University of Southampton

FACULTY OF MATHEMATICAL STUDIES

JACKKNIFE PROCEDURES FOR SAMPLE SURVEYS

by

Nazir Ahmed Chaudhry

Thesis submitted for the degree of Master of Philosophy

1985

In the memory of my Father (1981)

and

To Rifa'at, Asim, Sumeera, Arshia,
Adnan and my Mother.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MATHEMATICAL STUDIES

MATHEMATICS

Master of Philosophy

JACK-KNIFE PROCEDURES FOR SAMPLE SURVEYS

by Nazir Ahmed Chaudhry

A great deal of current research into the foundations of survey sampling has been devoted to the application of various inference criteria. A distinction is drawn between the randomisation based and the model based approaches to inference for finite populations. It is argued that the model based approach is effective and offers inferential benefits over the randomisation based approach for a realistic model.

Methods of variance estimation under both the randomisation and model based approaches are considered. The randomisation based approach does provide a variety of methods for variance estimation. The randomisation based variances are unconditional and do not lead to a meaningful inference from an observed sample. The model based variances are conditional and provide a suitable framework for inference. The ratio estimator provides an important example of statistic for which the randomisation and the model based approaches lead to different variance estimators.

Problems concerning the estimation of correlation from complex surveys are discussed. Implications of the complexity of the design on the behaviour of simple, partial and multiple correlations are studied. It is shown that the complexity of the design affects the behaviour of the simple, partial and multiple correlations. The relative biases for the multiple correlations are, in general, smaller than those for the simple and partial correlations when the design variable is strongly related to one of the x variables under study. It is argued that different surveys with different stratifications may produce quite different results.

The jack-knife method finds applications in a variety of fields both as a means of bias reduction and variance estimation. A comprehensive review of the applications of the jack-knife method to ratio estimation and the linear regression model is given. A distinction is drawn between the unweighted and the weighted jack-knife procedures for regression estimation. A simple generalisation of the weighted jack-knife for regression estimation is given. It is shown that under the model with more general variance structure, the weighted jack-knife regression estimator is unbiased and the weighted jack-knifed variance estimator is biased to $O(n^{-1})$.

Using the idea of empirical influence function, the weighted jack-knife technique is extended to ratio estimation. A weighted jack-knifed variance estimator for the ratio estimator is developed. Using the prediction theory approach, the properties of the weighted jack-knifed variance estimator are examined. The implications of the failures of regression model on the behaviour of the weighted jack-knifed variance estimator are studied. Comparison of the bias reduction properties of the weighted and the unweighted jack-knifed ratio estimators is made using Taylor Series expansion.

ACKNOWLDGEMENTS

I express my gratitude to my supervisor, Professor T. M. F. Smith, for his guidance and encouragement throughout the course of this study.

I am indebted to Sind Agriculture University Tandojam, in Pakistan, for financial support during the first three years; Dr. M. Yousaf Chaudhry for his generous financial support during the last year and to Dr. Wali Muhammad, Southdown and Gunter Trusts for financial assistance in the final phase of this research.

I am grateful to Dr. A. T. R. Fuller of the University Health Centre and Mr. A. R. Elkington of the Southampton Eye Hospital for their sincere efforts in keeping my troubled eyes working during the write-up of this thesis.

I also wish to express my thanks to Hazel Paul for patiently and competently typing a difficult manuscript and Dr. M. Yar for his invaluable help in reading the typed script and the arrangement of references.

My wife, Rifaat, has earned my sincere gratitude for being patient and most understanding during the periods of financial crisis. I should like to thank my four sweet children, Asim, Sumeera, Arshia and Adnan for all their innocent activities throughout the period of my research studies and our stay in Southampton. I should also like to thank all the members of our family in Pakistan and England as well as our friends and well-wishers for their patience, love and friendship, which was greatly appreciated and needed - through the completion of this work.

Finally, for the strength, patience and peace that only HE can give, I thank God, WHO is the source of all wisdom and strength.

CONTENTS

Page No.

CHAPTER 1

INTRODUCTION

There has been a rapid growth in sample surveys during the past four decades and a considerable amount of literature regarding the design and analysis of survey data is now available. The essence of survey sampling consists of the selection of a part of a finite population and on the basis of this selected part making inferences about the entire population. Thus the main objective in any sample survey is to compute the sample estimates so that inferences could be drawn about the finite population under investigation.

