \bar{y}_{L,st}| < \Delta) \\ \geq P(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} \frac{w_{Lh}}{(n_{Lh} - 1)} \{ |y_{Lhi} - \bar{y}_{Lh}| + |\bar{y}_{Lh} - \bar{y}_{Lh}| \} < \Delta) \end{aligned}$$

$$\geq P\left(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} \frac{w_{Lh}}{(n_{Lh}-1)} |y_{Lhi} - \bar{y}_{Lh}| < \frac{\Delta}{2}, \max_{1 \leq h \leq L} \frac{w_{Lh}}{(n_{Lh}-1)} |\bar{y}_{Lh} - \bar{\bar{y}}_{Lh}| < \frac{\Delta}{2}\right)$$

$$\geq 1 - P\left(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} \frac{w_{Lh}}{(n_{Lh}-1)} |y_{Lhi} - \bar{y}_{Lh}| \geq \frac{\Delta}{2}\right)$$

$$= P\left(\max_{1 \leq h \leq L} \frac{w_{Lh}}{(n_{Lh}-1)} |\bar{y}_{Lh} - \bar{\bar{y}}_{Lh}| \geq \frac{\Delta}{2}\right) \quad (4.54)$$

$$\rightarrow 1 \quad (4.54)$$

since under C3 (using Chebyshev's inequality and the inequality

$$(n_{Lh}-1)^{-1} \leq 2/n_{Lh})$$

$$P\left(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} \frac{w_{Lh}}{(n_{Lh}-1)} |y_{Lhi} - \bar{y}_{Lh}| \geq \frac{\Delta}{2}\right)$$

$$\leq \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} P\left(\frac{w_{Lh}}{(n_{Lh}-1)} |y_{Lhi} - \bar{y}_{Lh}| \geq \frac{\Delta}{2}\right)$$

$$\leq \frac{4}{\Delta^2} \sum_{h=1}^L \sum_{i=1}^{n_{Lh}} \frac{w_{Lh}^2}{(n_{Lh}-1)^2} \gamma_{Lh}$$

$$\leq \frac{16}{\Delta^2} \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \gamma_{Lh}$$

$$\rightarrow 0 \quad (4.55)$$

and since (by a similar argument)

$$\begin{aligned}
 & P\left(\max_{1 \leq h \leq L} \frac{W_{Lh}}{(n_{Lh}-1)}, |\bar{y}_{Lh} - \bar{Y}_{Lh}| \geq \frac{\Delta}{2}\right) \\
 & \leq \frac{8}{\Delta^2} \sum_{h=1}^L \frac{W_{Lh}^2}{n_{Lh}} \gamma_{Lh} \\
 & \rightarrow 0. \quad \triangleright \quad (4.56)
 \end{aligned}$$

Now since $|\bar{y}_{L,st}^{hi} - \mu| \leq |\bar{y}_{L,st}^{hi} - \bar{y}_{L,st}| + |\bar{y}_{L,st} - \mu|$,

$$\begin{aligned}
 & P\left(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} |\bar{y}_{L,st}^{hi} - \mu| < 2\Delta\right) \\
 & \geq P\left(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} |\bar{y}_{L,st}^{hi} - \bar{y}_{L,st}| + |\bar{y}_{L,st} - \mu| < 2\Delta\right) \\
 & \geq P\left(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} |\bar{y}_{L,st}^{hi} - \bar{y}_{L,st}| < \Delta, |\bar{y}_{L,st} - \mu| < \Delta\right) \\
 & \rightarrow 1 \quad (4.57)
 \end{aligned}$$

by (4.54) and (4.52). Combining (4.52) and (4.57),

$$P(\bar{y}_{L,st}, \bar{y}_{L,st}^{hi} \in I; i = 1, \dots, n_{Lh}; h = 1, \dots, L) \rightarrow 1. \quad (4.58)$$

By the argument following (4.40), it may be tacitly assumed in what follows that the events described in (4.58) hold.

For $\bar{y}_{L,st}^{hi} \in I$ ($i = 1, \dots, n_{Lh}$; $h = 1, \dots, L$)

$$g(\bar{y}_{L,st}^{hi}) = g(\bar{y}_{L,st}) + (\bar{y}_{L,st}^{hi} - \bar{y}_{L,st}) g'(\bar{y}_{L,st}) + \frac{1}{2} (\bar{y}_{L,st}^{hi} - \bar{y}_{L,st})^2 g''(\xi_L^{hi}) \quad (4.59)$$

where ξ_L^{hi} lies between $\bar{y}_{L,st}^{hi}$ and $\bar{y}_{L,st}$. Substituting (4.59)

in (4.44) and noting that $\frac{1}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \bar{y}_{L,st}^{hi} = \bar{y}_{L,st}$ for $h = 1, \dots, L$ leads to

$$\sqrt{n_L} (J_1(\hat{\theta}_L) - \theta_L) = \sqrt{n_L} (\hat{\theta}_L - \theta_L) - \frac{1}{2} \frac{n_L}{\sqrt{n_L}} \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} g''(\xi_L^{hi}). \quad (4.60)$$

Since the first term on the right side of (4.60) is asymptotically normally distributed with mean zero and variance $|g'(\mu)|^2 \gamma$ by Theorem 4.9, the result will follow from Slutsky's theorem if it can be shown that the second term on the right $\xrightarrow{P} 0$.

By Theorem 4.9

$$\frac{1}{n_L} \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} \xrightarrow{P} \gamma \quad (4.61)$$

Then since $|g''(\xi_r^{hi})| < M < \infty$ ($i = 1, \dots, n_{Lh}$; $h = 1, \dots, L$) when the events in (4.58) hold, the second term on the right side of (4.60) $\xrightarrow{P} 0$ due to the extra $\sqrt{n_L}$ term in the denominator (with

ξ_L^{hi} defined arbitrarily when the events in (4.58) do not hold).

Consider next the estimator $J_2(\hat{\theta}_L)$. Substituting (4.59) in (4.47) leads to

$$\sqrt{n_L}(J_2(\hat{\theta}_L) - \theta_L) = \sqrt{n_L}(\hat{\theta}_L - \theta_L) - \frac{1}{2} \frac{n_L}{\sqrt{n_L}} \sum_{h=1}^L \frac{n_{Lh}}{n_L} \frac{w_{Lh}^2}{n_{Lh}} \sum_{h=1}^L \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} g''(\xi_L^{hi}) \quad (4.62)$$

where the second term on the right side of (4.62) again $\xrightarrow{p} 0$

since for $L = 1, 2, \dots$, $\max_{1 \leq h \leq L} (n_{Lh}/n_L) < 1$. ///

Theorem 4.13: Under C1-C5 for $i = 1, 3$

$$n_L v_{LJ}^{(1)} \xrightarrow{p} \sum_{k=1}^p \sum_{k'=1}^p g_k g_{k'} \gamma_{kk'}$$

Proof: Details of the proof again need only be given for the case $p = 1$.

Consider first the estimator $v_{LJ}^{(3)}$. As in the previous theorem

$$P(\bar{y}_{L,st}, \bar{y}_{L,st}^{hi} \in I; i = 1, \dots, n_{Lh}; h = 1, \dots, L) \rightarrow 1 \quad (4.63)$$

where $I = (\mu - \Delta, \mu + \Delta)$, $\Delta > 0$, is any neighborhood of μ in which $g'(\cdot)$ is continuous. For $\bar{y}_{L,st}, \bar{y}_{L,st}^{hi} \in I$ ($i = 1, \dots, n_{Lh}; h = 1, \dots, L$)

$$g(\bar{y}_{L,st}^{hi}) = g(\bar{y}_{L,st}) + (\bar{y}_{L,st}^{hi} - \bar{y}_{L,st}) g'(\xi_L^{hi}) \quad (4.64)$$

where ξ_L^{hi} lies between $\bar{y}_{L,st}^{hi}$ and $\bar{y}_{L,st}$. Since $g'(\cdot)$ is continuous at μ , for every $\epsilon > 0$ there exists $\Delta_\epsilon > 0$ such that $|g'(\xi_L^{hi}) - g'(\mu)| < \epsilon$ whenever $|\xi_L^{hi} - \mu| < \Delta_\epsilon$. Thus using (4.63) (with ξ_L^{hi} defined arbitrarily when the events in (4.63) do not hold)

$$\begin{aligned} & \lim P(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} |g'(\xi_L^{hi}) - g'(\mu)| < \epsilon) \\ & \geq \lim P(\max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} |\xi_L^{hi} - \mu| < \Delta_\epsilon) = 1 \end{aligned} \quad (4.65)$$

Substituting (4.64) in (4.49) leads to

$$\begin{aligned} n_{L^v}^{(3)} &= n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} (n_{Lh} - 1)^{-1} \sum_{i=1}^{n_{Lh}} \{ (y_{Lhi} - \bar{y}_{Lh}) g'(\mu) \\ & \quad + (y_{Lhi} - \bar{y}_{Lh}) (g'(\xi_L^{hi}) - g'(\mu)) \}^2 \\ &= |g'(\mu)|^2 n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} \\ & \quad + n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{\{ (y_{Lhi} - \bar{y}_{Lh}) (g'(\xi_L^{hi}) - g'(\mu)) \}^2}{(n_{Lh} - 1)} \\ & \quad + \text{cross-product term.} \end{aligned} \quad (4.66)$$

Under C3 the first term on the right side of (4.66) $\xrightarrow{p} |g'(\mu)|^2 \gamma$

by Theorem 4.9. Since

$$\begin{aligned} & n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} (g'(\xi_L^{hi}) - g'(\mu))^2 \\ & \leq \max_{\substack{1 \leq i \leq n_{Lh} \\ 1 \leq h \leq L}} |g'(\xi_L^{hi}) - g'(\mu)|^2 n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} \quad (4.67) \end{aligned}$$

the second term on the right side of (4.66) $\xrightarrow{p} 0$ by (4.65). A

similar argument shows that the cross-product term in (4.66)

also $\xrightarrow{p} 0$ so that

$$n_L v_{LJ}^{(3)} \xrightarrow{p} |g'(\mu)|^2 \gamma. \quad (4.68)$$

Consider next the estimator $v_{LJ}^{(1)}$. From (4.47) and (4.49)

$$n_L \dot{v}_{LJ}^{(1)} = n_L v_{LJ}^{(3)} - n_L \sum_{h=1}^L (n_{Lh} - 1) (\hat{\theta}_L^h - \hat{\theta}_L)^2. \quad (4.69)$$

Substituting (4.64) in the remainder term in (4.69) and using

Lemma 4.3 ($f(x) = x^2$ is convex on $(-\infty, \infty)$) leads to

$$\begin{aligned} & n_L \sum_{h=1}^L (n_{Lh} - 1) (\hat{\theta}_L^h - \hat{\theta}_L)^2 \\ & = n_L \sum_{h=1}^L \frac{w_{Lh}^2}{(n_{Lh} - 1)} \left[\sum_{i=1}^{n_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})(g'(\xi_L^{hi}) - g'(\mu))}{n_{Lh}} \right]^2 \end{aligned}$$

$$\leq n_L \sum_{h=1}^L \frac{w_{Lh}^2}{(n_{Lh}-1)} \sum_{i=1}^{n_{Lh}} \frac{[(y_{Lhi} - \bar{y}_{Lh})(g'(\xi_L^{hi}) - g'(\mu))]^2}{n_{Lh}} \xrightarrow{P} 0 \quad (4.70)$$

as in (4.67). The result now follows from (4.68) - (4.70). ///

The consistency of $v_{LJ}^{(4)}$ may be established under C1-C5 and the additional condition

$$(C6) \quad \max_{1 \leq h \leq L} \frac{n_{Lh}}{m_{Lh}} = O(1)$$

as shown in Theorem 4.14 below. (Note that C6 is satisfied under bounded allocation.)

Theorem 4.14: Under C1-C6

$$n_L v_{LJ}^{(4)} \xrightarrow{P} \sum_{k=1}^p \sum_{k'=1}^p g_k g_{k'} \gamma_{kk'}$$

Proof: Only the necessary modification to the proof of Theorem 4.13 in the case of $v_{LJ}^{(3)}$ need be indicated. Since, as in (4.66),

$$n_L v_{LJ}^{(4)} = n_L \sum_{h=1}^L \frac{w_{Lh}^2}{m_{Lh}} (n_{Lh}-1)^{-1} \sum_{i=1}^{n_{Lh}} \{ (y_{Lhi} - \bar{y}_{Lh}) g'(\mu) + (y_{Lhi} - \bar{y}_{Lh})(g'(\xi_L^{hi}) - g'(\mu)) \}^2 \quad (4.71)$$

the result will follow as in (4.66)-(4.68) if it can be shown that

$$n_L \sum_{h=1}^L \frac{W_{Lh}^2}{m_{Lh}} \sum_{i=1}^{m_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} \xrightarrow{p} \gamma. \quad (4.72)$$

The proof of (4.72) proceeds along the lines of Theorem

4.9. Letting $\bar{y}_{Lh}^* = m_{Lh}^{-1} \sum_{i=1}^{m_{Lh}} y_{Lhi}$, write

$$\begin{aligned} & n_L \sum_{h=1}^L \frac{W_{Lh}^2}{m_{Lh}} \sum_{i=1}^{m_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} \\ &= n_L \sum_{h=1}^L \frac{W_{Lh}^2}{m_{Lh}} \sum_{i=1}^{m_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} - n_L \sum_{h=1}^L W_{Lh}^2 \frac{(\bar{y}_{Lh} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} \\ &+ 2n_L \sum_{h=1}^L \frac{W_{Lh}^2}{(n_{Lh} - 1)} (\bar{y}_{Lh} - \bar{y}_{Lh}) (\bar{y}_{Lh} - \bar{y}_{Lh}^*). \end{aligned} \quad (4.73)$$

Letting

$$x_{Lhi} = (n_L m_L)^{\frac{1}{2}} \frac{W_{Lh} (y_{Lhi} - \bar{y}_{Lh})}{(m_{Lh} (n_{Lh} - 1))^{\frac{1}{2}}} \quad (4.74)$$

where $m_L = \sum_{h=1}^L m_{Lh}$ it follows as in (4.24) - (4.26) that

$$\begin{aligned} & n_L \left\{ \sum_{h=1}^L \frac{W_{Lh}^2}{m_{Lh}} \sum_{i=1}^{m_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} - \sum_{h=1}^L \frac{W_{Lh}^2 \gamma_{Lh}}{(n_{Lh} - 1)} \right\} \\ & \xrightarrow{p} 0 \end{aligned} \quad (4.75)$$

since C6 implies $n_L/m_L = O(1)$. By (4.75) and (4.32), the proof

of (4.72) will follow from (4.73) if it can be shown that

$$n_L \sum_{h=1}^L \frac{w_{Lh}^2}{(n_{Lh}-1)} (\bar{y}_{Lh} - \bar{Y}_{Lh}) (\bar{y}_{Lh} - \bar{y}_{Lh}^*) \xrightarrow{P} 0. \quad (4.76)$$

To establish (4.76) let

$$X_{Lh} = L n_L w_{Lh}^2 \frac{(\bar{y}_{Lh} - \bar{Y}_{Lh})(\bar{y}_{Lh} - \bar{y}_{Lh}^*)}{(n_{Lh} - 1)} \quad (4.77)$$

and for convenience choose $0 < \delta' \leq \min \left\{ \frac{\delta}{2}, \frac{1}{2} \right\}$. Then as in (4.28)

$$\begin{aligned} & \frac{1}{L} \sum_{h=1}^L E |X_{Lh}|^{1+\delta'/2} \\ & \leq 2^{1+\delta'/2} \frac{L^{\delta'/2}}{n_L^{\delta'/2}} \left(\frac{n_L}{N_L} \max_{1 \leq h \leq L} \frac{N_{Lh}}{n_{Lh}} \right)^{1+\delta'} \\ & \quad \cdot \sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |(\bar{y}_{Lh} - \bar{Y}_{Lh})(\bar{y}_{Lh} - \bar{y}_{Lh}^*)|^{1+\delta'/2} \\ & = O(1) \end{aligned} \quad (4.78)$$

provided it can be shown that

$$\sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |(\bar{y}_{Lh} - \bar{Y}_{Lh})(\bar{y}_{Lh} - \bar{y}_{Lh}^*)|^{1+\delta'/2} = O(1). \quad (4.79)$$

To establish (4.79) note that by the Cauchy-Schwarz inequality

$$\sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |(\bar{y}_{Lh} - \bar{y}_{Lh})(\bar{y}_{Lh} - \bar{y}_{Lh}^*)|^{1+\delta'/2}$$

$$\leq \left\{ \sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |\bar{y}_{Lh} - \bar{y}_{Lh}|^{2+\delta'} \cdot \sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |\bar{y}_{Lh} - \bar{y}_{Lh}^*|^{2+\delta'} \right\}^{\frac{1}{2}} \quad (4.80)$$

where $\sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |\bar{y}_{Lh} - \bar{y}_{Lh}|^{2+\delta'} = O(1)$ as in (4.29) and

(using Lemma 4.4)

$$\sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |\bar{y}_{Lh} - \bar{y}_{Lh}^*|^{2+\delta'}$$

$$\leq 2^{1+\delta'} \left\{ \sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |\bar{y}_{Lh} - \bar{y}_{Lh}|^{2+\delta'} + \sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |\bar{y}_{Lh}^* - \bar{y}_{Lh}|^{2+\delta'} \right\}$$

(4.81)

where

$$\sum_{h=1}^L w_{Lh} n_{Lh}^{\delta'/2} E |\bar{y}_{Lh}^* - \bar{y}_{Lh}|^{2+\delta'}$$

$$\leq \left(\max_{1 \leq h \leq L} \frac{n_{Lh}}{m_{Lh}} \right)^{\delta'/2} \sum_{h=1}^L w_{Lh} m_{Lh}^{\delta'/2} E |\bar{y}_{Lh}^* - \bar{y}_{Lh}|^{2+\delta'}$$

$$= O(1)$$

(4.82)

under C6 since $\sum_{h=1}^L w_{Lh} m_{Lh}^{\delta'/2} E |\bar{y}_{Lh}^* - \bar{y}_{Lh}|^{2+\delta'} = O(1)$ as in (4.29)-

(4.31). Since (by direct calculation) $E(X_{Lh}) = 0$ it now follows

from (4.77) - (4.82) (invoking Theorem 4.1) that

$$\frac{1}{L} \sum_{h=1}^L x_{Lh} = n_L \sum_{h=1}^L w_{Lh}^2 \frac{(\bar{y}_{Lh} - \bar{y}_{Lh})(\bar{y}_{Lh} - \bar{y}_{Lh}^*)}{(n_{Lh} - 1)}$$

$$\xrightarrow{P} 0. \quad /// \quad (4.83)$$

The consistency of $v_{LJ}^{(2)}$ may be established under C1-C5 and the additional condition

$$(C7) \quad \frac{1}{n_L^2} \left[\sum_{h=1}^L n_{Lh}^2 (n_{Lh} - 1) \right] \sum_{h=1}^L \frac{1}{(n_{Lh} - 1)} = O(1) \quad (4.84)$$

(Note that this condition is satisfied under bounded allocation.)

Theorem 4.15: Under C1 - C5 and C7

$$n_L v_{LJ}^{(2)} \xrightarrow{P} \sum_{k=1}^P \sum_{k'=1}^P g_k g_{k'} \gamma_{kk'}.$$

Proof: Only the necessary modifications to the proof of Theorem 4.13 in the case of $v_{LJ}^{(3)}$ need be indicated. As in (4.66)

$$\begin{aligned} n_L v_{LJ}^{(4)} &= n_L \sum_{h=1}^L \frac{1}{n_{Lh} (n_{Lh} - 1)} \sum_{i=1}^{n_{Lh}} \{ w_{Lh} (y_{Lhi} - \bar{y}_{Lh}) g'(\mu) \\ &\quad + w_{Lh} (y_{Lhi} - \bar{y}_{Lh}) (g'(\xi_L^{hi}) - g'(\mu)) \\ &\quad - \frac{1}{n_L} \sum_{l=1}^L w_{Ll} \sum_{j=1}^{n_{Ll}} (y_{Llj} - \bar{y}_{Ll}) (g'(\xi_L^{lj}) - g'(\mu)) \}^2 \end{aligned}$$

$$\begin{aligned}
 &= |g'(\mu)|^2 n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{(y_{Lhi} - \bar{y}_{Lh})^2}{(n_{Lh} - 1)} \\
 &+ n_L \sum_{h=1}^L \frac{w_{Lh}^2}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{\{(y_{Lhi} - \bar{y}_{Lh})(g'(\xi_L^{hi}) - g'(\mu))\}^2}{(n_{Lh} - 1)} \\
 &+ n_L \sum_{h=1}^L \frac{1}{(n_{Lh} - 1)} \left[\sum_{\ell=1}^L \sum_{j=1}^{n_{L\ell}} \frac{w_{L\ell}}{n_L} (y_{L\ell j} - \bar{y}_{L\ell})(g'(\xi_L^{\ell j}) - g'(\mu)) \right]^2 \\
 &+ \text{cross-product terms.} \tag{4.85}
 \end{aligned}$$

Since the first term on the right side of (4.85) $\xrightarrow{P} |g'(\mu)|^2 \gamma$ the result will follow if it can be shown that the remaining terms $\xrightarrow{P} 0$.

As in (4.67) the second term on the right side of (4.85) $\xrightarrow{P} 0$. Applying the Cauchy-Schwarz inequality to the third term

$$\begin{aligned}
 &n_L \sum_{h=1}^L \frac{1}{(n_{Lh} - 1)} \left[\sum_{\ell=1}^L \sum_{j=1}^{n_{L\ell}} \frac{w_{L\ell}}{n_L} \frac{(y_{L\ell j} - \bar{y}_{L\ell})(g'(\xi_L^{\ell j}) - g'(\mu))}{\sqrt{n_{L\ell}(n_{L\ell} - 1)}} \sqrt{n_{L\ell}(n_{L\ell} - 1)} \right]^2 \\
 &\leq n_L \sum_{\ell=1}^L \frac{w_{L\ell}^2}{n_{L\ell}} \sum_{j=1}^{n_{L\ell}} \frac{(y_{L\ell j} - \bar{y}_{L\ell})^2 (g'(\xi_L^{\ell j}) - g'(\mu))^2}{(n_{L\ell} - 1)} \left[\sum_{h=1}^L \frac{n_{Lh}^2 (n_{Lh} - 1)}{n_L^2} \right] \sum_{h=1}^L \frac{1}{(n_{Lh} - 1)} \\
 &\leq \max_{\substack{1 \leq j \leq n_L \\ 1 \leq \ell \leq L}} |g'(\xi_L^{\ell j}) - g'(\mu)|^2 n_L \text{var}(\bar{y}_{L, st}) \frac{1}{n_L^2} \left[\sum_{h=1}^L n_{Lh}^2 (n_{Lh} - 1) \right] \sum_{h=1}^L \frac{1}{(n_{Lh} - 1)} \\
 &\xrightarrow{P} 0 \tag{4.86}
 \end{aligned}$$

under C7 by 4.65.

By an argument similar to that of (4.67), the first cross-product term in (4.85) $\xrightarrow{P} 0$. Since $\sum_{i=1}^{n_{Lh}} y_{Lhi} = \bar{y}_{Lh}$, the second cross-product term is identically zero. Finally, the third cross-product term $\xrightarrow{P} 0$ using an argument similar to that in (4.86). ///

Remark 4.3: While C7 is satisfied in the case of bounded allocation, this condition is not satisfied in general. For example, if $n_{Lh} = 2$ for $h = 1, \dots, L-1$ and $n_{LL} = L$ then

$$\begin{aligned} & \frac{1}{n_L^2} \left[\sum_{h=1}^L n_{Lh} (n_{Lh} - 1) \right] \sum_{h=1}^L \frac{1}{(n_{Lh} - 1)} \\ &= (3L - 2)^{-2} (L^2 + 4) (L^2 - 2L + 2) \\ &= O(L^2) . \end{aligned} \tag{4.87}$$

Finally, the consistency of $v_{LJ}^{(5)}$ is established.

Theorem 4.16: Under C1 - C5

$$n_L v_{LJ}^{(5)} \xrightarrow{P} \sum_{k=1}^P \sum_{k'=1}^P g_k g_{k'} \gamma_{kk'} .$$

Proof: As in (4.85)

$$\begin{aligned}
 n_L v_{LJ}^{(5)} &= n_L \sum_{h=1}^L \frac{(n_{Lh}-1)}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \left\{ W_{Lh} \frac{(y_{Lhi} - \bar{y}_{Lh})}{(n_{Lh}-1)} g'(\mu) \right. \\
 &\quad \left. + W_{Lh} \frac{(y_{Lhi} - \bar{y}_{Lh})}{(n_{Lh}-1)} (g'(\xi_L^{hi}) - g'(\mu)) \right. \\
 &\quad \left. - \frac{1}{n_L} \sum_{\ell=1}^L W_{L\ell} \sum_{j=1}^{n_{L\ell}} \frac{(y_{L\ell j} - \bar{y}_{L\ell})}{(n_{L\ell}-1)} (g'(\xi_L^{\ell j}) - g'(\mu)) \right\}^2. \quad (4.88)
 \end{aligned}$$

Noting that (as in (4.86))

$$\begin{aligned}
 n_L \sum_{h=1}^L (n_{Lh}-1) &\left[\frac{1}{n_L} \sum_{\ell=1}^L W_{L\ell} \sum_{j=1}^{n_{L\ell}} \frac{(y_{L\ell j} - \bar{y}_{L\ell})}{(n_{L\ell}-1)} (g'(\xi_L^{\ell j}) - g'(\mu)) \right]^2 \\
 &\leq \max_{\substack{1 \leq j \leq n_{L\ell} \\ 1 \leq \ell \leq L}} |g'(\xi_L^{\ell j}) - g'(\mu)|^2 n_L \text{var}(\bar{y}_{L,st}) \frac{1}{n_L^2} \left[\sum_{h=1}^L (n_{Lh}-1) \right] \sum_{h=1}^L \frac{n_{Lh}^2}{(n_{Lh}-1)} \\
 &\xrightarrow{P} 0 \quad (4.89)
 \end{aligned}$$

since (using the inequality $(n_{Lh}-1)^{-1} \leq 2/n_{Lh}$)

$$\frac{1}{n_L^2} \left[\sum_{h=1}^L (n_{Lh}-1) \right] \sum_{h=1}^L \frac{n_{Lh}^2}{(n_{Lh}-1)} < 2, \quad (4.90)$$

the result now follows as in the previous theorem. ///

Remark 4.4: By Theorem 4.10 and Theorems 4.13-4.16, the jackknife

variance estimator $v_{LJ}^{(i)}$ may be used in conjunction with $\hat{\theta}_L$ in that $(\hat{\theta}_L - \theta_L) / \sqrt{v_{LJ}^{(i)}} \xrightarrow{L} N(0,1)$ for $i = 1, \dots, 5$.

4.5. Balanced Repeated Replication (BRR)

When $n_{Lh} = 2$ ($h = 1, \dots, L$) McCarthy (1966, 1969a) has developed a replication method of variance estimation in which the replicates are half-samples formed by deleting one unit in each stratum. The set of R_L half-samples used is defined by an $R_L \times L$ matrix $((\delta_{Lrh}))$ where for $r = 1, \dots, R_L$ and $h = 1, \dots, L$

$$\delta_{Lrh} = \begin{cases} +1 & \text{if unit 1 in the } h^{\text{th}} \text{ stratum is in the } r^{\text{th}} \text{ half-sample} \\ -1 & \text{if unit 2 in the } h^{\text{th}} \text{ stratum is in the } r^{\text{th}} \text{ half-sample.} \end{cases}$$

The set of half-samples is chosen so that for $h \neq h' = 1, \dots, L$

$$\sum_{r=1}^{R_L} \delta_{Lrh} = 0 \quad (4.91)$$

and

$$\sum_{r=1}^{R_L} \delta_{Lrh} \delta_{Lrh'} = 0. \quad (4.92)$$

(A set of half-samples satisfying (4.91) and (4.92) will be called "fully balanced".) Plackett & Burman (1946) provide a method of constructing a set of R_L fully balanced half-samples where $L+1 \leq R_L \leq L+4$.

Letting $\bar{y}_{L,st}^r = (\bar{y}_{L1,st}^r, \dots, \bar{y}_{Lp,st}^r)$ denote the stratified sample mean based on the r^{th} half-sample ($r = 1, \dots, R_L$) and noting that

$$\bar{y}_{Lk,st}^r - \bar{y}_{Lk,st} = \sum_{h=1}^L w_{Lh} \delta_{Lrh} d_{Lhk} \quad (4.93)$$

where $d_{Lhk} = (y_{Lh1k} - y_{Lh2k})/2$, it follows from (4.91) - (4.93) that

$$R_L^{-1} \sum_{r=1}^{R_L} \bar{y}_{Lk,st}^r = \bar{y}_{Lk,st} \quad (4.94)$$

and

$$\begin{aligned} R_L^{-1} \sum_{r=1}^{R_L} (\bar{y}_{Lk,st}^r - \bar{y}_{Lk,st}) (\bar{y}_{Lk',st}^r - \bar{y}_{Lk',st}) \\ = \text{cov}(\bar{y}_{Lk,st}, \bar{y}_{Lk',st}) \end{aligned} \quad (4.95)$$

for $k, k' = 1, \dots, p$.

To apply the method in the nonlinear case let $\hat{\theta}_L^{(r)} = g(\bar{y}_{L,st}^r)$ and $\hat{\theta}_L'^{(r)} = g(\bar{y}_{L,st}'^r)$ where $\bar{y}_{L,st}^r$ is the stratified sample mean based on the complement of the r^{th} half-sample. Then McCarthy (1966, 1969a) has proposed

$$H_1(\hat{\theta}_L) = R_L^{-1} \sum_{r=1}^{R_L} \hat{\theta}_L^{(r)} \quad (4.96)$$

and

$$H_2(\hat{\theta}_L) = (H_1(\hat{\theta}_L) + H_1'(\hat{\theta}_L))/2 \quad (4.97)$$

as estimators of θ_L (where $H_1'(\hat{\theta}_L) = R_L^{-1} \sum_{r=1}^{R_L} \hat{\theta}_L'^{(r)}$) and

$$v_{LH}^{(1)} = R_L^{-1} \sum_{r=1}^{R_L} (\hat{\theta}_L^{(r)} - \hat{\theta}_L)^2, \quad (4.98)$$

$$v_{LH}^{(2)} = R_L^{-1} \sum_{r=1}^{R_L} (\hat{\theta}_L^{(r)} - H_1(\hat{\theta}_L))^2, \quad (4.99)$$

and

$$v_{LH}^{(3)} = (4R_L)^{-1} \sum_{r=1}^{R_L} (\hat{\theta}_L^{(r)} - \hat{\theta}_L'^{(r)})^2 \quad (4.100)$$

as estimators of the variance of $\hat{\theta}_L$ or $H_1(\hat{\theta}_L)$, $i = 1, 2$.

Remark 4.5: The use of $H_1(\hat{\theta}_L)$ and $v_{LH}^{(1)}$ ($i = 1, 2$) involves at most $L+4$ replicates while the use of $H_2(\hat{\theta}_L)$ and $v_{LH}^{(3)}$ involves at most $2(L+4)$ replicates. By way of comparison all the jackknife estimators of the previous section involve $2L$ replicates (with the exception of $v_{LJ}^{(4)}$ which involves only L replicates).

The asymptotic normality of $H_1(\hat{\theta}_L)$ and $H_2(\hat{\theta}_L)$ is shown in Theorem 4.17 and the consistency of the BRR variance estimators $v_{LH}^{(i)}$ ($i = 1, 2, 3$) is shown in Theorem 4.18. Note that since $n_{Lh} = 2$ ($h = 1, \dots, L$) $C2$ reduces to $C2'$.

Theorem 4.17: Under $C1 - C4$ and $C5'$ for $i = 1, 2$

$$\sqrt{2L} (H_i(\hat{\theta}_L) - \theta_L) \xrightarrow{L} N(0, \sum_{k=1}^p \sum_{k'=1}^p g_k g_{k'} \gamma_{kk'}).$$

Proof: As in the previous section details of the proofs need only be given for the case $p = 1$.

Consider the estimator $H_1(\hat{\theta}_L)$ first. Suppressing the index k , let $I = (\mu - 2\Delta, \mu + 2\Delta)$, $\Delta > 0$, be any neighborhood of μ in which $g''(\cdot)$ is bounded. Let

$$X_{Lrh} = L W_{Lh} (y_{Lh1_r} - \bar{y}_{Lh}) \quad (4.101)$$

where

$$y_{Lh1_r} = \begin{cases} y_{Lh1} & \text{if } \delta_{Lrh} = +1 \\ y_{Lh2} & \text{if } \delta_{Lrh} = -1 \end{cases} \quad (4.102)$$

Noting that

$$\begin{aligned} L^{-1} \sum_{h=1}^L E |X_{Lrh}|^{2+\delta} &= L^{1+\delta} \sum_{h=1}^L W_{Lh}^{2+\delta} E |y_{Lh1_r} - \bar{y}_{Lh}|^{2+\delta} \\ &\leq L^{1+\delta} \max_{1 \leq h \leq L} W_{Lh}^{1+\delta} \sum_{h=1}^L W_{Lh} E |y_{Lh1} - \bar{y}_{Lh}|^{2+\delta} \\ &= O(1), \end{aligned} \quad (4.103)$$

since $\max_{1 \leq h \leq L} W_{Lh}^{1+\delta} = O(L^{-(1+\delta)})$ under $C2'$, it follows from

Theorem 4.1 that

$$\begin{aligned} P(|L^{-1} \sum_{h=1}^L X_{Lrh}| \geq \Delta) &= P(|\bar{y}_{L,st}^T - \bar{y}_L| \geq \Delta) \\ &= O(L^{-(1+\delta/2)}). \end{aligned} \quad (4.104)$$

Then using the fact that $\bar{y}_{L,st}^T$ ($r = 1, \dots, R_L$) are identically distributed random variables and (4.104)

$$\begin{aligned}
 & P\left(\max_{1 \leq r \leq R_L} |\bar{y}_{L,st}^r - \bar{Y}_L| \geq \Delta\right) \\
 & \leq \sum_{r=1}^{R_L} P(|\bar{y}_{L,st}^r - \bar{Y}_L| \geq \Delta) \\
 & = R_L P(|\bar{y}_{L,st}^1 - \bar{Y}_L| \geq \Delta) \\
 & = O(L^{-\delta/2})
 \end{aligned} \tag{4.105}$$

since $R_L = O(L)$. Then

$$\begin{aligned}
 & P\left(\max_{1 \leq r \leq R_L} |\bar{y}_{L,st}^r - \mu| < 2\Delta\right) \\
 & \geq P\left(\max_{1 \leq r \leq R_L} |\bar{y}_{L,st}^r - \bar{Y}_L| + |\bar{Y}_L - \mu| < 2\Delta\right) \\
 & \geq P\left(\max_{1 \leq r \leq R_L} |\bar{y}_{L,st}^r - \bar{Y}_L| < \Delta, |\bar{Y}_L - \mu| < \Delta\right) \\
 & \rightarrow 1
 \end{aligned} \tag{4.106}$$

under C4 by (4.105). Combining (4.106) and (4.52)

$$P(\bar{y}_{L,st}, \bar{y}_{L,st}^r \in I; r = 1, \dots, R_L) \rightarrow 1. \tag{4.107}$$

By the argument following (4.40), it may be tacitly assumed in what follows that the events described in (4.107) hold.

For $\bar{y}_{L,st}, \bar{y}_{L,st}^r \in I$ ($r = 1, \dots, R_L$),

$$\begin{aligned}
 g(\bar{y}_{L,st}^r) &= g(\bar{y}_{L,st}) + (\bar{y}_{L,st}^r - \bar{y}_{L,st}) g'(\bar{y}_{L,st}) \\
 &\quad + \frac{1}{2} (\bar{y}_{L,st}^r - \bar{y}_{L,st})^2 g''(\xi_L^r)
 \end{aligned} \tag{4.108}$$

where ξ_L^r lies between $\bar{y}_{L,st}^r$ and $\bar{y}_{L,st}$. Substituting (4.108) in (4.96) and using (4.94) leads to

$$\sqrt{2L} (H_1(\hat{\theta}_L) - \theta_L) = \sqrt{2L} (\hat{\theta}_L - \theta_L) + \epsilon_L \quad (4.109)$$

where

$$\epsilon_L = \frac{1}{2} \frac{2L}{\sqrt{2L}} R_L^{-1} \sum_{r=1}^{R_L} (\bar{y}_{L,st}^r - \bar{y}_{L,st})^2 g''(\xi_L^r). \quad (4.110)$$

Then since $|g''(\xi_L^r)| < M < \infty$ ($r = 1, \dots, R_L$) when the events in (4.107) hold it follows from (4.95), (4.61) and (4.110) that

$\epsilon_L \xrightarrow{P} 0$ due to the $\sqrt{L}$ term in the denominator in (4.110) (with ξ_L^r defined arbitrarily when the events in (4.107) do not hold).

The required result now follows from (4.109) using Slutsky's theorem since $\sqrt{2L} (\hat{\theta}_L - \theta_L) \xrightarrow{L} N(0, |g'(\mu)|^2 \gamma)$ by Theorem 4.10.

Consider next the estimator $H_2(\hat{\theta}_L)$. As in (4.109)

$$\sqrt{2L} (H_1'(\hat{\theta}_L) - \theta_L) = \sqrt{2L} (\hat{\theta}_L - \theta_L) + \epsilon_L' \quad (4.111)$$

where $\epsilon_L' \xrightarrow{P} 0$. Substituting (4.109) and (4.111) in (4.97) leads to

$$\sqrt{2L} (H_2(\hat{\theta}_L) - \theta_L) = \sqrt{2L} (\hat{\theta}_L - \theta_L) + \epsilon_L^* \quad (4.112)$$

where $\epsilon_L^* = (\epsilon_L + \epsilon_L')/2 \xrightarrow{P} 0$. ///

Theorem 4.18: Under C1 - C5 for $i = 1, 2, 3$

$$2L v_{LH}^{(i)} \xrightarrow{P} \sum_{k=1}^P \sum_{k'=1}^P g_k g_{k'} \gamma_{kk'}.$$

Proof: Consider first the estimator $v_{LH}^{(1)}$. As in the previous theorem

$$P(\bar{y}_{L,st}, \bar{y}_{L,st}^r \in I; r = 1, \dots, R_L) \rightarrow 1 \quad (4.113)$$

where $I = (\mu - \Delta, \mu + \Delta)$, $\Delta > 0$, is any neighborhood of μ in which $g'(\cdot)$ is continuous. For $\bar{y}_{L,st}, \bar{y}_{L,st}^r \in I$ ($r=1, \dots, R_L$)

$$g(\bar{y}_{L,st}^r) = g(\bar{y}_{L,st}) + (\bar{y}_{L,st}^r - \bar{y}_{L,st})g'(\xi_L^r) \quad (4.114)$$

where ξ_L^r lies between $\bar{y}_{L,st}^r$ and $\bar{y}_{L,st}$. Then as in (4.65)

$$\max_{1 \leq r \leq R_L} |g'(\xi_L^r) - g'(\mu)| \xrightarrow{P} 0 \quad (4.115)$$

(with ξ_L^r defined arbitrarily when the events in (4.113) do not hold).