The theory of inference from finite populations has seen controversial developments during the last three decades. As a consequence of these developments a model based approach to inference has emerged as an alternative to the traditional randomisation based approach. In the randomisation based approach the only source of probabilities is the artificial randomisation induced by the sample design. Other sources of probabilities such as superpopulation models are also prevalent but only as evaluation techniques with the probability sampling models for special purposes. The basic criterion and definitions such as bias and standard errors refer to probability sampling models. In contrast the model based approach to survey sampling takes the form of assuming a superpopulation model with the characteristics of the units of the finite population under study being the realisations of the assumed super-population model. The sampling plan need not be a probability sampling plan and the estimators need not be randomisation

consistent. The model based inference concerns the prediction of the actual outcome of the population characteristics and is made conditional on the outcome of the sampling. Thus from the view point of statistical inference the most interesting aspect of these developments in the field of sample survey is that by a shift to the model based approach a primary object of sample surveys is turned from estimation to prediction. The model based approach is one avenue through which important new methods are currently being added to the existing survey sampling literature. The model based approach is used extensively in this thesis.

Many sample surveys today are characterised by complex clustered and stratified samples and often depart drastically from a simple random sample. This raises important questions about how far standard statistical procedures based implicitly on simple random sampling assumptions are appropriate when a population has complex structure. Pioneer work for estimating analytical statistics such as correlation and regression coefficients from complex surveys is due to Kish and Frankel (1974). These authors studied the behaviour of simple, partial and multiple correlation from clustered stratified samples. Kish and Frankel concentrated mainly on the effect of clustering and ignored the effects of stratification. We have studied the effect of stratification on the behaviour of correlations.

During the last few years a great deal of attention has been focussed on the development of robust statistical methods that do not depend upon the assumptions of normality but may retain many desirable features of this assumption. A class of methods known as jack-knife techniques were introduced by Quenouille (1949, 1956) as a bias reducing device

and later developed for variance estimation and setting up confidence intervals. The essence of the technique is that pseudovalues of the estimated parameters are computed by deleting a single observation or a set of observations and the estimate of the variance of the estimated parameters is obtained from the variance of the pseudovalues.

The main subject of this thesis concentrates on the applications of the jack-knife method to ratio estimation. A considerable amount of literature on the jack-knife technique as a method of bias reduction and robust interval estimation is available in balanced situations (Chapter 5). However, much still remains to be investigated in unbalanced situations. Hinkley (1977b) proposed the weighted jack-knife for regression estimation. We extend the application of the weighted jack-knife to ratio estimation. We also study the properties of the weighted jack-knifed ratio estimator and the weighted jack-knifed variance estimator using the prediction theory approach.

1.1 Outline of Thesis

An outline of the study reported in this thesis is as follows:

In Chapter 2 we present a review of the randomisation based and the model based approaches to inference. We feel that such a review is necessary in order to gain an insight into the structure of inference problems concerning the characteristics of a finite population. It is argued that the model based approach is highly effective and offers inferential benefits over the randomisation based approach for a realistic and valid model.

The variance of an estimator depends on the framwork for inference. The randomisation based and the model based approaches provide different frameworks for defining variances. We consider variance estimation under both these approaches in chapter 3. We can think of this chapter as divided into two parts. The first part deals with the methods of variance estimation under the randomisation based approach. We note that for a given estimator $\hat{T}$ of T there are two variances, the true variance $V(\hat{T})$ and the estimator of the true variance. First, we discuss the methods commonly used for obtaining the expressions for the true variances. Then we discuss the methods for estimating the variances of estimates. Our main concern is with the jack-knife procedure. However, it seems to us that a knowledge of useful alternative techniques will be useful in a better understanding of the subject and will also facilitate a comparison of these techniques. Thus we discuss a number of competing techniques developed from time to time in an attempt to obtain accurate variance estimates necessary for survey sampling inference. We have, however confined our discussion to the main features of these techniques pointing out the usefulness and the limitations of each technique. We also introduce in this part a useful technique, known as the weighted jack-knife, developed recently as an alternative to the ordinary jack-knife method for variance estimation. It is observed that the randomisation approach to variance estimation offers several choices but it has been recognised that the randomisation based variances may not provide a suitable framework for inference from an observed sample. In contrast, under the prediction approach to finite population sampling, the variance of an estimated value is determined conditionally from the

set s of samples units actually observed. The prediction approach, by conditioning on the sample actually observed, reveals the relationships that are important for inference. Thus in the second part of this chapter we consider the variance estimation under the prediction approach. Our discussion is mainly centred round the ratio estimator for which the prediction approach has suggested useful alternatives to the randomisation based variance estimators. This chapter thus consolidates the material on the variance estimation from complex surveys under two main approaches, which lie scattered in the literature.