Substituting (4.114) in (4.98) leads to

$$\begin{aligned} 2L v_{LH}^{(1)} &= \frac{2L}{R_L} \sum_{r=1}^{R_L} \{(\bar{y}_{L,st}^r - \bar{y}_{L,st}) g'(\mu) \\ &\quad + (\bar{y}_{L,st}^r - \bar{y}_{L,st})(g'(\xi_L^r) - g'(\mu))\}^2 \\ &= |g'(\mu)|^2 \frac{2L}{R_L} \sum_{r=1}^{R_L} (\bar{y}_{L,st}^r - \bar{y}_{L,st})^2 \\ &\quad + \frac{2L}{R_L} \sum_{r=1}^{R_L} \{(\bar{y}_{L,st}^r - \bar{y}_{L,st})(g'(\xi_L^r) - g'(\mu))\}^2 \\ &\quad + \text{cross-product term.} \end{aligned} \quad (4.116)$$

By (4.95) and (4.61), the first term on the right side of (4.116)

$\xrightarrow{P} |g'(\mu)|^2 \gamma$. Since

$$\begin{aligned} & \frac{2L}{R_L} \sum_{r=1}^{R_L} \{(\bar{y}_{L,st}^r - \bar{y}_{L,st})(g'(\xi_L^r) - g'(\mu))\}^2 \\ & \leq \max_{1 \leq r \leq R_L} |g'(\xi_L^r) - g'(\mu)|^2 \frac{2L}{R_L} \sum_{r=1}^{R_L} (\bar{y}_{L,st}^r - \bar{y}_{L,st})^2 \quad (4.117) \end{aligned}$$

the second term on the right side of (4.116) $\xrightarrow{P} 0$ by (4.115). A similar argument shows that the cross-product term $\xrightarrow{P} 0$ so that

$$2L v_{LH}^{(1)} \xrightarrow{P} |g'(\mu)|^2 \gamma \quad (4.118)$$

Consider next the estimator $v_{LH}^{(2)}$. From (4.98) and (4.99)

$$2L v_{LH}^{(2)} = 2L v_{LH}^{(1)} - 2L(H_1(\hat{\theta}_L) - \hat{\theta}_L)^2 \quad (4.119)$$

Substituting (4.114) in the remainder term in (4.119) and using (4.94) and Lemma 4.3 ($f(x) = x^2$ is convex on $(-\infty, \infty)$) leads to

$$\begin{aligned} 2L(H_1(\hat{\theta}_L) - \hat{\theta}_L)^2 &= 2L \left\{ R_L^{-1} \sum_{r=1}^{R_L} (\bar{y}_{L,st}^r - \bar{y}_{L,st})(g'(\xi_L^r) - g'(\mu)) \right\}^2 \\ &\leq \frac{2L}{R_L} \sum_{r=1}^{R_L} \{(\bar{y}_{L,st}^r - \bar{y}_{L,st})(g'(\xi_L^r) - g'(\mu))\}^2 \\ &\xrightarrow{P} 0 \quad (4.120) \end{aligned}$$

as in (4.117). The consistency of $v_{LH}^{(2)}$ now follows from (4.118) - (4.120).

Finally consider the estimator $v_{LH}^{(3)}$. As in (4.114) and (4.115)

$$g(\bar{y}_{L,st}^{(r)}) = g(\bar{y}_{L,st}) + (\bar{y}_{L,st}^{(r)} - \bar{y}_{L,st}) g'(\xi_L^{(r)}) \quad (4.121)$$

where

$$\max_{1 \leq r \leq R_L} |g'(\xi_L^{(r)}) - g'(\mu)| \xrightarrow{P} 0. \quad (4.122)$$

Substituting (4.121) and (4.114) in (4.100) and noting that, (see (4.93)) $\bar{y}_{L,st}^{(r)} - \bar{y}_{L,st} = -(\bar{y}_{L,st}^{(r)} - \bar{y}_{L,st})$ leads to

$$\begin{aligned} 2L v_{LH}^{(3)} &= \frac{2L}{4R_L} \sum_{r=1}^{R_L} \{ 2(\bar{y}_{L,st}^{(r)} - \bar{y}_{L,st}) g'(\mu) \\ &\quad + (\bar{y}_{L,st}^{(r)} - \bar{y}_{L,st}) (g'(\xi_L^{(r)}) - g'(\mu)) \\ &\quad + (\bar{y}_{L,st}^{(r)} - \bar{y}_{L,st}) (g'(\xi_L^{(r)}) - g'(\mu))^2 \} \quad (4.123) \end{aligned}$$

so that $2L v_{LH}^{(3)} \xrightarrow{P} |g'(\mu)|^2 \gamma$ as in (4.116) - (4.118). ///

Remark 4.6: Frankel (1971) has also considered the variance estimator

$$v_{LH}^{(4)} = (v_{LH}^{(1)} + v_{LH}^{(1)})/2 \quad (4.124)$$

where $v_{LH}^{(1)} = R_L^{-1} \sum_{r=1}^{R_L} (\hat{\theta}_L^{(r)} - \hat{\theta}_L)^2$. The consistency of $v_{LH}^{(4)}$ and

$$v_{LH}^{(5)} = (v_{LH}^{(2)} + v_{LH}^{(2)})/2 \quad (4.125)$$

(where $v_{LH}^{(2)} = R_L^{-1} \sum_{r=1}^{R_L} (\hat{\theta}_L^{(r)} - H_L'(\hat{\theta}_L))^2$) follow immediately from Theorem 4.18.

Remark 4.7: In the original development of the BRR method of variance estimation McCarthy (1966) employed a set of half-samples satisfying (4.92) but not necessarily (4.91). (Such a set of half-samples will be called "balanced".) It may be noted that the proof of Theorem 4.18 goes through in the case $v_{LH}^{(1)}$ and $v_{LH}^{(3)}$ even if a set of balanced (rather than fully balanced) half-samples is used. The methods of Plackett & Burman (1946) yield a set of R_L balanced half-samples where $L \leq R_L \leq L + 3$.

Remark 4.8: By Theorems 4.10 and 4.18 the BRR variance estimators may be used in conjunction with $\hat{\theta}_L$ in that $(\hat{\theta}_L - \theta_L) / \sqrt{v_{LH}^{(1)}} \xrightarrow{L} N(0,1)$ for $i = 1, \dots, 5$.

The BRR method may also be applied when $n_{Lh} = p > 2$ ($h = 1, \dots, L$) when p is a prime. Gurney & Jewett (1975) describe a method of constructing replicates comprised of one unit in stratum such that (as in (4.94) and (4.95))

$$R_L^{-1} \sum_{r=1}^{R_L} \bar{y}_{Lk,st} = \bar{y}_{Lk,st} \quad (4.126)$$

and

$$\begin{aligned}
 (p-1)^{-1} R_L^{-1} \sum_{r=1}^{R_L} (\bar{y}_{Lk,st}^r - \bar{y}_{Lk,st}) (\bar{y}_{Lk',st}^r - \bar{y}_{Lk',st}) \\
 = \text{cov} (\bar{y}_{Lk,st}, \bar{y}_{Lk',st})
 \end{aligned}
 \tag{4.127}$$

for $k, k' = 1, \dots, p$. The number of replicates required is $R_L = p^{m_L}$ where m_L is the smallest integer such that

$$m_L \geq \ln((p-1)L + 1) / \ln p \tag{4.128}$$

from which it follows that $R_L = O(L)$. The asymptotic normality of $H_1(\hat{\theta}_L)$ and the consistency of $v_{LH}^{(1)} / (p-1)$ in this general case follow using precisely the same arguments as in Theorems 4.17 and 4.18.

4.6. Stratified Multi-stage Sampling.

The results of sections 4.2 - 4.6 may be extended to any stratified multi-stage design in which the psu's may be selected with arbitrary probabilities provided that the sampling is done with replacement at the first stage and provided that independent and unbiased estimates of the totals for selected psu's are available.

Let $\underline{Y}_{Lhi} = (Y_{Lhi1}, \dots, Y_{Lhip})$ denote the vector of totals for the i^{th} psu in the h^{th} stratum in π_L ($i = 1, \dots, M_{Lh}$;

$h = 1, \dots, L$) and let $\underline{Y}_{Lh} = \sum_{i=1}^{M_{Lh}} \underline{Y}_{Lhi} = (Y_{Lh1}, \dots, Y_{Lhp})$ denote the total of all N_{Lh} units in the h^{th} stratum ($h = 1, \dots, L$).

The total of all $N_L = \sum_{h=1}^L N_{Lh}$ units in the population is then

$$\underline{Y}_L = \sum_{h=1}^L \underline{Y}_{Lh} = (Y_{L1}, \dots, Y_{Lp}).$$

As in section 4.2, let $\bar{\underline{Y}}_{Lh} = \underline{Y}_{Lh}/N_{Lh} = (\bar{Y}_{Lh1}, \dots, \bar{Y}_{Lhp})$ denote the mean of all N_{Lh} units in the h^{th} stratum ($h = 1, \dots, L$) and let $\bar{\underline{Y}}_L = \underline{Y}_L/N_L = (\bar{Y}_{L1}, \dots, \bar{Y}_{Lp})$ denote the population mean,

Suppose now that the i^{th} psu in the h^{th} stratum is selected

with probability p_{Lhi} ($i = 1, \dots, M_{Lh}$; $h = 1, \dots, L$) where $\sum_{i=1}^{M_{Lh}} p_{Lhi} = 1$

($h = 1, \dots, L$) and let $\hat{\underline{Y}}_{Lhi} = (\hat{Y}_{Lhi1}, \dots, \hat{Y}_{Lhip})$ be an unbiased estimator of $\underline{Y}_{Lhi}$ ($i = 1, \dots, M_{Lh}$; $h = 1, \dots, L$). Then the

asymptotic results of sections 4.2 - 4.6 remain valid provided that

y_{Lhi} is replaced by $\hat{y}_{Lhi}/(N_{Lh} p_{Lhi})$.

For example, replacing y_{Lhi} by $\hat{y}_{Lhi}/(N_{Lh} p_{Lhi})$ in the expression for $\bar{y}_{L,st}$ given in section 4.2 yields (with $W_{Lh} = N_{Lh}/N_L$ as before)

$$\begin{aligned}\bar{y}_{L,st} &= \sum_{h=1}^L \frac{W_{Lh}}{n_{Lh}} \sum_{i=1}^{n_{Lh}} y_{Lhi} \\ &= \frac{1}{N_L} \sum_{h=1}^L \frac{1}{n_{Lh}} \sum_{i=1}^{n_{Lh}} \frac{\hat{y}_{Lhi}}{p_{Lhi}}\end{aligned}\quad (4.129)$$

which is the usual unbiased estimator of $\bar{y}_L$ (Raj, 1968, p.120).

The dispersion matrix of $\Gamma_{Lh} = ((y_{Lhkk'}))$ of $y_{Lhi} = \hat{y}_{Lhi}/(N_{Lh} p_{Lhi})$ is defined by

$$\begin{aligned}y_{Lhkk'} &= E\left(\frac{\hat{y}_{Lhik}}{N_{Lh} p_{Lhi}} - \frac{y_{Lhk}}{N_{Lh}}\right)\left(\frac{\hat{y}_{Lhik'}}{N_{Lh} p_{Lhi}} - \frac{y_{Lhk'}}{N_{Lh}}\right) \\ &= N_{Lh}^{-2} \sum_{i=1}^{M_{Lh}} p_{Lhi} \left(\frac{y_{Lhik}}{p_{Lhi}} - y_{Lhk}\right) \left(\frac{y_{Lhik'}}{p_{Lhi}} - y_{Lhk'}\right) \\ &\quad + N_{Lh}^{-2} \sum_{i=1}^{M_{Lh}} p_{Lhi} E\left\{\left(\frac{\hat{y}_{Lhik}}{p_{Lhi}} - y_{Lhik}\right) \left(\frac{\hat{y}_{Lhik'}}{p_{Lhi}} - y_{Lhik'}\right) \mid h, i\right\}\end{aligned}\quad (4.130)$$

for $k, k' = 1, \dots, p$ where $E(\cdot \mid h, i)$ denotes the conditional

expectation within the i^{th} psu, in the h^{th} stratum. Replacing y_{Lhik} by $\hat{y}_{Lhik}/(N_{Lh} \pi_{Lhi})$ in (4.7) yields the usual unbiased estimator of $\gamma_{Lhkk'}$. The dispersion matrix of $\bar{y}_{L,st}$ in (4.129) is given by (4.8) and may be estimated as in (4.9).

4.7. Summary

The large sample results for stratified samples presented in this chapter may be appropriate in cases where relatively few p.s.u.'s are selected within each of a large number of strata. Since the efficiency of an estimator generally increases with the number of strata for the same sample size, survey designs of this type are of considerable importance and are frequently employed in practice.

The asymptotic results are obtained as $L \rightarrow \infty$ within the context of a sequence of finite populations $\{\Pi_L\}$ with L strata in Π_L . Under not unreasonable regularity conditions, the asymptotic normality of both linear and nonlinear statistics is established (Theorems 4.8 and 4.10). Consistency of the usual variance estimator in the linear case is shown (Theorem 4.9) as is the consistency of the linearization variance estimator in the nonlinear case (Theorem 4.11).

The class of nonlinear statistics covered by Theorems 4.10 and 4.11 includes those statistics which may be expressed as a function of several means. By defining new variables in terms of squares and cross-products of the original variables, these results may be applied to many of the commonly encountered nonlinear statistics such as ratios and regression and correlation coefficients (Examples 4.4 and 4.5).

The jackknife statistics proposed by McCarthy (1966) and Jones (1974) as a means of bias reduction in stratified samples are asymptotically normally distributed with the same variance as the unjackknifed statistic (Theorem 4.12). The five jackknife variance estimators proposed by various authors are all consistent (Theorems 4.13-4.15). Similar results hold in the case of BRR : the average of the half-sample estimates is asymptotically normally distributed with the same variance as the estimator based on the full sample (Theorem 4.17) and the alternative BRR variance estimators are all consistent (Theorem 4.18). Letting $\hat{\theta}_L$ denote the estimator of the parameter of interest θ_L in Π_L , it follows from these results that

$$(\hat{\theta}_L - \theta_L) / \sqrt{v_L} \xrightarrow{L} N(0,1)$$

as $L \rightarrow \infty$, where v_L is the variance estimator obtained using either the linearization, jackknife or BRR method.

While detailed proofs are presented only for stratified simple random sampling, the results remain valid for any stratified multi-stage design in which the p.s.u.'s are selected with replacement (section 4.6).

CHAPTER 5

STRATIFIED SAMPLING: SMALL SAMPLE RESULTS

In this chapter, properties of the linearization, jackknife and BRR methods applied to ratio estimation are investigated in small samples under a particular model. Comparisons of the bias of four alternative ratio estimators are made in section 5.1 while the bias of six alternative variance estimators is studied in section 5.2. The stabilities of the BRR and linearization variance estimators are compared in section 5.3.

5.1. Bias of Ratio Estimators

Suppose (y_{hi}, x_{hi}) ($i = 1, \dots, n_h$) denotes a simple random sample of n_h observations on the characters y and x from the N_h units in stratum $h = 1, \dots, L$. Within stratum h let $\bar{y}_h$ and $\bar{x}_h$ denote the sample means for the characters y and x and let $\bar{Y}_h$ and $\bar{X}_h$ denote the population means. The overall population means are then $\bar{Y} = \sum W_h \bar{Y}_h$ and $\bar{X} = \sum W_h \bar{X}_h$ where $W_h = N_h/N$ and $N = \sum N_h$. The classical combined ratio estimator of $R = \bar{Y}/\bar{X}$ is $r_1 = \bar{y}_{st}/\bar{x}_{st}$ where $\bar{y}_{st} = \sum W_h \bar{y}_h$ and $\bar{x}_{st} = \sum W_h \bar{x}_h$.

Jones' (1974) jackknife ratio estimator r_2 and McCarthy's (1966) jackknife r_3 are defined by (see section 4.4)

$$\left\{ \begin{aligned} r_2 &= [1 + E(n_h - 1)]r_1 - E(n_h - 1)r_1^h \end{aligned} \right. \quad (5.1)$$

and

$$r_3 = \frac{(\sum n_h^2)}{n} r_1 - \frac{\sum n_h (n_h - 1) r_1^h}{n} \quad (5.2)$$

respectively, where $n = \sum n_h$ and $r_1^h = n_h^{-1} \sum_{i=1}^{n_h} r_1^{hi}$ where r_1^{hi} denotes the combined ratio estimator omitting (y_{hi}, x_{hi}) .

If $n_h = p$ (a prime) for all h and a set of k balanced replicates consisting of one unit in each stratum is formed (section 4.5), then a combined ratio estimator r_{1i} may be associated with each replicate $i = 1, \dots, k$. The mean of the r_{1i} defined by $r_4 = k^{-1} \sum r_{1i}$ may also be considered as an estimator of R .

The biases of the estimators r_1, r_2, r_3 and r_4 are now investigated under proportional allocation ($n_h = W_h n$) with the model

$$y_{hi} = \alpha_h + \beta_h x_{hi} + e_{hi}, \quad (5.3)$$

$$E(e_{hi} | x_{hi}) = 0,$$

$$E(e_{hi}^2 | x_{hi}) = \delta_h x_{hi}^{c_h},$$

$$E(e_{hi} e_{hj} | x_{hi}, x_{hj}) = 0, \quad (i \neq j)$$

$$E(e_{hi} e_{h'j} | x_{hi}, x_{h'j}) = 0, \quad (h \neq h')$$

($i, j = 1, \dots, n_h$ and $h, h' = 1, \dots, L$) where x_{hi} has a gamma distribution with mean a_h .

For a variety of natural and artificial populations Rao & Kuzik (1974) found the coefficient of variation of the character x to lie between 0.4 and 1.0. Since the coefficient of variation of x_{hi} is $a_h^{-1/2}$, small values of a_h are thus of most interest. In practice t_h has often been found to lie between 0 and 2. To simplify the discussion it is assumed that the strata sizes N_h are infinite.

Under the model (5.3) with proportional allocation the bias of the combined ratio estimator may be expressed using Lemmas 3.1 and 3.2 as

$$\begin{aligned} B(r_1) &= E(r_1) - R \\ &= \frac{n \bar{\alpha}}{m(m-1)} \\ &= D_1 \bar{\alpha}, \end{aligned} \quad (5.4)$$

provided $m > 1$, where $m = \sum m_h$, $m_h = n_h a_h$ and $\bar{\alpha} = n^{-1} \sum n_h \alpha_h$. Similarly the bias of r_4 is

$$\begin{aligned} B(r_4) &= \frac{np \bar{\alpha}}{m(m-p)} \\ &= D_4 \bar{\alpha} \end{aligned} \quad (5.5)$$

provided $m > p$. The bias of r_4 relative to that of r_1 is

$$B(r_4)/B(r_1) = p(m-1)/(m-p) > 1 \quad (5.6)$$

so that for large m (corresponding to large L) $B(r_4) \approx pB(r_1)$.

Assuming $a_h = a$ and $n_h = p$ for all h so that $B(r_2)$ and $B(r_3)$ do not depend on the β_h the biases of r_2 and r_3 may be expressed as

$$\begin{aligned} B(r_2) &= n\{m^{-1}(m-1)^{-1} + L(p-1)(m-1)^{-1} \\ &\quad - L(p-1)I_\lambda(b,c)\} \bar{\alpha} \\ &= D_2 \bar{\alpha} \end{aligned} \quad (5.7)$$

and

$$\begin{aligned} B(r_3) &= \{Lp^2(m-1)^{-1} - nm^{-1} \\ &\quad - p(p-1)L I_\lambda(b,c)\} \bar{\alpha} \\ &= D_3 \bar{\alpha} \end{aligned} \quad (5.8)$$

provided $m > 1$. Here $b = pa(L-1)$, $c = a(p-1)$, $\lambda = p/(p-1)$ and $I_\lambda(a_1, a_2) = E(X_1 + \lambda X_2)^{-1}$ where X_1 and X_2 are independent gamma variates with means a_1 and a_2 respectively. I_λ is evaluated explicitly in the following theorem.

Theorem 5.1: Let X_1 and X_2 be independent gamma variates with means a and b respectively. Then for any positive, constant $\lambda \neq 1$ and integral values of a and b

$$\begin{aligned} I_\lambda(a, b) &= E(X_1 + \lambda X_2)^{-1} \\ &= I^{(1)} + I^{(2)} + I^{(3)} \end{aligned}$$

where

$$I^{(1)} = \sum_{k=1}^{a-1} (-1)^{k+1} \lambda^{k-1} \frac{\Gamma(b+k-1)}{\Gamma(b)} \frac{\Gamma(a-k)}{\Gamma(a)},$$

$$I^{(2)} = (-1)^{a+1} \lambda^{a-1} \frac{\Gamma(b+a-1)}{\Gamma(b)\Gamma(a)} \sum_{k=1}^{b+a-2} \frac{(-1)^{k+1}}{(\lambda-1)^k (b+a-k-1)},$$

$$I^{(3)} = (-1)^{b+1} \lambda^{a-1} \frac{\Gamma(b+a-1)}{\Gamma(b)\Gamma(a)} \frac{\ln \lambda}{(\lambda-1)^{b+a-1}}.$$

Proof: The proof utilizes a method introduced by P.S.R.S. Rao (1974, p.87). For $t \geq 0$ let

$$\begin{aligned} \phi(t) &= E[e^{-t(X_1 + \lambda X_2)}] \\ &= (1+t)^{-a} (1+\lambda t)^{-b}. \end{aligned} \quad (5.9)$$

Then

$$\begin{aligned} I_{\lambda}(a,b) &= \int_0^{\infty} \phi(t) dt \\ &= \int_0^{\infty} (1+t)^{-a} (1+\lambda t)^{-b} dt. \end{aligned} \quad (5.10)$$

Applying integration by parts for $a \geq 2$ yields

$$\begin{aligned} I_{\lambda}(a,b) &= \frac{1}{(a-1)} - \frac{b}{(a-1)} I_{\lambda}(a-1, b+1) \\ &= I^{(1)} + (-1)^{a+1} \lambda^{a-1} \frac{\Gamma(b+a-1)}{\Gamma(b)\Gamma(a)} I_{\lambda}(1, b+a-1). \end{aligned} \quad (5.11)$$

By partial fractions, for $b+a \geq 3$,

$$\begin{aligned} I_{\lambda}(1, b+a-1) &= \frac{1}{(\lambda-1)(b+a-2)} - \frac{1}{(\lambda-1)} I_{\lambda}(1, b+a-2) \\ &= \sum_{k=1}^{b+a-2} \frac{(-1)^{k+1}}{(\lambda-1)^k (b+a-k-1)} + \frac{(-1)^{b+a}}{(\lambda-1)^{b+a-2}} I_{\lambda}(1,1). \end{aligned} \quad (5.12)$$

The result now follows from (5.10) - (5.12) since $I_\lambda(1,1) = (\lambda-1)^{-1} \ln \lambda$. ///

Note that since $E(X_1 + \lambda X_2)^{-1} = \lambda^{-1} E(X_2 + \lambda^{-1} X_1)^{-1}$ the number of terms in the expression for $I_\lambda(a,b)$ given in Theorem 5.1 may be minimized by taking $a \leq b$.

Since the expressions for $B(r_2)$ and $B(r_3)$ given in (5.7) and (5.8) are not in closed form the biases of r_1, r_2, r_3 and r_4 with $p=2$ are compared for selected values of a in Table 5.1. While both the jackknife estimators r_2 and r_3 have smaller absolute bias than r_1 , Jones' jackknife r_2 appears particularly effective as a means of bias reduction while D_1/D_3 tends to one as L increases.

When $\beta_h = \beta$ for all h , the bias of both r_2 and r_3 does not depend on β (for arbitrary a_h and n_h). When the β_h and the a_h are not constant the bias of both r_2 and r_3 will in general involve the β_h . In this case $B(r_2) = LB(r_3)$ for $\bar{\alpha} = 0$ and $n_h = p$ for all h . Thus while r_2 is preferable to r_3 with respect to bias when $a_h = a$ for all h (Table 5.1) r_3 is preferable to r_2 when $\bar{\alpha} = 0$ and the a_h and β_h are not constant. Further, when $\bar{\alpha} = 0$ r_1 is unbiased while both r_2 and r_3 will be biased in the case of unequal a_h and β_h .

Table 5.1. Coefficients D_i in $B(r_1)$, $i = 1, 2, 3, 4$, for $p = 2$

(original values multiplied by 1000)

L	a = 1				a = 2				a = 3			
	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4
2	333	-90	121	1000	71	-6.7	32	167	30	-1.7	14	67
3	200	-13	129	500	45	-.78	30	100	20	-.15	13	42
4	143	-1.8	107	333	33	.13	25	71	14	.07	11	30
5	111	0.9	89	250	26	.30	21	56	11	.11	9.2	24
6	91	1.6	76	200	22	.31	18	45	9.5	.10	8.0	20
7	77	1.6	66	167	19	.28	16	38	8.1	.09	7.0	17
8	67	1.5	59	143	16	.24	14	33	7.1	.08	6.2	14
9	59	1.4	52	125	14	.21	13	29	6.3	.07	5.6	13
10	53	1.3	47	111	13	.19	12	26	5.6	.06	5.1	11
11	48	1.1	43	100	12	.16	11	24	5.1	.04	4.7	10
12	43	1.0	40	91	11	.14	10	22	4.7	+	+	10

† Values obtained are not accurate.

5.2. Bias of Variance Estimators

Applying Lemmas 3.1 and 3.2, the mean square error of r_1 under the model (5.3) with proportional allocation is

$$\begin{aligned} \text{MSE}(r_1) &= \frac{n^2(m+2)}{m^2(m-1)(m-2)} \bar{\alpha}^{-2} + \sum \frac{n_h(\beta_h - \bar{\beta})^2}{m(m+1)} \\ &+ \sum \frac{n_h f_h(t_h) \delta_h}{(m+t_h-1)(m+t_h-2)} \end{aligned} \quad (5.13)$$

provided $m > 2$ where $\bar{\beta} = \sum n_h \beta_h / m$ and $f_h(t) = \Gamma(a_h + t) / \Gamma(a_h)$.

The linearization estimator of $\text{MSE}(r_1)$ is

$$v_1 = \bar{x}_{st}^{-2} [v(\bar{y}_{st}) - 2r_1 \text{cov}(\bar{y}_{st}, \bar{x}_{st}) + r_1^2 v(\bar{x}_{st})] \quad (5.14)$$

where

$$v(\bar{y}_{st}) = \sum_h \frac{w_h^2}{n_h} \sum_i \frac{(y_{hi} - \bar{y}_h)^2}{(n_h - 1)}, \quad (5.15)$$

$$v(\bar{x}_{st}) = \sum_h \frac{w_h^2}{n_h} \sum_i \frac{(x_{hi} - \bar{x}_h)^2}{(n_h - 1)} \quad (5.16)$$

and

$$\text{cov}(\bar{y}_{st}, \bar{x}_{st}) = \sum_h \frac{w_h^2}{n_h} \sum_i \frac{(y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h)}{(n_h - 1)} \quad (5.17)$$

McCarthy's (1966, 1969a) balanced repeated replication (BRR) estimators of $\text{MSE}(r_1)$ when $n_h = p = 2$ for all h are (see section 4.5)

$$v_2 = \sum (r_{1i} - r_1)^2 / k, \quad (5.18)$$

$$v_3 = \sum (r_{1i} - r_4)^2 / k \quad (5.19)$$

and

$$v_4 = \sum (r_{1i} - r_{1i}^c)^2 / (4k) \quad (5.20)$$

where r_{1i}^c denotes the combined ratio estimator based on the complement of the i^{th} replicate.

Jones' (1974) and Frankel's (1971) jackknife estimators of $MSE(r_1)$ are (see section 4.4)

$$v_5 = \sum_h \frac{(n_h - 1)}{n_h} \sum_i (r_1^{hi} - r_1^h)^2 \quad (5.21)$$

and

$$v_6 = \sum_h \frac{(n_h - 1)}{n_h} \sum_i (r_1^{hi} - r_1)^2. \quad (5.22)$$

After considerable algebra the bias of v_1 under the model (5.3) with proportional allocation may be expressed, using Lemmas 3.1 and 3.2, as

$$\begin{aligned} B(v_1) = & - \frac{n^2 (3m + 2) \frac{\sigma^2}{\alpha^2}}{m^2 (m - 1) (m - 2)} - \frac{3 \sum m_h (\beta_h - \bar{\beta})^2}{m(m+1)(m+3)} \\ & - \sum \frac{n_h f_h(t_h) (t_h^2 + t_h(2m+1) - m) \delta_h}{(m+t_h+1)(m+t_h)(m+t_h-1)(m+t_h-2)} \end{aligned} \quad (5.23)$$

provided $m > 2$. Similarly, after some tedious algebra the bias of v_2 may be expressed, using Lemmas 3.1-3.4, as

$$B(v_2) = \frac{2n^2(m+2)(3m-4)\bar{\alpha}^2}{m^2(m-1)(m-2)^2(m-4)} + 4 \sum \frac{f_h(t_h)\delta_h}{(m+2t_h-2)} \left\{ \frac{1}{(m+2t_h-4)} - \frac{1}{(m+t_h-2)} \right\} \quad (5.24)$$

provided $m > 4$. Using Lemmas 3.1 and 3.2,

$$B(v_4) = \frac{n^2(3m^2+4m-16)\bar{\alpha}^2}{m^2(m-1)(m-2)^2(m-4)} - \frac{\sum m_h(\beta_h - \bar{\beta})^2}{m(m+1)(m+2)} + 2 \sum f_h(t_h)\delta_h \left\{ \frac{1}{(m+2t_h-2)(m+2t_h-4)} - \frac{1}{(m+t_h-1)(m+t_h-2)} \right\} \quad (5.25)$$

provided $m > 4$.

From (5.23), it is easily seen that $B(v_1) < 0$ when $t_h \geq 1/2$ for all h . From (5.24), $B(v_2) \geq 0$ when $t_h \leq 2$ for all h . From (5.25), $B(v_4) > 0$ when $t_h \leq 3/2$ and $\beta_h = \bar{\beta}$ for all h provided $m \geq 5$. Comparing (5.23) - (5.25) when $n_h = 2$ and $\beta_h = \bar{\beta}$ for all h shows $B(v_2) > B(v_4) > 0$ when $t_h \leq 3/2$ for all h and $B(v_2) > B(v_4) > |B(v_1)|$ when $t_h = 1$ for all h provided $m \geq 5$. When $n_h = 2$ for all h and $\bar{\alpha} = 0$ $|B(v_1)| > |B(v_4)| > B(v_2) = 0$ when $t_h = 2$ for all h .

Expressions for the biases of v_3 , v_5 and v_6 are now given for the case $n_h = p = 2$, $a_h = a$, $\beta_h = \beta$ and $t_h = t$ for all h . If a set of k balanced replicates is constructed with the properties that (i) the number of units L_0 common to each pair

of replicates is constant and (ii) each unit is included in precisely half the replicates then the bias of v_3 may be expressed, using Lemmas 3.1 and 3.2, as

$$\begin{aligned} B(v_3) &= n^2 \left\{ \frac{(k-1)}{k(m-2)(m-4)} - \frac{(m+2)}{m^2(m-1)(m-2)} - \frac{(k-1)}{4k} E_{110}(b, b, c) \right\} \bar{\alpha}^2 \\ &\quad + Lf(t) \left\{ \frac{4(k-1)}{k(m+2t-2)(m+2t-4)} - \frac{2}{(m+t-1)(m+t-2)} \right. \\ &\quad \left. - \frac{(k-2)}{2k} E_{110}(b, b, c+t) \right\} \bar{\delta} \\ &= F_3 \bar{\alpha}^2 + H_3 \bar{\delta} \end{aligned} \quad (5.26)$$

provided $m > 4$. Here $f(t) = \Gamma(a+t)/\Gamma(a)$, $\bar{\delta} = E \delta_h / L$, $b = (L - L_0)a$, $c = L_0 a$ and $E_{110}(a_1, a_2, a_3) = E(X_1 + X_3)^{-1} (X_2 + X_3)^{-1}$ where X_1 , X_2 and X_3 are independent gamma variates with means a_1 , a_2 and a_3 respectively. (E_{110} is evaluated explicitly in Theorem 3.5). If the Plackett & Burman (1946) method of constructing orthogonal replicates is employed, then the required properties may be obtained whenever $(L+1)$ is a multiple of four with $k = L+1$ replicates.

The bias of the jackknife variance estimators v_5 and v_6 may be expressed as

$$\begin{aligned} B(v_5) &= n^2 \left\{ \frac{L}{2} [I_2'(b, a) - I^*(a, a, b)] - \frac{(m+2)}{m^2(m-1)(m-2)} \right\} \bar{\alpha}^2 \\ &\quad + Lf(t) \left\{ (L-1) [I_2'(b+t, a) - I^*(a, a, b+t)] \right\} \end{aligned}$$

$$\begin{aligned}
 & + 2[I_2'(b, a+t) - (m+t-1)^{-1}(m+t-2)^{-1}] \bar{\delta} \\
 & = F_5 \bar{\alpha}^2 + H_5 \bar{\delta}
 \end{aligned} \tag{5.27}$$

and

$$\begin{aligned}
 B(v_6) &= n^2 \left\{ L[I_2'(b, a) - 2I_2''(b, a, a)] + \frac{Lm^2 - (m+2)}{m^2(m-1)(m-2)} \right\} \bar{\alpha}^2 \\
 & + 2f(t) \{ (L-1)[I_2'(b+t, a) - 2I_2''(b+t, a, a) + (m+t-1)^{-1}(m+t-2)^{-1}] \\
 & + 2[I_2'(b, a+t) - I_2''(b, a+t, a)] \} \bar{\delta} \\
 & = F_6 \bar{\alpha}^2 + H_6 \bar{\delta}
 \end{aligned} \tag{5.28}$$

provided $m > 2$. Here $b = 2a(L-1)$, $I_\lambda'(a_1, a_2) = E(X_1 + \lambda X_2)^{-2}$, $I_\lambda''(a_1, a_2, a_3) = E(X_1 + \lambda X_2)^{-1}(X_1 + X_2 + X_3)^{-1}$ and $I^*(a_1, a_2, a_3) = E(2X_1 + X_3)^{-1}(2X_2 + X_3)^{-1}$, where $\lambda > 0$ ($\lambda \neq 1$) and X_1, X_2 and X_3 are independent gamma variates with means a_1, a_2 and a_3 respectively. Explicit expressions for I_λ' , I_λ'' and I^* are given in the following three theorems.

Theorem 5.2: Let X_1 and X_2 be independent gamma variates with means a and b respectively. Then for any positive constant $\lambda \neq 1$ and integral values of a and b with $b \geq 2$,

$$\begin{aligned}
 I_\lambda'(a, b) &= E(X_1 + \lambda X_2)^{-2} \\
 &= I^{(1)} + I^{(2)} + I^{(3)}
 \end{aligned}$$

where

$$I^{(1)} = \sum_{k=1}^{a-1} (-1)^{k+1} \lambda^{k-1} \frac{\Gamma(b+k-1)\Gamma(a-k)}{\Gamma(b)\Gamma(a)} I_\lambda(a-k, b+k-1),$$

$$I^{(2)} = (-1)^{a+1} \lambda^{a-2} \frac{\Gamma(b+a-1)}{\Gamma(b)\Gamma(a)} \sum_{k=1}^{b+a-3} \frac{(-1)^{k+1}}{(\lambda-1)^k (b+a-k-1)(b+a-k-2)},$$

$$I^{(3)} = (-1)^b \frac{\lambda^{a-1}}{(\lambda-1)^{b+a-2}} \frac{\Gamma(b+a-1)}{\Gamma(b)\Gamma(a)} \left[\frac{\ln \lambda}{(\lambda-1)} - \frac{1}{\lambda} \right].$$

and I_λ is defined in Theorem 5.1.

Proof: For $t_1, t_2 \geq 0$ let

$$\begin{aligned} \phi(t_1, t_2) &= E[e^{-(t_1+t_2)(X_1 + \lambda X_2)}] \\ &= (1+t_1+t_2)^{-a} (1+\lambda t_1+\lambda t_2)^{-b}. \end{aligned} \quad (5.29)$$

Then

$$\begin{aligned} I'_\lambda(a, b) &= \int_0^\infty \int_0^\infty \phi(t_1, t_2) dt_1 dt_2 \\ &= \int_0^\infty \int_0^\infty (1+t_1+t_2)^{-a} (1+\lambda t_1+\lambda t_2)^{-b} dt_1 dt_2. \end{aligned} \quad (5.30)$$

Applying integration by parts for $a \geq 2$ yields

$$\begin{aligned} I'_\lambda(a, b) &= \frac{I_\lambda(a-1, b)}{(a-1)} - \frac{\lambda b}{(a-1)} I'_\lambda(a-1, b+1) \\ &= I^{(1)} + (-1)^{a+1} \lambda^{a-1} \frac{\Gamma(b+a-1)}{\Gamma(b)\Gamma(a)} I'_\lambda(1, b+a-1). \end{aligned} \quad (5.31)$$

By partial fractions for $b+a \geq 4$

$$\begin{aligned} I'_\lambda(1, b+a-1) &= \frac{1}{\lambda(\lambda-1)(b+a-2)(b+a-3)} - \frac{I'_\lambda(1, b+a-2)}{(\lambda-1)} \\ &= \frac{1}{\lambda} \sum_{k=1}^{b+a-3} \frac{(-1)^{k+1}}{(\lambda-1)^k (b+a-k-1)(b+a-k-2)} + \frac{(-1)^{b+a-1}}{(\lambda-1)^{b+a-3}} I'_\lambda(1, 2). \end{aligned} \quad (5.32)$$

The result now follows from (5.30) - (5.32) since $I'_\lambda(1,2) = (\lambda-1)^{-1}[(\lambda-1)^{-1} \ln \lambda - \lambda^{-1}]$. ///

Note that since $E(X_1 + \lambda X_2)^{-2} = \lambda^{-2} E(X_2 + \lambda^{-1} X_1)^{-2}$ the number of terms in the expression for $I'_\lambda(a,b)$ given in Theorem 5.2 may be minimized by taking $a \leq b$.