In Chapter 4 we consider the problems concerning the estimation of correlation from complex surveys. Frankel (1971), and Kish and Frankel (1974) studied the behaviour of simple, partial and multiple correlation and regression from clustered stratified samples. They conjectured that high relative biases for the multiple correlations were due to the basic defects of the estimator rather than the complexity of the design. For the estimation of population correlation coefficient from a stratified design, Koop (1970) and Wakimoto (1971 a, b) derived unbiased estimates for the population variances and covariances. Using these results we derive expressions for the biases in Kish and Frankel's estimates for variance and covariance. It is shown that for a self-weighting design, Koop and Wakimoto's results for the estimation of variances and covariance reduce to those of Kish and Frankel's estimate plus bias (Section 4.3). These theoretical results indicate that stratification can cause bias in

the estimation of correlation and that the bias may be of substantial magnitude. Kish and Frankel have mainly concentrated on the effects of clustering on variances and have ignored the effects of stratification on the bias. They were thus lead into misinterpretation of results. Our conclusion is that high relative biases in Kish and Frankel's estimates of correlation may be due to the effects of stratification. This conclusion finds support from Warren's (1971) analysis of the effect of sampling method on the estimation of correlation and regression summarised in Section 4.4. Warren concludes that non-simple random sampling can cause bias in the estimation of correlation and that this bias may be of substantial magnitude. Similar views have also been expressed by Smith (1974, discussion of Kish and Frankel). We have also investigated the effects of stratification on the behaviour of simple, partial and multiple correlations through a simulation study using data from Frankel (1971). The analysis of these results also supports our claim that the high relative biases in Kish and Frankel's estimates of correlation are due to the effect of stratification, rather than the basic defect in the estimator.

The Jack-knife method introduced originally by Quenouille as a method of bias reduction was extended progressively to obtain estimation of variance. This method has been shown to have robust qualities in a wide variety of fields. The most significant feature of the jack-knife technique is that bias reduction as well as variance estimation can be achieved without detailed knowledge of the sample distribution or involved analysis of the estimation method. We only

need a sample and an estimation definition. Miller (1974 b) has given an excellent review of the theoretical developments whereas Bissel and Ferguson (1975) discuss some practical aspects. However, in those reviews the authors restrict themselves to a brief discussion of the most important developments in the field of survey sampling. In Chapter 5 we attempt a review that focusses on the main developments concerning the applications of the jack-knife method in survey sampling. One of the earliest applications of the jack-knife method was in the case of ratio estimators which are frequently employed to improve the efficiency of estimation in sample surveys. We discussed the bias and variance estimation properties of the jack-knifed method for the ratio estimation under the model based approach in chapter 3. In this chapter we address ourselves to the problem of bias and variance estimation for the jack-knifed ratio estimator under the randomisation approach. The most important contribution of the jack-knife method is to produce an estimate of variance where alternative techniques appear to be cumbersome. The method has also been extended for variance component estimation in complex surveys (Folsom et al., 1971). We find that along with the theoretical developments, considerable empirical work has also been undertaken to compare the efficiencies of the jack-knifed and other competing variance estimators in small and moderate size samples from natural and artificial populations. This, however, lies dispersed in literature. We also attempt to summarise the empirical work on the jack-knifed ratio estimation under both the randomisation based and the model based approaches. Tukey proposed the jack-knife as a method of robust interval estimation.

Developments along this line are also discussed in this chapter. From this review we gather that the jack-knife appears to behave well both as a means of bias reduction and variance estimation where the estimator is a function of all data values concerned. This includes situations involving variances, maximum likelihood or least squares, regression, correlation and ratios. However, it does not perform well in the case of order statistics. We also note that the standard jack-knife procedure is sensitive to deviant data points and there is a need for considering alternative more robust methods.

Chapter 6 presents a discussion of the jack-knife application to the linear regression model $\underline{Y} = \underline{X}\underline{\beta} + \underline{e}$, where $\underline{Y}$ is a vector of n independent observations, $\underline{X}$ is a matrix of known values, $\underline{\beta}$ is a vector of unknown parameters and $\underline{e}$ is a vector of independent and identically distributed error variables. Following Hinkely (1977b) the properties of the jack-knifed regression estimator and the jack-knifed variance estimate are examined. It is found that the jack-knifed estimates of regression parameters are unbiased but less efficient and the jack-knifed variance estimator is biased.