Theorem 5.3: Let X_1, X_2 and X_3 be independent gamma variates with means a, b and c respectively. Then for any positive constant λ

$$\begin{aligned} I''_\lambda(a,b,c) &= E(X_1 + \lambda X_2)^{-1} (X_1 + X_2 + X_3)^{-1} \\ &= (a+b+c-2)^{-1} I_\lambda(a,b) \end{aligned}$$

where $I_\lambda(a,b)$ is defined in Theorem 5.1.

Proof: Consider the transformation $S_1 = W_1/W_3$, $S_2 = W_2/W_3$ and $W_3 = X_1 + X_2 + X_3$ where $W_1 = X_1 + \lambda X_2$ and $W_2 = X_2$. Since the Jacobian of the transformation is W_3^2 , integrating over W_3 yields

$$E(X_1 + \lambda X_2)^{-1} (X_1 + X_2 + X_3)^{-1} = \frac{\Gamma(a+b+c-2)}{\Gamma(a)\Gamma(b)\Gamma(c)} M(a,b,c) \quad (5.33)$$

where

$$M(a,b,c) = \int_0^1 \int_{\lambda S_2}^{1+(\lambda-1)S_2} S_1^{-1} (S_1 - \lambda S_2)^{a-1} S_2^{b-1} (1-S_1+(\lambda-1)S_2)^{c-1} dS_1 dS_2. \quad (5.34)$$

By the same transformation

$$E(X_1 + \lambda X_2)^{-1} = \frac{\Gamma(a+b+c-1)}{\Gamma(a)\Gamma(b)\Gamma(c)} M(a,b,c) . \quad (5.35)$$

The result now follows from (5.33) and (5.35). ///

Theorem 5.4: Let X_1, X_2 and X_3 be independent gamma variates with means a, b and c respectively. Then for integral values of a, b and c

$$\begin{aligned} I^*(a,b,c) &= E(2X_1 + X_3)^{-1} (2X_2 + X_3)^{-1} \\ &= I^{(1)} + I^{(2)} + I^{(3)} \end{aligned}$$

where

$$I^{(1)} = \sum_{k=1}^{a-1} \frac{(-1)^{k+1}}{2^k} \frac{\Gamma(c+k-1)}{\Gamma(c)} \frac{\Gamma(a-k)}{\Gamma(a)} I_2(c+k-1, b) ,$$

$$I^{(2)} = \frac{(-1)^{a+1}}{2^{a-1}} \sum_{k=1}^{b-1} \frac{(-1)^{k+1}}{2^k} \frac{\Gamma(c+a+k-2)}{\Gamma(c)} \frac{\Gamma(b-k)}{\Gamma(b)} I_2(c+a+k-2, 1) ,$$

$$I^{(3)} = \frac{(-1)^{a+b}}{2^{a+b-2}} \frac{\Gamma(c+b+a-2)}{\Gamma(c)\Gamma(b)\Gamma(a)} \frac{I_2(c+b+a-2, 1)}{(c+b+a-2)} ,$$

and I_2 is defined in Theorem 5.1.

Proof: For $t_1, t_2 \geq 0$ let

$$\begin{aligned} \phi(t_1, t_2) &= E[e^{-t_1(2X_1+X_3)-t_2(2X_2+X_3)}] \\ &= (1+2t_1)^{-a} (1+2t_2)^{-b} (1+t_1+t_2)^{-c} . \end{aligned} \quad (5.36)$$

Then

$$I^*(a,b,c) = \int_0^\infty \int_0^\infty \phi(t_1, t_2) dt_1 dt_2$$

$$= \int_0^\infty \int_0^\infty (1+2t_1)^{-a} (1+2t_2)^{-b} (1+t_1+t_2)^{-c} dt_1 dt_2. \quad (5.37)$$

Applying integration by parts for $a \geq 2$ yields

$$I^*(a,b,c) = \frac{I_2(c,b)}{2(a-1)} - \frac{c}{2(a-1)} I^*(a-1,b,c+1)$$

$$= I(1) + \frac{(-1)^{a+1}}{2^{a-1}} \frac{\Gamma(c+a-1)}{\Gamma(c)\Gamma(a)} I^*(1,b,c+a-1). \quad (5.38)$$

By partial fractions for $b \geq 2$

$$I^*(1,b,c+a-1) = \frac{I_2(c+a-1,1)}{2(b-1)} - \frac{(c+a-1)}{2(b-1)} I^*(1,b-1,c+a)$$

$$= \sum_{k=1}^{b-1} \frac{(-1)^{k+1}}{2^k} \frac{\Gamma(c+a+k-2)}{\Gamma(c+a-1)} \frac{\Gamma(b-k)}{\Gamma(b)} I_2(c+a+k-2,1)$$

$$+ \frac{(-1)^{b+1}}{2^{b-1}} \frac{\Gamma(c+b+a-2)}{\Gamma(c+a-1)\Gamma(b)} I^*(1,1,c+b+a-2). \quad (5.39)$$

The result now follows from (5.37) - (5.39) since

$$I^*(1,1,c+b+a-2) = I_2(c+b+a-2,1)/(c+b+a-2). \quad ///$$

The biases of v_1 , v_2 and v_4 may also be expressed as $B(v_1) = F_1 \bar{\alpha}^2 + H_1 \bar{\delta}$ in the case $n_h = 2$, $a_h = a$, $\beta_h = \beta$ and $t_h = t$ for all h , where F_1 and H_1 ($i = 1, 2, 4$) are defined by (5.23) - (5.25). Since the expressions for F_1 and H_1 ($i=3, 5, 6$) are not in closed form the biases of the variance estimators v_i ($i = 1, 2, 3, 4, 5, 6$) are compared for selected values of t , a and L in Tables 5.2 - 5.5.

As noted above, $B(v_1) < 0$ when $t \geq 1/2$. From Tables 5.2 - 5.5 both the jackknife variance estimators v_5 and v_6 also underestimate $MSE(r_1)$ when $t = 1$ or 2 and $L > 4$ with $|B(v_1)| > |B(v_5)| > |B(v_6)|$ in this case. For $t \leq 3/2$ it was shown above that both the BRR variance estimators v_2 and v_4 are overestimates. From Tables 5.2 - 5.5 v_3 is also an overestimate when $t = 0$ or 1 with $B(v_2) > B(v_3) > B(v_4)$ in this case. When $t = 1$, $B(v_2) > B(v_3) > B(v_4) > |B(v_1)| > |B(v_5)| > |B(v_6)|$ provided $L > 4$.

When in addition $\bar{\alpha} = 0$ it is seen from Tables 5.3 - 5.5 that all the variance estimators are overestimates for $t = 0$ with $B(v_2) > B(v_3) > B(v_4) > B(v_6) > B(v_5) > B(v_1)$. For $t = 2$ all the variance estimators are underestimates when $\bar{\alpha} = 0$ with the absolute biases following the reverse order to that for $t = 0$.

Table 5.2. Coefficients F_i in $B(v_i)$, $i = 1, 2, 3, 4, 5, 6$ for $p = 2$

(original values multiplied by 1000)

L	a = 1						a = 2						a = 3					
	F_1	F_2	F_3^*	F_4	F_5	F_6	F_1	F_2	F_3^*	F_4	F_5	F_6	F_1	F_2	F_3^*	F_4	F_5	F_6
2	-467	††		††	819	1791	-17	99		52	7	22	-3.0	11.3		5.9	.7	2.7
3	-143	1400	1018	725	20	121	-7	25	18	13	-1	2	-1.2	3.6	2.6	1.9	-.3	.2
4	-69	397		206	-14	17	-3.5	11.2		5.8	-1.3	-1	-.65	1.76		.90	-.27	-.07
5	-40	181		94	-15	-1	-2.2	6.3		3.2	-1.1	-.5	-.41	1.03		.53	-.21	-.11
6	-27	102		53	-13	-5	-1.5	4.0		2.0	-.8	-.5	-.28	.68		.34	-.17	-.11
7	-19	65	51	34	-10	-6	-1.1	2.7	2.2	1.4	-.7	-.5	-.20	.48	.38	.24	-.13	-.10
8	-14	45		23	-8	-5	-.8	2.0		1.0	-.5	-.4	-.15	.36		.18	-.11	-.08
9	-11	33		17	-7	-5	-.6	1.5		.8	-.4	-.3	-.12	.27		.14	-.09	-.07
10	-9	25		13	-6	-4	-.5	1.2		.6	-.4	-.3	-.10	.22		.11	-.07	-.06
11	-7	20	16	10	-5	-4	-.4	1.0	.8	.5	-.3	-.3	-.08	.18	†	.09	†	†
12	-6	16		8	-4	-3	-.3	.8		.4	-.3	-.2	-.07	.15		.07	†	†

† Values obtained are not accurate.

†† Not defined for $L \leq 2$.

* F_3 computed only for (Ltd) a multiple of four.

Table 5.3. Coefficients H_i in $B(v_i)$, $i = 1, 2, 3, 4, 5, 6$ for $p=2$ and $t=0$.

(original values multiplied by 1000)

L	a = 1						a = 2						a = 3					
	H_1	H_2	H_3^*	H_4	H_5	H_6	H_1	H_2	H_3^*	H_4	H_5	H_6	H_1	H_2	H_3^*	H_4	H_5	H_6
2	133	††		††	788	1259	11	111			32	44	3	20		14	7	10
3	43	750	608	450	111	151	4	30	26		8	11	1.2	6.7	5.9	4.7	2.2	2.7
4	21	222		143	41	52	2	14		10	4	5	.6	3.3		2.4	1.0	1.2
5	13	104		69	21	25	1.4	7.7		5.5	2.1	2.5	.4	2.0		1.4	.6	.7
6	8	60		41	13	15	.9	5.0		3.6	1.3	1.5	.3	1.3		.9	.4	.4
7	6	39	36	27	8	10	.7	3.5	3.3	2.5	.9	1.0	.20	.92	.88	.67	.26	.29
8	4	27		19	6	7	.5	2.5		1.8	.7	.7	.15	.69		.50	.19	.21
9	3	20		14	5	5	.4	1.9		1.4	.5	.6	.12	.53		.39	.15	.16
10	3	15		11	4	4	.3	1.5		1.1	.4	.4	.10	.42		.31	.12	.13
11	2	12	12	9	3	3	.3	1.2	1.2	.9	.3	.4	.08	.35	†	.26	†	†
12	2	10		7	2	2	.2	1.0		.8	.3	.3	.07	.29		.21	†	†

† Values obtained are not accurate.

†† Not defined for $L_a \leq 2$.

* H_3 computed only for $(L+1)$ a multiple of four.

Table 5.4. Coefficients H_1 in $B(v_1)$, $i = 1, 2, 3, 4, 5, 6$ for $p=2$ and $t=1$

(original values multiplied by 1000)

L	a = 1						a = 2						a = 3					
	H_1	H_2	H_3^*	H_4	H_5	H_6	H_1	H_2	H_3^*	H_4	H_5	H_6	H_1	H_2	H_3^*	H_4	H_5	H_6
2	-67	++		++	30	90	-16	48		24	3	13	-7	18		9	.8	.5
3	-29	100	77	50	-2	13	-7	18	14	9	-1	2	-3.1	7.4	5.9	3.7	-8	.5
4	-16	48		24	-4	2	-3.9	9.5		4.8	-1.5	-.3	-1.7	4.0		2.0	-.7	-.2
5	-10	28		14	-4	-1	-2.5	5.8		2.9	-1.2	-.6	-1.1	2.5		1.2	-.6	-.3
6	-7	18		9	-3	-2	-1.7	4.0		2.0	-1.0	-.6	-.8	1.7		.8	-.5	-.3
7	-5	13	11	6	-3	-2	-1.3	2.8	2.6	1.4	-.8	-.6	-.6	1.2	1.1	.6	-.4	-.3
8	-4	10		5	-2	-2	-1.0	2.2		1.1	-.7	-.5	-.43	.93		.46	-.30	-.23
9	-3.1	7.4		3.7	-2.0	-1.4	-.8	1.7		.8	-.5	-.4	-.34	.73		.36	-.25	-.20
10	-2.5	5.8		2.9	-1.7	-1.3	-.6	1.3		.7	-.5	-.4	-.28	.58		.29	-.21	-.17
11	-2.1	4.8	4.4	2.4	-1.4	-1.1	-.5	1.1	1.0	.6	-.4	-.3	-.23	.48	+	.24	+	+
12	-1.7	4.0		2.0	-1.2	-1.0	-.43	.93		.46	-.33	-.28	-.19	.40		.20	+	+

† Values obtained are not accurate.

†† Not defined for $L \leq 2$.

* H_3 computed only for $(L+1)$ a multiple of four.

Table 5.5. Coefficients H_i in $B(v_i)$, $i = 1, 2, 3, 4, 5, 6$ for $p=2$ and $t=2$

(original values multiplied by 1000)

L	a = 1						a = 2						a = 3					
	H_1	H_2	H_3^*	H_4	H_5	H_6	H_1	H_2	H_3^*	H_4	H_5	H_6	H_1	H_2	H_3^*	H_4	H_5	H_6
2	-171	++		++	-95	-49	-91	0		-33	-53	-32	-62	0		-22	-37	-24
3	-95	0	-20	-36	-60	-40	-46	0	-8	-16	-32	-23	-30	0	-5	-11	-21	-16
4	-61	0		-22	-41	-30	-28	0		-10	-20	-16	-18	0		-6	-13	-11
5	-42	0		-15	-30	-24	-19	0		-6	-14	-12	-12	0		-4	-9	-8
6	-31	0		-11	-23	-19	-13	0		-5	-11	-9	-8.3	.0		-2.8	-6.8	-6.1
7	-24	0	-2	-8	-18	-15	-10	0	-1	-3	-8	-7	-6.2	.0	-.4	-2.1	-5.2	-4.7
8	-19	0		-7	-15	-13	-7.8	0.0		-2.7	-6.5	-5.9	-4.8	.0		-1.6	-4.1	-3.8
9	-15	0		-5	-12	-11	-6.2	0.0		-2.1	-5.3	-4.9	-3.8	.0		-1.3	-3.3	-3.1
10	-12	0		-4	-10	-9	-5.1	0.0		-1.7	-4.4	-4.1	-3.1	.0		-1.1	-2.8	-2.6
11	-10	0	-1	-4	-9	-8	-4.3	0.0	-.2	-1.5	-3.7	-3.5	-2.6	.0	+	-.9	+	+
12	-9	0		-3	-7	-7	-3.6	0.0		-1.2	-3.2	-3.0	-2.2	.0		-.7	+	+

† Values obtained are not accurate.

++ Not defined for $L \leq 2$.

* H_3 computed only for $(L+1)$ a multiple of four.

5.3. Stability of Variance Estimators

The mean square errors of the linearization estimator v_1 and the BRR estimator v_2 are now derived under the model (5.3) with normally distributed errors for the case $n_h = p = 2$, $a_h = a$, $t_h = t$, $\beta_h = \beta$ and $\delta_h = \delta$ for all h . Noting that

$$\text{MSE}(v_1) = E(v_1^2) - 2 \text{MSE}(r_1)B(v_1) - \text{MSE}^2(r_1), \quad (5.40)$$

only $E(v_1^2)$ ($i = 1, 2$) need be evaluated in view of (5.13), (5.23) and (5.24). After considerable algebra $E(v_1^2)$ may be expressed, using Lemmas 3.1 and 3.2, as

$$E(v_1^2) = J_1' \alpha^4 + K_1' \alpha^2 \delta + L_1' \delta^2 \quad (5.41)$$

where

$$J_1' = \frac{64a L^5 [a(L+2) + 3]}{(m+3)^{[8]}}, \quad (5.42)$$

$$\begin{aligned} K_1' = 16L^3 f(t) & \left\{ \frac{2a(L+2)+3t(t+1)}{(m+t+1)^{[6]}} - 6t \frac{[2a(L+2)+(t+1)(t+2)]}{(m+t+2)^{[7]}} \right. \\ & \left. + \frac{12a[t(t+1)(L+2)+3L(a+1)+aL(L-1)]+3t(t+1)(t+2)(t+3)}{(m+t+3)^{[8]}} \right\}, \quad (5.43) \end{aligned}$$

$$\begin{aligned} L_1' = 6Lf(2t) & \left\{ \frac{1}{(m+2t-1)^{[4]}} - \frac{8t}{(m+2t)^{[5]}} + \frac{4[3t(2t+1)+a(L+2)]}{(m+2t+1)^{[6]}} \right. \\ & - 16t \frac{[a(L+2)+(t+1)(2t+1)]}{(m+2t+2)^{[7]}} \\ & \left. + 4 \frac{[2at(L+2)(2t+1)+3La(a+1)+a^2L(L-1)+t(t+1)(2t+1)(2t+3)]}{(m+2t+3)^{[8]}} \right\} \end{aligned}$$

$$\begin{aligned}
 & + 2Lf^2(t) \left\{ \frac{(2L+1)}{(m+2t-1)^{[4]}} - \frac{8t(L-1)}{(m+2t)^{[5]}} \right. \\
 & + 4 \frac{[a(L+2)(2L-3)+t(4L-7)+8t^2(L-1)]}{(m+2t+1)^{[6]}} \\
 & - 48t(L-1) \frac{[a(L+2) + (t+1)^2]}{(m+2t+2)^{[7]}} \\
 & + 12[3La(a+1)(2L-1)+a^2L(L-1)(2L-1)] + 2at(L+2) \\
 & + 3t(t+1)+4at(t+1)(L-1)(L+2)+t(t+1)(t+2)(t+3)(L-1) \\
 & \left. + t^2(t+1)^2(L-1)]/(m+2t+3)^{[8]} \right\} \quad (5.44)
 \end{aligned}$$

provided $m > 4$, where $f(t) = \Gamma(a+t)/\Gamma(a)$ and $z^{[s]} = z(z-1)\dots(z-s+1)$.

If a set of k balanced replicates is selected with the property that the number of units L_0 common to each pair of replicates is constant then it follows after considerable algebra (using the methods illustrated in Lemmas 3.1 - 3.4) that

$$E(v_2^2) = J_2' \alpha^4 + K_2' \alpha^2 \delta + L_2' \delta^2 \quad (5.45)$$

where

$$\begin{aligned}
 J_2' = & n^4 \left\{ \frac{1}{16k(a-1)^{[4]}} - \frac{1}{2k(a-1)^{[3]}(m-4)} + \frac{(k+2)}{2k(a-1)^{[2]}(m-3)^{[2]}} \right. \\
 & - \frac{2}{(a-1)(m-2)^{[3]}} + \frac{1}{(m-1)^{[4]}} + \frac{(k-1)}{16k} E_{220}(b,b,c) \\
 & \left. - \frac{(k-1)}{2k} E_{211}^*(b,b,c,c) + \frac{(k-1)}{k} E_{112}^*(b,b,c,c) \right\}, \quad (5.46)
 \end{aligned}$$

$$\begin{aligned}
 K_2' = n^2 f(t) \{ & \frac{3L}{2k(a_+ + t - 1)^{[4]}} - \frac{9L}{k(a_+ + t - 1)^{[3]}(m + t - 4)} \\
 & + \frac{13L(k+2)}{2k(a_+ + t - 1)^{[2]}(m + t - 3)^{[2]}} + \frac{L(k+2)}{2k(a_+ - 1)^{[2]}(m + t - 4)^{[2]}} \\
 & - \frac{18L}{(a_+ + t - 1)(m + t - 2)^{[3]}} - \frac{6L}{(a_+ - 1)(m + t - 2)^{[3]}} + \frac{12L}{(m + t - 1)^{[4]}} \\
 & + \frac{(k-1)(L-L_0)}{2k} E_{220}(b+t, b, c) + \frac{3L_0(k-1)}{2k} E_{220}(b, b, c+t) \\
 & - \frac{9L_0(k-1)}{k} E_{211}^*(b, b, c+t, c) - \frac{(k-1)(L-L_0)}{k} E_{211}^*(b, b+t, c, c) \\
 & - \frac{4(k-1)(L-L_0)}{k} E_{211}^*(b+t, b, c, c) \\
 & + 10 \frac{(k-1)(L-L_0)}{k} E_{112}^*(b+t, b, c, c) + 13L_0 \frac{(k-1)}{k} E_{112}^*(b, b, c+t, c) \\
 & + L_0 \frac{(k-1)}{k} E_{112}^*(b, b, c, c+t) \} , \tag{5.47}
 \end{aligned}$$

$$\begin{aligned}
 L_2' = & \frac{3L[f(2t) + (L-1)f^2(t)]}{k(a_+ + 2t - 1)^{[4]}} - \frac{12L[f(2t) + (L-1)f^2(t)]}{k(a_+ + 2t - 1)^{[3]}(m + 2t - 4)} \\
 & + 2 \frac{(k+2)}{k} \left\{ \frac{3L[f(2t) + (L-1)f^2(t)]}{(a_+ + 2t - 1)^{[2]}(m + 2t - 3)^{[2]}} + \frac{L^2 f^2(t)}{(a_+ + t - 1)^{[2]}(m + 2t - 3)^{[2]}} \right\} \\
 & - 12 \left\{ \frac{L[f(2t) + (L-1)f^2(t)]}{(a_+ + 2t - 1)(m + 2t - 2)^{[3]}} + \frac{L^2 f^2(t)}{(a_+ + t - 1)(m + 2t - 2)^{[3]}} \right\} \\
 & + \frac{6L[f(2t) + (2L-1)f^2(t)]}{(m + 2t - 1)^{[4]}}
 \end{aligned}$$

$$\begin{aligned}
& + \frac{(k-1)}{k} (L-L_0)^2 f^2(t) E_{220}(b+t, b+t, c) \\
& + 2L_0(L-L_0) \frac{(k-1)}{k} f^2(t) E_{220}(b+t, b, c+t) \\
& + 3L_0 \frac{(k-1)}{k} [f(2t) + (L_0-1)f^2(t)] E_{220}(b, b, c+2t) \\
& - 12L_0(L-L_0) \frac{(k-1)}{k} f^2(t) E_{211}^*(b+t, b, c+t, c) \\
& - 4 \frac{(k-1)}{k} (L-L_0)^2 f^2(t) E_{211}^*(b+t, b+t, c, c) \\
& - 4L_0(L-L_0) \frac{(k-1)}{k} f^2(t) E_{211}^*(b, b+t, c+t, c) \\
& - 12L_0 \frac{(k-1)}{k} [f(2t) + (L_0-1)f^2(t)] E_{211}^*(b, b, c+2t, c) \\
& + 8 \frac{(k-1)}{k} (L-L_0)^2 f^2(t) E_{112}^*(b+t, b+t, c, c) \\
& + 24L_0(L-L_0) \frac{(k-1)}{k} f^2(t) E_{112}^*(b+t, b, c+t, c) \\
& + 12L_0 \frac{(k-1)}{k} [f(2t) + (L_0-1)f^2(t)] E_{112}^*(b, b, c+2t, c) \\
& + 4L_0^2 \frac{(k-1)}{k} f^2(t) E_{112}^*(b, b, c+t, c+t) \tag{5.48}
\end{aligned}$$

provided $a = \sum a_h > 4$. Here $b = a(L-L_0)$, $c = aL_0$ and

$$E_{b_1 b_2 b_3}^*(a_1, a_2, a_3, a_4) = E[(X_1+X_3)^{-b_1} (X_2+X_3)^{-b_2} (X_1+X_2+X_3+X_4)^{-b_3}] \tag{5.49}$$

where X_1, X_2, X_3 and X_4 are independent gamma variates with means a_1, a_2, a_3 and a_4 respectively. ($E_{220}(a_1, a_2, a_3)$ is defined in (3.19) and evaluated in Theorem 3.7.)

In order to express the mean square errors of v_1 and v_2 in the form

$$\text{MSE}(v_i) = J_i \bar{\alpha}^4 + K_i \bar{\alpha}^2 \delta + L_i \delta^2 \quad (5.50)$$

(i = 1,2) it remains only to evaluate the expectations E_{112}^* and E_{211}^* in (5.46), (5.47) and (5.48). The following theorem shows that these expectations may be obtained using the expressions for E_{110} and E_{210} (see (3.19)) given in Theorems 3.5 and 3.6 respectively.

Theorem 5.5: Let X_1, X_2, X_3, X_4 be independent gamma variates with means a, b, c, d respectively. Then for $b_1, b_2, b_3 \geq 0$ with $a + b + c + d > b_1 + b_2 + b_3$

$$K_{b_1 b_2 b_3}^*(a, b, c, d) = \frac{\Gamma(a+b+c+d-b_1-b_2-b_3)}{\Gamma(a+b+c+d-b_1-b_2)} K_{b_1 b_2 0}(a, b, c)$$

where $K_{b_1 b_2 0}$ is defined in (3.19).

Proof: Consider the transformation. $S_1 = W_1/W_4$, $S_2 = W_2/W_4$, $S_3 = W_3/W_4$ and $W_4 = X_1 + X_2 + X_3 + X_4$ where $W_1 = X_1 + X_3$, $W_2 = X_2 + X_3$ and $W_3 = X_4$. Since the Jacobian of the transformation is W_4^3 , integrating over W_4 yields

$$K_{b_1 b_2 b_3}^*(a, b, c, d) = \Gamma(a+b+c+d-b_1-b_2-b_3) H_{b_1 b_2}(a, b, c, d) \quad (5.51)$$

where

$$H_{b_1 b_2}(a, b, c, d) = \int_0^1 \int_0^{1-S_3} \int_0^{1-S_2-S_3} \frac{(1-S_2-S_3)^{a-1} (1-S_1-S_3)^{b-1} (S_1+S_2+S_3-1)^{c-1} S_3^{d-1}}{S_1^{b_1} S_2^{b_2} \Gamma(a) \Gamma(b) \Gamma(c) \Gamma(d)} dS_1 dS_2 dS_3. \quad (5.52)$$

Similarly

$$E_{b_1 b_2^0}(a, b, c) = \Gamma(a+b+c+d-b_1-b_2) H_{b_1 b_2}(a, b, c, d) . \quad (5.53)$$

The result now follows from (5.51) - (5.53). ///

The coefficients J_1 , K_1 and L_1 in the mean square error of v_1 ($i = 1, 2$) are given in Tables 5.6 - 5.8. These coefficients decrease as L or a increase so that the mean square errors of v_1 and v_2 decrease as the number of strata increases or as the coefficient of variation of the x population within each stratum decreases.

Tables 5.6-5.8 indicate that $J_1 < J_2$, $K_1 < K_2$ and $L_1 < L_2$ for all values of t , a and L so that v_1 is more stable than v_2 . The ratios J_2/J_1 , K_2/K_1 and L_2/L_1 all decrease however as L or a increase. The ratio L_2/L_1 is particularly close to one for moderate values of La , in which case the stability of v_1 is comparable to that of v_2 when $\bar{\alpha} = 0$. Since L_2/L_1 decreases as t increases, $MSE(v_2)/MSE(v_1)$ decreases as t increases when $\bar{\alpha} = 0$.

Table 5.6. Coefficients J_1 in $MSE(v_i)$, $i = 1, 2$, for $p = 2$

(original values multiplied by 10^6)

L	a = 1		a = 2		a = 3	
	J_1	J_2	J_1	J_2	J_1	J_2
3	297143	++	969	30075	55	502
4	64723	++	336	5148	21	142
7	6507	94124	50	198	3.5	8.9
8	4000	47995	33	131	2.3	6.2
11	1321	6706	12	28	0.8	+
12	985	5112	9	23	0.6	+

+ Values obtained are not accurate.

++ Not defined for $L \leq 4$.

Table 5.7. Coefficients K_1 in $MSE(v_1)$, $i = 1, 2$, for $p = 2$

L	t	a = 1		a = 2		a = 3	
		K_1	K_2	K_1	K_2	K_1	K_2
3	0	1011429	++	5538	88931	439	2303
	1	148571	++	4179	34622	664	2679
	2	83018	++	6719	36774	1775	5872
4	0	195090	++	1780	15023	156	641
	1	47842	++	1649	10263	272	984
	2	38858	++	3278	16207	844	2705
7	0	16795	140352	240	640	24	45
	1	6978	38596	288	700	49	90
	2	9460	38216	763	1661	186	322
8	0	10046	70626	152	411	15	30
	1	4530	24275	191	489	33	64
	2	6728	28226	532	1232	129	236
11	0	3140	10534	53	97	6	†
	1	1653	4820	72	128	13	†
	2	2939	7350	222	369	52	†
12	0	2313	7850	40	77	4	†
	1	1260	3892	56	105	10	†
	2	2334	6230	174	308	41	†

† Values obtained are not accurate.

++ Not defined for $L_a \leq 4$.

Table 5.8. Coefficients L_A in $MSE(v_i)$, $i = 1, 2$, for $p = 2$
(original values multiplied by 10^6)

L	t	a = 1		a = 2		a = 3	
		L_1	L_2	L_1	L_2	L_1	L_2
3	0	555714	++	5438	28070	632	1437
	1	40476	++	6753	14196	2643	4283
	2	54374	++	36635	56518	30715	42734
4	0	99495	++	1683	5365	220	429
	1	15584	++	2781	5587	1103	1767
	2	30212	++	18634	30142	14890	21350
7	0	7717	25421	217	340	32	43
	1	2677	5016	508	697	204	252
	2	8845	12944	4602	5814	3393	4024
8	0	4534	13162	137	213	21	28
	1	1776	3389	339	476	136	172
	2	6482	9883	3247	4205	2356	2846
11	0	1370	2504	47	61	7	+
	1	672	996	130	159	52	+
	2	3006	3939	1387	1626	973	+
12	0	1001	1824	35	46	6	+
	1	516	787	100	125	40	+
	2	2421	3254	1094	1305	762	+

+ Values obtained are not accurate.

++ Not defined for $L_a \leq 4$.

5.4. Summary

Properties of the linearization, jackknife and BRR methods applied to ratio estimation are investigated in stratified samples involving a small or moderate number of strata L under the linear regression model in (5.3) with proportional allocation.

Results concerning the jackknife method as a means of bias reduction in stratified samples are not clearcut. Under equal allocation and the assumption that the distribution of the x population is the same in all strata, Jones' jackknife is particularly effective as a means of bias reduction while the absolute bias of McCarthy's jackknife is comparable to that of the classical combined ratio estimator for moderate L (Table 5.1). When the distribution of the x population is not the same in all strata, however, McCarthy's jackknife can have smaller absolute bias than Jones'. Moreover, both jackknife estimators may be biased in situations where the classical estimator is unbiased.

Since the bias of the classical ratio estimator has been found to be negligible for the most part in many large scale surveys (thereby limiting the need for an estimator with reduced bias), the exact mean square errors of the jackknife estimators are not derived. (Note, however, that as shown in Chapter 4, all the ratio estimators considered here have the same asymptotic variance as $L \rightarrow \infty$.)

When $t_h \geq 1/2$ in each stratum h , the linearization variance estimator v_1 underestimates the mean square error of the classical ratio estimator. (Recall that since $V(e_{hi} | x_{hi}) = \delta_h x_{hi}^{t_h}$ in (5.3), the parameter t_h reflects the degree of heteroskedasticity of the error variance in stratum h .) When $t_h \leq 2$ in each stratum h , the BRR variance estimator v_2 is an overestimate.

Further results on the biases of the alternative variance estimators are obtained under some simplifying assumptions for the case of two units selected in each stratum. When the distribution of the x populations and the slopes β_h in (5.3) are the same in each stratum h , both jackknife variance estimators v_5 and v_6 tend to underestimate the mean square error of the classical ratio estimator for $t_h = t = 1$ or 2 . When $t_h = t = 0$ or 1 , the three BRR variance estimators v_2 , v_3 and v_4 are all overestimates (Tables 5.2 - 5.5). When in addition the average $\bar{\alpha}$ of the intercepts α_h in (5.3) is zero (as in the case of regression through the origin in each stratum), all six variance estimators are overestimates for $t_h = t = 0$ and underestimates for $t_h = t = 2$ (with v_2 unbiased in the latter case).

The mean square errors of the linearization variance estimator v_1 and the BRR variance estimator v_2 are derived under the assumption that all parameters in (5.3) are the same in each stratum h (with the exception of the intercepts α_h). The mean square errors of both estimators decrease as the number of strata L

increases or as the coefficient of variation C_x of the x populations decreases. The stability of v_1 is better than that of v_2 , although for $\bar{\alpha} = 0$ the two variance estimators are of somewhat comparable stability provided L is moderately large or C_x is relatively small (Tables 3.6-3.8). Analytic results on the stabilities of the remaining variance estimators could not be obtained.

In an empirical study using data from the U.S. Bureau of the Census' Current Population Survey, Kish & Frankel (1974) found little difference in the mean square errors of v_1 and v_2 . On the basis of the present analysis, Kish & Frankel's finding may be attributable to a regression approximately through the origin and the small coefficients of variation of the x populations involved ($C_x = 0.19, 0.13$ and 0.076).

CHAPTER 6

STRATIFIED SAMPLING: SOME INEFFICIENT REPLICATION VARIANCE

ESTIMATORS REQUIRING A REDUCED NUMBER OF REPLICATES

Application of both the BRR and jackknife methods of variance estimation can become cumbersome as well as expensive when the number of replicates involved is large. While a number of replication methods of variance estimation requiring a reduced number of replicates have been proposed (see McCarthy [1966], Kish & Frankel [1970], Lee [1972, 1973b] and Lemeshow [1976]), the reduction in computational cost is accompanied by a loss in precision. In this chapter, this loss in precision is evaluated in the linear case. The results obtained may be useful in practice in balancing precision against computational effort.

In section 6.1 the precision of the usual estimator of the variance of the stratified sample mean is evaluated. In sections 6.2-6.5 the alternative replication variance estimators are discussed and the efficiency of each estimator relative to the usual estimator is evaluated. (In section 6.2 a replication variance estimator based on randomly selected subsamples is considered. In sections 6.3-6.5 the idea of randomly splitting the sample in each stratum into two groups of equal size and then applying BRR or jackknife methods is explored.) A summary of the results obtained is provided in section 6.6.

6.1. Variance Estimation in the Linear Case

Suppose $\{y_{h1}, \dots, y_{hn_h}\}$ denotes a simple random sample of size n_h selected with replacement from the N_h units in stratum $h = 1, \dots, L$, (The set of all $\sum n_h$ units in the sample will be denoted by s .) Let $\bar{y}_h$ and $\bar{Y}_h$ denote the sample and population means respectively in stratum $h = 1, \dots, L$. The overall population mean is then $\bar{Y} = \sum W_h \bar{Y}_h$, where $W_h = N_h/N$ and $N = \sum N_h$.

Letting σ_h^2 denote the population variance in stratum h , the variance of $\bar{y}_{st}$ is

$$V(\bar{y}_{st}) = \sum W_h^2 \sigma_h^2 / n_h \quad (6.1)$$

which is estimated by

$$v(\bar{y}_{st}) = \sum W_h^2 s_h^2 / n_h, \quad (6.2)$$

where $s_h^2 = (n_h - 1)^{-1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$ provided $n_h \geq 2$

($h = 1, \dots, L$) . Noting that (see Wilks [1962, p.199])

$$E(s_h^2)^2 = \frac{\mu_{4,h}}{n_h} + \frac{(n_h^2 - 2n_h + 3)}{n_h(n_h - 1)} \sigma_h^4, \quad (6.3)$$

where $\mu_{4,h}$ denotes the fourth central population moment in stratum h , the variance of $v(\bar{y}_{st})$ may be expressed as

$$V(v(\bar{y}_{st})) = \sum \frac{W_h^4}{n_h} [E(s_h^2)^2 - \sigma_h^4]$$

$$= \sum \frac{(W_h \sigma_h)^4}{n_h^3} \left[\beta_h - \frac{(n_h - 3)}{(n_h - 1)} \right] \quad (6.4)$$

where $\beta_h = \mu_{4,h}/\sigma_h^4$ denotes the kurtosis in stratum h .

The efficiencies of the replication variance estimators described in subsequent sections of this chapter relative to $v(\bar{y}_{st})$ will be evaluated under the simplifying assumptions (McCarthy [1966], Lee [1972])

$$\left. \begin{aligned} W_h \sigma_h &= M \text{ (a constant),} \\ \beta_h &= 3 \text{ (as for the normal distribution), and} \\ n_h &= p \text{ (an even positive integer)} \end{aligned} \right\} \quad (6.5)$$

for $h = 1, \dots, L$. Under (6.5), (6.4) reduces to

$$V(v(\bar{y}_{st})) = \frac{2L M^4}{p^2(p-1)} \quad (6.6)$$

6.2. Randomly Selected Subsamples

When $n_h = 2$ ($h = 1, \dots, L$) McCarthy (1966) has considered a replication method of variance estimation in which the replicates are half-samples randomly selected (with replacement) from the set of all 2^L possible half-samples. This method is now extended to the case of arbitrary $\{n_h\}$.

Suppose that a subsample is formed by randomly selecting m_h ($1 \leq m_h < n_h$) units from the n_h sampled units in stratum $h = 1, \dots, L$. Suppose that k such partial samples are formed independently and let $\bar{y}_{hi}$ denote the mean of the m_h units in the h^{th} stratum in the i^{th} partial sample ($i = 1, \dots, k$). With adjusted stratum weights W'_h defined by

$$W'_h = \left(\frac{m_h}{n_h - m_h} \right)^{\frac{1}{2}} W_h, \quad (6.7)$$

let $\bar{y}'_{st} = \sum W'_h \bar{y}_h$ and $\bar{y}'_{st,i} = \sum W'_h \bar{y}_{hi}$ ($i = 1, \dots, k$). Then Theorem 6.1 below shows that

$$v_R(\bar{y}_{st}) = \frac{1}{k} \sum_{i=1}^k (\bar{y}'_{st,i} - \bar{y}'_{st})^2 \quad (6.8)$$

is an unbiased estimator of $V(v(\bar{y}_{st}))$. An expression for the variance of $v_R(\bar{y}_{st})$ is given in Theorem 6.2.

Theorem 6.1: $E(v_R(\bar{y}_{st})) = V(\bar{y}_{st})$.

Proof: Let $E|_s$ and $V|_s$ denote the expectation and the variance conditional on the sample s and let E and V denote the

expectation and the variance over all possible samples s . Then,
for $i = 1, \dots, k$,

$$\begin{aligned} E(\bar{y}'_{st,i} - \bar{y}'_{st})^2 &= E(W'_h)^2 E E|_s (\bar{y}_{hi} - \bar{y}_h)^2 \\ &= E(W'_h)^2 E V|_s (\bar{y}_{hi}) \\ &= E(W'_h)^2 E \left(\frac{n_h - m_h}{n_h} \right) \frac{s_h^2}{n_h} \\ &= E W_h^2 \frac{\sigma_h^2}{n_h} \\ &= V(\bar{y}_{st}) \quad /// \end{aligned} \quad (6.9)$$

Theorem 6.2: $V(v_R(\bar{y}_{st})) = V(v(\bar{y}_{st}))$

$$\begin{aligned} &+ \frac{4}{k} \sum_{h < h'} W_h^2 W_{h'}^2 \frac{\sigma_h^2}{n_h} \frac{\sigma_{h'}^2}{n_{h'}} \\ &+ \frac{1}{k} \sum_{h=1}^L W_h^4 E \left[\left(\frac{m_h}{n_h - m_h} \right) (\bar{y}_{hi} - \bar{y}_h)^2 - \frac{s_h^2}{n_h} \right]^2. \end{aligned}$$

Proof: The variance of $u_i = (\bar{y}'_{st,i} - \bar{y}'_{st})^2$ is first obtained using the result

$$V(u_i) = E V|_s u_i + V E|_s u_i. \quad (6.10)$$

The first term on the right side of (6.10) may be expressed

as

$$\begin{aligned}
 E V|_s u_i &= E V|_s (\bar{y}'_{st,i} - \bar{y}'_{st})^2 \\
 &= E E|_s \{ (\bar{y}'_{st,i} - \bar{y}'_{st})^2 - \sum_h W_h^2 \frac{s_h^2}{n_h} \}^2 \\
 &= E E|_s \left\{ \sum_h W_h^2 \left[\left(\frac{m_h}{n_h - m_h} \right) (\bar{y}_{hi} - \bar{y}_h)^2 - \frac{s_h^2}{n_h} \right] \right. \\
 &\quad \left. + 2 \sum_{h < h'} W_h' W_{h'}' (\bar{y}_{hi} - \bar{y}_h) (\bar{y}_{h'i} - \bar{y}_{h'}) \right\}^2 \\
 &= \sum_h W_h^4 E \left[\left(\frac{m_h}{n_h - m_h} \right) (\bar{y}_{hi} - \bar{y}_h)^2 - \frac{s_h^2}{n_h} \right]^2 \\
 &\quad + 4 \sum_{h < h'} W_h^2 W_{h'}^2 \frac{\sigma_h^2}{n_h} \frac{\sigma_{h'}^2}{n_{h'}}. \tag{6.11}
 \end{aligned}$$

The second term on the right side of (6.10) is given by

$$\begin{aligned}
 V E|_s u_i &= V E|_s (\bar{y}'_{st,i} - \bar{y}'_{st})^2 \\
 &= V \left(\sum_h W_h^2 \frac{s_h^2}{n_h} \right) \\
 &= V(v(\bar{y}_{st})) \tag{6.12}
 \end{aligned}$$

Replacing u_i by $\bar{u} = k^{-1} \sum u_i = v_R(\bar{y}_{st})$ in (6.10) and noting that, conditional on s , $u_1, \dots, u_k$ are independent identically distributed random variables yields

$$\begin{aligned}
 V(\bar{u}) &= E k^{-2} \sum_{i=1}^k V|_s u_i + V k^{-1} \sum_{i=1}^k E|_s u_i \\
 &= k^{-1} E V|_s u_i + V E|_s u_i \tag{6.13}
 \end{aligned}$$

The result now follows from (6.11) - (6.13). ///

Remark 6.1: When $n_h = 2$ and $m_h = 1$, $(\bar{y}_{hi} - \bar{y}_h) \equiv s_h^2/2$ so that $V(v_R(\bar{y}_{st}))$ reduces to

$$V(v_R(\bar{y}_{st})) = V(v(\bar{y}_{st})) + k^{-1} \sum_{h < h'} w_h^2 w_{h'}^2 \sigma_h^2 \sigma_{h'}^2, \quad (6.14)$$

as previously obtained by McCarthy (1966, p.15).

While it is clear from Theorem 6.2 that $v_R(\bar{y}_{st})$ is less efficient than $v(\bar{y}_{st})$, the following result is required in order to quantify this loss in efficiency. (The special case $n_h = 3$ is not required here as all efficiency comparisons are made under (6.5) with $n_h = p$, an even positive integer.)

Theorem 6.3: For $h = 1, \dots, L$ and $i = 1, \dots, k$

$$E \left[\left(\frac{m_h}{n_h - m_h} (\bar{y}_{hi} - \bar{y}_h)^2 - \frac{s_h^2}{n_h} \right)^2 \right] = \frac{\sigma_h^4 (8n_h - 3)}{n_h m_h (n_h - m_h)} - 4 \frac{\mu_{4,h}}{n_h^3} + 2 \frac{(n_h^2 + 4n_h - 6)}{n_h^3 (n_h - 1)} \sigma_h^4$$

provided $n_h \geq 4$ or $n_h = 2$.

Proof: For $h = 1, \dots, L$ and $i = 1, \dots, k$

$$\begin{aligned}
 & E_s E_s \left[\left(\frac{m_h}{n_h - m_h} \right) (\bar{y}_{hi} - \bar{y}_h)^2 - \frac{s_h^2}{n_h} \right]^2 \\
 &= E_s V_s \left(\frac{m_h}{n_h - m_h} \right) (\bar{y}_{hi} - \bar{y}_h)^2 \\
 &= \left(\frac{m_h}{n_h - m_h} \right)^2 E_s \{ E_s (\bar{y}_{hi} - \bar{y}_h)^4 - [E_s (\bar{y}_{hi} - \bar{y}_h)^2]^2 \} \\
 &= \left(\frac{m_h}{n_h - m_h} \right)^2 E_s E_s (\bar{y}_{hi} - \bar{y}_h)^4 - E_s \frac{(s_h^2)^2}{n_h} . \quad (6.15)
 \end{aligned}$$

Now, for $n_h \geq 4$, Kendall & Stuart (1969, p.302) have shown that

$$\begin{aligned}
 E_s (\bar{y}_{hi} - \bar{y}_h)^4 &= \{ \alpha_h^{(3)} - \frac{4}{n_h} \alpha_h^{(2)} + \frac{6}{n_h^2} \alpha_h^{(1)} - \frac{3(n_h-1)}{n_h(n_h+1)} (\alpha_h^{(1)})^2 \} K_h(4) \\
 &\quad + 3(\alpha_h^{(1)})^2 \frac{(n_h-1)}{(n_h+1)} (K_h(2))^2 \quad (6.16)
 \end{aligned}$$

where $\alpha_h^{(t)} = m_h^{-t} - n_h^{-t}$ ($t = 1, 2, 3$) and $K_h(2)$ and $K_h(4)$ are the second and fourth sample k-statistics (see Collins & Graham [1973]) in stratum h . Since (see Kendall & Stuart [1969, p.71, 301])

$$E_s K_h(4) = \kappa_h(4) = \mu_{4,h} - 3\sigma_h^4 \quad (6.17)$$

(where $\kappa_h(4)$ denotes the fourth population cumulant stratum h) and since $K_h(2) = s_h^2$, the result in the case $n_h \geq 4$ now follows from (6.15) - (6.17) and (6.3)..

When $n_h = 2$ and $m_h = 1$, the expectation to be evaluated is zero (see Remark 6.1). It is easily verified that the statement of the theorem remains valid in this case. ///

Remark 6.2: When $\beta_h = 3$ ($h = 1, \dots, L$) it follows from Theorems 6.2 and 6.3 that $V(v_R(\bar{y}_{st}))$ does not depend on the choice of m_h ($h = 1, \dots, L$). (When $n_h > 2$ and $\beta_h \neq 3$ the choice of m_h which minimizes $V(v(\bar{y}_{st}))$ depends on the sign of $(\beta_h - 3)$ and is easily calculated using Theorem 6.3.)

It now follows from (6.6) and Theorems 6.2 and 6.3 that

$$\frac{V(v_R(\bar{y}_{st}))}{V(v(\bar{y}_{st}))} = 1 + \frac{L(p-1)}{k} - \frac{1}{k} \quad (6.18)$$

under (6.5).

Remark 6.3: The covariance between $\bar{y}'_{st,i}$ and $\bar{y}'_{st,j}$ ($i, j = 1, \dots, k$) is given by

$$\begin{aligned} \text{Cov}(\bar{y}'_{st,i}, \bar{y}'_{st,j}) &= E(\bar{y}'_{st,i} \bar{y}'_{st,j}) - E(\bar{y}'_{st,i}) E(\bar{y}'_{st,j}) \\ &= E\{E_s(\bar{y}'_{st,i}) E_s(\bar{y}'_{st,j})\} - \{E_s E_s(\bar{y}'_{st,i})\} \{E_s E_s(\bar{y}'_{st,j})\} \\ &= E_s(\bar{y}'_{st})^2 - E_s^2(\bar{y}'_{st}) \\ &= V(\bar{y}'_{st}) \\ &= \sum W_h^2 \frac{m_h}{n_h} \frac{\sigma_h^2}{(n_h - m_h)}. \end{aligned} \quad (6.19)$$

Since $V(\bar{y}'_{st,i}) = \sum W_h^2 \sigma_h^2 / (n_h - m_h)$ it follows from (6.19) that when $m_h/n_h = \rho$ ($h = 1, \dots, L$), the correlation between $\bar{y}'_{st,i}$ and $\bar{y}'_{st,j}$

is equal to p . This result is of use in adapting the sign test to the correlated random variables $\bar{y}_{st,1}, \dots, \bar{y}_{st,k}$ as proposed by McCarthy (1966, 1969a). Application of the sign test is particularly simple when $p = 1/2$ since David & Six (1971) have shown that (for a certain class of symmetric distributions) the probability is exactly $(u+1)/(k+1)$ that at most u of the k estimates $\bar{y}_{st,1}, \dots, \bar{y}_{st,k}$ are greater than (or less than) the median of the distribution of the $\bar{y}_{st,1}$.

6.3. Balanced Half-Samples

When the number of units selected in each stratum is even, Kish & Frankel (1970, p.1087-8) have suggested randomly dividing the n_h units in stratum h into two groups of size $n_h = n_h/2$ ($h = 1, \dots, L$) and then applying BRR (section 4.5) to the pairs thus formed. This modification of the BRR method has recently been employed by Casady (1975) and Lemeshow (1976).

As noted previously (Remark 4.7) a set of k_B balanced half-samples may always be constructed using the methods of Plackett & Burman (1946), where $L \leq k_B \leq L+3$. Letting $\bar{y}_{st}^{-1}$ denote the stratified sample mean based on the i^{th} half-sample ($i = 1, \dots, k_B$) Kish & Frankel's variance estimator is given by

$$v_B(\bar{y}_{st}) = \frac{1}{k_B} \sum_{i=1}^{k_B} (\bar{y}_{st}^{-1} - \bar{y}_{st})^2 \quad (6.20)$$

In evaluating the expected value and variance of this estimator, it will be convenient to work with a different form of $v_B(\bar{y}_{st})$ than that expressed in (6.20). The set of balanced half-samples used may be defined by a $k_B \times L$ matrix $((\delta_{ih}))$ such that for $i = 1, \dots, k_B$ and $h = 1, \dots, L$

$$\delta_{ih} = \begin{cases} +1 & \text{if group 1 in the } h^{\text{th}} \text{ stratum is in the } i^{\text{th}} \text{ half-sample} \\ -1 & \text{if group 2 in the } h^{\text{th}} \text{ stratum is in the } i^{\text{th}} \text{ half-sample} \end{cases}$$

and such that for $h \neq h' = 1, \dots, L$

$$\sum_{i=1}^{k_B} \delta_{ih} \delta_{ih'} = 0 \quad (6.21)$$

Letting $d_h = \bar{y}_h^{(1)} - \bar{y}_h^{(2)} = (\bar{y}_h^{(1)} - \bar{y}_h)$, where $\bar{y}_h^{(1)} = m_h^{-1} \sum_{i=1}^{m_h} y_{hi}$ and $\bar{y}_h^{(2)} = m_h^{-1} \sum_{i=m_h+1}^{n_h} y_{hi}$ denote the means of the two groups in stratum $h = 1, \dots, L$, and noting that

$$\bar{y}_{st}^{(1)} - \bar{y}_{st}^{(2)} = \sum_{h=1}^L W_h \delta_{ih} d_h \quad (6.22)$$

for $i = 1, \dots, k_B$, it follows from (6.20) - (6.22) that

$$v_B(\bar{y}_{st}) = \sum W_h^2 d_h^2 \quad (6.23)$$

As in (6.9), it follows that $E(d_h^2) = \sigma_h^2/n_h$ so that

$E(v_B(\bar{y}_{st})) = V(v(\bar{y}_{st}))$. An expression for the variance of $v_B(\bar{y}_{st})$ is given in Theorem 6.4.

Theorem 6.4: $V(v_B(\bar{y}_{st})) = V(v(\bar{y}_{st})) + \sum W_h^4 E\left[d_h^2 - \frac{\sigma_h^2}{n_h}\right]^2$.

Proof: From (6.23)

$$\begin{aligned} V(v_B(\bar{y}_{st})) &= V\left(\sum W_h^2 d_h^2\right) \\ &= \sum W_h^4 V(d_h^2) \\ &= \sum W_h^4 E\left[d_h^2 - \frac{\sigma_h^2}{n_h}\right]^2 \\ &= \sum W_h^4 E\left[d_h^2 - \frac{\sigma_h^2}{n_h}\right]^2 + \sum W_h^4 E\left[\frac{\sigma_h^2}{n_h} - \frac{\sigma_h^2}{n_h}\right]^2 \\ &= \sum W_h^4 E\left[d_h^2 - \frac{\sigma_h^2}{n_h}\right]^2 + V(v(\bar{y}_{st})) \quad /// \quad (6.24) \end{aligned}$$

3

OF / DE

3

It now follows from (6.6) and Theorems 6.3 and 6.4 that

$$\frac{V(v_B(\bar{y}_{st}))}{V(v(\bar{y}_{st}))} \leq p - 1 \quad (6.25)$$

under (6.5). When $p = 2$, $v_B(\bar{y}_{st})$ is of course fully efficient in the linear case.

It follows from (6.25) and (6.18) that

$$\frac{V(v_R(\bar{y}_{st}))}{V(v_B(\bar{y}_{st}))} = \frac{1}{k} \left[L + \frac{k-1}{p-1} \right] \quad (6.26)$$

under (6.5). When $k = k_B = L$ (recall that $L \leq k_B \leq L+3$), it follows from (6.26) that $v_B(\bar{y}_{st})$ is more precise than $v_R(\bar{y}_{st})$. When $k = k_B = L+t$ ($t = 1, 2, 3$), $v_B(\bar{y}_{st})$ is more precise than $v_R(\bar{y}_{st})$ if and only if

$$p < \frac{(L+t-1)}{t} + 1. \quad (6.27)$$

When $k=L$ is large

$$\frac{V(v_R(\bar{y}_{st}))}{V(v_B(\bar{y}_{st}))} \approx \frac{p}{p-1}. \quad (6.28)$$

6.4. Partially Balanced Half-Samples

While the use of $v_B(\bar{y}_{st})$ requires at most $L+3$ half-samples, even this amount of replication may be prohibitive when L is large. In this case the use of a set of "partially balanced" half-samples has been considered by McCarthy (1966) and Lee (1972, 1973b) as a means of reducing the number of half-samples required.

To construct a set of partially balanced half-samples, the L strata are first divided into g groups ($2 \leq g \leq L/2$) with L/g strata in each group. A partially balanced set of k_{PB} half-samples is then obtained by constructing a set of k_{PB} balanced half-samples ($L/g \leq k_{PB} \leq (L/g)+3$) in each group (using the same construction in each group). The set of partially balanced half-samples may again be defined by a $k_{PB} \times L$ matrix $((\delta_{ih}))$ where

$$\delta_{ih} = \begin{cases} +1 & \text{if group } i \text{ in the } h^{\text{th}} \text{ stratum is in the } i^{\text{th}} \text{ half-sample} \\ -1 & \text{if group } i \text{ in the } h^{\text{th}} \text{ stratum is not in the } i^{\text{th}} \text{ half-sample.} \end{cases}$$

The matrix $((\delta_{ih}))$ satisfies

$$\sum_{i=1}^{k_{PB}} \delta_{ih} \delta_{ih'} = 0 \quad (6.29)$$

whenever $h \neq h'$ belong to the same group or do not correspond to the same column in two different groups. For example when $L = 4$ and $g = 2$ the matrix

$$((\delta_{ih})) = \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \end{pmatrix} \quad (6.30)$$

defines a set of $k_{PB} = 2$ partially balanced half-samples.

As in (6.20), the variance estimator based on a set of k_{PB} partially balanced half-samples is

$$v_{PB}(\bar{y}_{st}) = \frac{1}{k_{PB}} \sum_{i=1}^{k_{PB}} (\bar{y}_{st}^i - \bar{y}_{st})^2 \quad (6.31)$$

It now follows from (6.29), (6.31) and (6.22) that

$$v_{PB}(\bar{y}_{st}) = \sum W_h^2 d_h^2 + 2 \sum^* W_h W_{h'} d_h d_{h'} \quad (6.32)$$

where $\sum^*$ denotes summation over all $h < h'$ such that h and h' are from different groups of L/g strata and h and h' represent corresponding columns in the two groups. Since $E(d_h d_{h'}) = E(d_h) E(d_{h'}) = 0$ it follows from (6.32) that $E(v_{PB}(\bar{y}_{st})) = V(v_{PB}(\bar{y}_{st}))$. An expression for the variance of $v_{PB}(\bar{y}_{st})$ is given in Theorem 6.5.

Theorem 6.5: $V(v_{PB}(\bar{y}_{st})) = V(v_B(\bar{y}_{st})) + 4 \sum^* W_h^2 W_{h'}^2 \frac{\sigma_h^2 \sigma_{h'}^2}{n_h n_{h'}}.$

Proof: From (6.32)

$$V(v_{PB}(\bar{y}_{st})) = V(\sum W_h^2 d_h^2) + 4 \sum^* W_h^2 W_{h'}^2 V(d_h d_{h'}) \quad (6.33)$$

Since d_h and $d_{h'}$ are independent with $E(d_h) = E(d_{h'}) = 0$.

$$\begin{aligned} V(d_{h,h'}) &= E(d_h^2) E(d_{h'}^2) \\ &= \frac{\sigma_h^2}{n_h} \frac{\sigma_{h'}^2}{n_{h'}} \end{aligned} \quad (6.34)$$

The result now follows from (6.33), (6.34) and (6.24). ///

It is clear from Theorem 6.5 that $v_{PB}(\bar{y}_{st})$ is less precise than $v_B(\bar{y}_{st})$ and that the precision of $v_{PB}(\bar{y}_{st})$ depends on the arrangement of strata into groups as well as on the value of g . (Note that there are $(L/g)_2 C_2 = L(g-1)/2$ terms in the summation Σ^* .) When $n_h = 2$ and the value of $W_h \sigma_h$ is known ($h = 1, \dots, L$), Lee (1972, 1973b) has considered several methods of arranging the strata into groups in order to minimize the variance of $v_{PB}(\bar{y}_{st})$.

It follows from (6.6) and Theorems 6.3 and 6.5 that

$$\frac{V(v_{PB}(\bar{y}_{st}))}{V(v(\bar{y}_{st}))} = g(p-1) \quad (6.35)$$

under (6.5). When $g = 1$, the set of half-samples is balanced and (6.35) reduces to (6.25). When $p = 2$, $V(v_{PB}(\bar{y}_{st}))/V(v(\bar{y}_{st})) = g$ as previously obtained by Lee (1972, p.327). Note that while the precision of $v_{PB}(\bar{y}_{st})$ increases as g decreases, the number of partially balanced half-samples required also increases as g decreases.

It follows from (6.35) and (6.18) that

$$\frac{V(v_R(\bar{y}_{st}))}{V(v_{PB}(\bar{y}_{st}))} = \frac{1}{gk} \left[L + \frac{k-1}{p-1} \right] \quad (6.36)$$

under (6.5). When $k = k_{PB} = L/g$ (recall that $L/g \leq k_{PB} \leq (L/g)+3$), it follows from (6.36) that $v_{PB}(\bar{y}_{st})$ is more precise than $v_R(\bar{y}_{st})$. When $k = k_{PB} + t$ ($t = 1, 2, 3$), $v_{PB}(\bar{y}_{st})$ is more precise than $v_R(\bar{y}_{st})$ if and only if

$$p < \frac{(L + g(t-1))}{g^2 t} + 1. \quad (6.37)$$

When L is large and $k = L/g$

$$\frac{V(v_R(\bar{y}_{st}))}{V(v_{PB}(\bar{y}_{st}))} \doteq 1 + \frac{1}{g(p-1)}. \quad (6.38)$$

6.5. Jackknife Methods

When the n_h units in stratum h are randomly divided into two groups each of size $m_h = n_h/2$ ($h = 1, \dots, L$) as in sections 6.3 and 6.4, the jackknife method may also be applied using the artificially created pairs. In this case the jackknife variance estimator is

$$v_J(\bar{y}_{st}) = \sum_{h=1}^L \frac{1}{2} \sum_{i=1}^2 (\bar{y}_{st}^{hi} - \bar{y}_{st})^2 \quad (6.39)$$

where $\bar{y}_{st}^{hi}$ denotes the stratified sample mean computed from the sample after deleting the i^{th} group in the h^{th} stratum ($i = 1, 2$; $h = 1, \dots, L$) . .

Since $(\bar{y}_{st}^{hi} - \bar{y}_{st})^2 = W_h^2 d_h^2$ it follows from (6.39) that

$$\begin{aligned} v_J(\bar{y}_{st}) &= \sum W_h^2 d_h^2 \\ &= v_B(\bar{y}_{st}) \end{aligned} \quad (6.40)$$

From (6.40) and (6.26), $v_J(\bar{y}_{st})$ is less efficient than $v_R(\bar{y}_{st})$ with $k = 2L$ if $p > 2$.

Since the use of $v_J(\bar{y}_{st})$ involves the use of $2L$ replicates the alternative jackknife variance estimator

$$v'_J(\bar{y}_{st}) = \sum_{h=1}^L (\bar{y}_{st}^{hi} - \bar{y}_{st})^2 \quad (6.41)$$

requiring only L replicates may be considered. As in (6.40),

$v_J'(\bar{y}_{st}) = v_B(\bar{y}_{st})$. From (6.40) and (6.26), $v_J'(\bar{y}_{st})$ is more efficient than $v_R(\bar{y}_{st})$ with $k = L$.

Remark 6.4: When $n_h = 2$ ($h = 1, \dots, L$), v_J and v_J' correspond respectively to Frankel's (1971) JRR-S and JRR-H jackknife variance estimators (see section 4.4).

6.6. Summary

Since the application of the BRR and jackknife methods of variance estimation may not be computationally feasible when the number of replicates involved is large, the use of several replication variance estimators requiring a reduced number of replicates is considered.

The stability of the variance estimator v_R based on k randomly selected subsamples is evaluated under arbitrary allocation in the linear case. While v_R is not fully efficient, any number of replicates may be used (although the stability of v_R may be unsatisfactory for small k).

When the number of units selected in each stratum is even, they may be randomly divided into two halves and BRR applied to the pairs thus formed as proposed by Kish & Frankel (1970). (Recall that the number of half-samples required to form a balanced set is approximately equal to the number of strata.) While this BRR variance estimator v_B is fully efficient when only two units are selected in each stratum, the stability of v_B decreases markedly as the sample size within each stratum increases. (Since the jackknife variance estimators v_J and v_{J^*} are algebraically equal to v_B in the linear case, the same conclusions hold for the jackknife.) While the variance estimator v_{PB} based on a partially balanced set of half-samples may be used with fewer replicates than v_B ,

the estimator v_{PB} lags behind v_B considerably in terms of stability. For the same number of replicates, v_R is generally less stable than v_B , v_{PB} and v_J .

CHAPTER 7

CONCLUDING REMARKS

7.1. Brief Summary

Analytic properties of the linearization, jackknife and balanced repeated replication (BRR) methods are investigated with results for both large and small samples presented. While some results are obtained under the assumption of simple random sampling, the main results of this research pertain to stratified samples.

The large sample theory presented here provides a basis for the construction of asymptotically robust tests of significance and confidence intervals through the use of an appropriate pivotal statistic. The results for simple random sampling are developed for the case of sampling without replacement (Chapter 2) while those for stratified samples are obtained as the number of strata $L \rightarrow \infty$ (Chapter 4). Estimators of the population parameter of interest obtained using the jackknife and BRR methods are shown to be asymptotically normally distributed with the same variance as the unjackknifed statistic. Variance estimators obtained using all three methods are shown to be consistent.

In view of the fact that the above three methods have similar asymptotic properties, the small sample results for ratio estimation may be more useful in assessing the relative merits of

these three methods. With simple random sampling (Chapter 3), the jackknife variance estimator is found to be somewhat less stable than the two commonly used linearization variance estimators, except under conditions particularly favorable to ratio estimation (i.e., regression through the origin and a small coefficient of variation in the x population). In stratified samples (Chapter 5), the linearization variance estimator tends to underestimate the mean square error of the combined ratio estimator while the BRR variance estimators tend to be overestimates. The jackknife variance estimators are also underestimates in some situations. A particular BRR variance estimator is found to be somewhat less stable than the linearization variance estimator in stratified samples, except under conditions favorable to ratio estimation.

Since the use of the jackknife and BRR methods of variance estimation can become both cumbersome and expensive when the number of replicates involved is large, the use of several replication variance estimators requiring a reduced number of replicates is explored (Chapter 7). The loss in efficiency suffered by these estimators is substantial.

7.2. Further Research

While it is hoped that the results of this research will be of use in assessing the relative merits of the linearization, jackknife and BRR methods, many areas remain in which further research might be profitably undertaken. Two such areas are now discussed.

The empirical investigations conducted by Frankel (1971) and Bean (1975) have revealed only small differences in the stabilities of the variance estimators obtained using the above three methods in stratified samples (although the variability is consistently lowest for the linearization method and highest for BRR). The small sample results for ratio estimation presented here, however, indicate that the linearization variance estimator is considerably more stable than the BRR variance estimator in situations not so favorable to ratio estimation. Further study of the stabilities of the variance estimators obtained using all three methods in a variety of situations is warranted.

The asymptotic results for stratified samples reported here are obtained as the number of strata tends to infinity. This approach could be used in developing large sample theory for chi-square statistics or other measures of association (Nathan [1973], Shuster & Downing [1976]) for stratified samples.

REFERENCES

- Arvesen, J. N. (1969). Jackknifing U-statistics. Ann. Math. Statist., 40, 2076-100.
- Arvesen, J. N. & Layard, M. W. J. (1975). Asymptotically robust tests in unbalanced variance component models. Ann. Statist., 3, 1122-34.
- Barton, D. E. & David, F. N. (1961). The central sampling moments of the mean in samples from a finite population (Aly's formulae and Madow's central limit). Biometrika, 48, 199-201.
- Bean, J. A. (1975). Distribution and properties of variance estimators for complex multistage probability samples. Vital and Health Statistics, Series 2, No. 65. Washington, D.C.: U.S. Government Printing Office.
- Bissel, A. E. & Ferguson, R. A. (1975). The jackknife - toy, tool or two-edged weapon? Statistician, 24, 79-100.
- Borack, J. I. (1971). A General Theory of Balanced 1/N Sampling. Unpublished Ph.D. thesis, Cornell University.
- Brewer, K. R. W. & Mellor, R. W. (1973). The effect of sample structure on analytical surveys. Aust. J. Statist., 15, 145-52.
- Casady, R. J. (1975). The estimation of variance components using balanced repeated replication. Proc. Amer. Statist. Ass. (Social Statistics Section), 352-7.
- Chakrabarty, R. P. (1968). Contributions to the Theory of Ratio-type Estimators. Unpublished Ph.D. thesis, Texas A & M University.
- Chung, K. L. (1974). A Course in Probability Theory. (Second edition). New York: Academic Press.
- Cochran, W. G. (1963). Sampling Techniques. (Second edition). New York: John Wiley & Sons.
- Cochran, W. G. (1975). Two recent areas of sample survey research. A Survey of Statistical Design and Linear Models (J. N. Srivastava, Ed.). New York: North-Holland. 101-15.
- Collins, B. T. & Graham, J. E. (1973). Theory of polykay statistics with applications to survey sampling. Carleton University Mathematical Series, No. 85., Dept. of Math. Carleton Univ. Ottawa.

David, H. A. & Six, F. B. (1971). Sign distribution of standard multinormal variables with equal positive correlation. Rev. Int. Statist. Inst., 39, 1-3.

Deming, W. E. (1956). On simplification of sampling design through replication with equal probabilities and without stages. J. Amer. Statist. Ass., 51, 24-53.

Deming, W. E. (1960). Sample Design in Business Research. New York: John Wiley & Sons, Inc.

Duncan, G. T. & Layard, M. W. J. (1973). A Monte-Carlo study of asymptotically robust tests for correlation coefficients. Biometrika, 60, 551-8.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. Biometrika, 46, 477-80.

Erdős, P. & Rényi, P. (1959). On a central limit theorem for samples from a finite population. Acta Math. Acad. Sci. Hung., 4, 49-61.

Frankel, M. R. (1971). An Empirical Investigation of Some Properties of Multivariate Statistical Estimates from Complex Samples. Ph.D. thesis, University of Michigan.

Fuller, W. A. (1975). Regression analysis for sample surveys. Sankhya, Ser. C., To appear.

Gnedenko, B. V. (1968). Theory of Probability (Fourth edition). New York: Chelsea Publishing Co.

Gray, H. L. & Shucany, W. R. (1972). The Generalized Jackknife Statistic. New York: Marcel Dekker, Inc.

Gurney, M. & Jewett, R. S. (1975). Constructing orthogonal replications for variance estimation. J. Amer. Statist. Ass., 70, 819-21.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. Acta Math. Acad. Sci. Hung., 5, 361-74.

Hartley, H. O. & Ross, A. (1954). Unbiased ratio estimators. Nature, 174, 270-1.

Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. Ann. Math. Statist., 40, 1977-91.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. Ann. Math. Statist., 19, 293-325.

Hutchison, M. C. (1971). A Monte Carlo comparison of some ratio estimators. Biometrika, 58, 313-21.

Jones, H. L. (1964). The jackknife method. Proc. IBM Sci. Computing Symp. Statist., New York: IBM Data Processing Division. 185-97.

Jones, H. L. (1974). Jackknife estimation of functions of stratum means. Biometrika, 61, 343-8.

Kendall, M. G. & Stuart, A. (1969). The Advanced Theory of Statistics, Vol. I, Second Edition. New York: Hafner Publishing Co.

Kish, L. (1965). Survey Sampling. New York: John Wiley & Sons.

Kish, L. & Frankel, M. R. (1970). Balanced repeated replications for standard errors. J. Amer. Statist. Ass., 65, 1071-94.

Kish, L. & Frankel, M. R. (1974). Inference from complex samples (with discussion). J. R. Statist. Soc., Ser. B., 36, 1-37.

Kish, L., Namboodiri, N. K. & Pillai, R. K. (1962). The ratio bias in surveys. J. Amer. Statist. Ass., 57, 863-76.

Lauh, E. & Williams, W. H. (1963). Some small sample results for the variance of a ratio. Proc. Amer. Statist. Ass. (Social Statistics Section), 273-83.

Lee, K. H. (1971). Variance Estimation for Survey Statistics Using Techniques Based on Partial Samples. Unpublished Ph.D. thesis, University of London.

Lee, K. H. (1972). Partially balanced designs for half sample replication method of variance estimation. J. Amer. Statist. Ass., 67, 324-34.

Lee, K. H. (1973a). Variance estimation in stratified sampling. J. Amer. Statist. Ass., 68, 336-42.

Lee, K. H. (1973b). Using partially balanced designs for the half sample replication method of variance estimation. J. Amer. Statist. Ass., 68, 612-14.

- Lemeshow, S. (1976). The use of unique statistical weights for estimating variances with the balanced half-sample technique. Proc. Amer. Statist. Ass. (Social Statistics Section), In press.
- Madow, W. G. (1948). On the limiting distribution of estimates based on samples from finite universes. Ann. Math. Statist., 19, 535-45.
- Mahalanobis, P. C. (1944). On large scale sample surveys. Phil. Trans. Roy. Soc., B231, 329-451.
- McCarthy, P. J. (1966). Replication: an approach to the analysis of data from complex surveys. Vital and Health Statistics, Series 2, No. 14. Washington, D.C.: U.S. Government Printing Office.
- McCarthy, P. J. (1969a). Pseudoreplication: further evaluation and application of the balanced half-sample technique. Vital and Health Statistics, Series 2, No. 31. Washington, D.C.: U.S. Government Printing Office.
- McCarthy, P. J. (1969b). Pseudoreplication: half-samples. Rev. Int. Statist. Inst., 37, 239-64.
- Mickey, M. R. (1959). Some finite population unbiased ratio and regression estimators. J. Amer. Statist. Ass., 54, 594-612.
- Miller, R. G. (1964). A trustworthy jackknife. Ann. Math. Statist., 35, 1594-1605.
- Miller, R. G. (1968). Jackknifing variances. Ann. Math. Statist., 39, 567-82.
- Miller, R. G. (1974a). The jackknife - a review. Biometrika, 61, 1-15.
- Miller, R. G. (1974b). An unbalanced jackknife. Ann. Statist., 2, 880-91.
- Mosteller, F. & Tukey, J. W. (1966). Data analysis, including statistics. Handbook of Social Psychology (G. Lindzey & E. Aronson, eds.). Reading, Mass.: Addison-Wesley. 134-60.
- Nandi, H. K. & Sen, P.K. (1963). On the properties of U-statistics when the observations are not independent. Part two: unbiased estimation of the parameters of a finite population. Calcutta Statist. Ass. Bull., 12, 124-48.

- Nathan, G. (1973). Approximate tests of independence in contingency tables from complex stratified cluster samples. Vital and Health Statistics, Series 2, no. 53. Washington, D.C.: U.S. Government Printing Office.
- Nutrition Canada (1973). Nutrition: A National Priority. A report by Nutrition Canada to Health & Welfare Canada. Ottawa, Health & Welfare Canada.
- Philippou, A. N. (1973). Exponential approximation of the likelihood function and hypothesis testing in the independent not identically distributed case. Commun. Statist., 2, 401-18.
- Plackett, R. L. & Burman, J. P. (1946). The design of optimum multifactorial experiments. Biometrika, 23, 305-25.
- Quenouille, M. H. (1949). Approximate tests of correlation in time-series. J. R. Statist. Soc., Ser. B., 11, 68-84.
- Quenouille, M. H. (1956). Notes on bias in estimation. Biometrika, 43, 353-60.
- Raj, D. (1968). Sampling Theory. New York: McGraw-Hill.
- Rao, C. R. (1973). Linear Statistical Inference and its Applications. (Second edition). New York: John Wiley & Sons.
- Rao, J. N. K. (1965). A note on the estimation of ratios by Quenouille's method. Biometrika, 52, 647-9.
- Rao, J. N. K. (1967). The precision of Mickey's unbiased ratio estimator. Biometrika, 54, 321-4.
- Rao, J. N. K. (1969). Ratio and regression estimators. New Developments in Survey Sampling (N. L. Johnson & H. Smith, Jr., eds.). New York: Wiley-Interscience. 213-34.
- Rao, J. N. K. (1975). Analytic studies of sample survey data. Survey Methodology, Supplementary issue. Ottawa: Statistics Canada.
- Rao, J. N. K. & Beegle, L. D. (1967). A Monte Carlo study of some ratio estimators. Sankhyā, Ser. B., 29, 47-56.
- Rao, J. N. K. & Kuzik, R. A. (1974). Sampling errors in ratio estimation. Sankhyā, Ser. C., 36, 43-58.
- Rao, J. N. K. & Ramachandran, V. (1974). Comparison of the separate and combined ratio estimators. Sankhyā, Ser. C., 36, 151-6.

- Rao, J. N. K. & Webster, J. T. (1966). On two methods of bias reduction in the estimation of ratios. Biometrika, 53, 571-7.
- Rao, P. S. R. S. (1969). Comparison of four ratio-type estimators under a model. J. Amer. Statist. Ass., 64, 574-80.
- Rao, P. S. R. S. (1974). Jackknifing the ratio estimator. Sankhyā, Ser. C., 36, 84-97.
- Rao, P. S. R. S. & Rao, J. N. K. (1971). Small sample results for ratio estimators. Biometrika, 58, 625-30.
- Rudin, W. (1966). Real and Complex Analysis. New York: McGraw-Hill.
- Sen, P. K. (1960). On some convergence properties of U-statistics. Calcutta Statist. Assoc. Bull., 10, 1-18.
- Sen, P. K. (1970). On some convergence properties of one-sample rank order statistics. Ann. Statist., 41, 2140-3.
- Sen, P. K. (1972). Finite population sampling and weak convergence to a Brownian bridge. Sankhyā, Ser. A., 34, 85-90.
- Sharot, T. (1976a). The generalized jackknife: finite samples and subsample sizes. J. Amer. Statist. Ass., 71, 451-4.
- Sharot, T. (1976b). Sharpening the jackknife. Biometrika, 63, 315-21.
- Shucany, W. R., Gray, H. L. & Owen, D. B. (1971). On bias reduction in estimation. J. Amer. Statist. Ass., 66, 524-33.
- Shuster, J. J. & Downing, D. J. (1976). Two-way contingency tables for complex sampling schemes. Biometrika, 63, 271-6.
- Tepping, B. J. (1968). The estimation of variance in complex surveys. Proc. Amer. Statist. Ass. (Social Statistics Section), 11-8.
- Thorburn, D. (1976). Some asymptotic properties of jackknife statistics. Biometrika, 63, 305-13.
- Tin, M. (1965). Comparison of some ratio estimators. J. Amer. Statist. Ass., 60, 294-307.
- Tukey, J. W. (1958). Bias and confidence in not-quite so large samples (Abstract). Ann. Math. Statist., 29, 614.

Wilks, S. S. (1962). Mathematical Statistics. New York: John Wiley & Sons, Inc.

Woodruff, R. S. (1971). Simple method for approximating the variance of a complex estimate. J. Amer. Statist. Ass., 66, 411-14.

Woodruff, R. S. & Causey, B. D. (1976). Computerized method for approximating the variance of a complex estimate. J. Amer. Statist. Ass., 71, 315-21.

END

07

11

77

FIN