The important point to consider in any data anlaysis is the impact that each observation has on the estimates. For example,

in some data sets one or more cases may have sufficient impact on the analysis and if these cases are removed from the data set completely different results would have been obtained. Such cases are known as influential cases. Study of influential cases is important from the inference viewpoint. In Chapter 7 we discuss the influence function and its role in regression analysis. We also introduce in this chapter various sample versions of the influence function which provide the basis and justification for various techniques used for the detection of influential cases. This chapter also introduces different methodologies for assessing the influence of individual or group of cases on regression analysis. However, we confine our discussion to those aspects of analysis that can be summarised by sample influence function.

In the unweighted jack-knife, discussed earlier in Chapter 6, the pseudovalues are defined symmetrically with respect to the observations whereas in the linear model situation the observations do not affect the least squares estimates in a symmetric manner. In order to adjust for this imbalance, Hinkley (1977 b) proposed weighted pseudovalues and hence the weighted jack-knife for regression estimation. We discuss the application of the weighted jack-knife to regression estimation in Chapter 8. First we give a suitable definition and justification for the weighted jack-knife. Then following the appraoch of Hinkley we

examine the properties of the weighted jack-knifed regression estimator and the weighted jack-knifed variance estimator. Hinkley assumed that the error variables e_i are uncorrelated. However, in practice, this may not be true in many situations. We give a simple generalisation of the weighted jack-knife for the general variance structure. I is shown that under this more general model the weighted jack-knifed regression estimator is unbiased and the weighted jack-knifed variance estimator is biased to $O(n^{-1})$.

In Chapter 9 we extend the weighted jack-knife technique to ratio estimation. Using the idea of empirical influence function, discussed in Chapter 7, we propose the weighted pseudovalues, the weighted jack-knifed ratio estimator and the weighted jack-knifed variance estimator. Using prediction theory approach we examine the properties of the weighted jack-knifed variance estimator. This variance estimator is shown to be approximatley unbiased under the model ξ for all variance functions with certain restrictions on n and f. Under the same conditions and assuming that the error distribution is normal, it is shown that the jack-knifed variance estimator is quite stable (Section 9.3). We also examine the implications of the failure of regression model on the behaviour of the weighted jack-knifed variance estimator. It is noted that the failure of the regression assumption $E(Y_i) = \beta x_i$ has a conservative

effect on the weighted jack-knifed variance estimator for the ratio estimate. Another important consideration with the jack-knife technique is that whether a given jack-knife procedure removes the first order bias. This important aspect of the weighted and the unweighted jack-knives is also examined in this chapter. It appears that the small sample performance of the weighted jack-knifed method for ratio estimation may be better than the unweighted jack-knife.

Finally, we consider some problems for further research work.

CHAPTER 2

RANDOMISATION-BASED AND MODEL-BASED APPROACHES FOR SURVEY SAMPLING INFERENCE

2.1 Introduction

Randomisation-based survey sampling theory, as presented in Cochran (1953, 1963, 1977); Hansen, Hurwitz and Madow (1953a, b) and several other similar text books has dominated the survey sampling literature until recently. The randomisation approach is based on the idea of repeated sampling from the same finite population. This approach is based upon the use of probability sampling designs and inferences based on the induced randomisation. The last two decades have brought a change in this outlook and as a consequence a model-based theory which refers to a superpopulation distribution and is based on a predictive argument as in Royall (1970), has emerged as a promising alternative. The model-based approach has contributed significantly towards a better understanding of a variety of survey sampling problems and has provided an avenue through which important new methods are currently being added to the existing survey sampling literature.

In this chapter, we review the inference problems concerning the characteristics of a finite population. In Section 2.2 we describe the basic concepts such as finite population and descriptive and analytic inferences. A great deal of current research in survey sampling is devoted to the investigation of an appropriate criterion of inference for finite populations. Through these efforts a considerable insight has been gained into the structure of inference for survey populations. We discuss the randomisation based and the model based approaches to inference for finite populations in Sections 2.3 and 2.4 respectively.

Our evaluation of these approaches is based on the important issues such as estimation, variance, asymptotic normality and conditional inference. Finally in Section 2.5 we give a summary and conclusions. It is argued that the model based approach is highly effective and offers inferential benefits over the randomisation approach for a realistic and valid model.

2.2 *Preliminaries*

We consider a finite population consisting of N units which are listed in a frame and are identifiable by labels that can be represented by $i = 1, 2, \ldots, N$. Associated with the unit i, is a vector of unknown values, y_i, which are to be measured in the survey. In addition some prior knowledge is available for each unit which can be utilized to design a representative sample. The prior information may include quantitative variables such as the measure of size and qualitative variables such as the membership of a cluster or a stratum. Let $\underline{Z}$ represent the matrix of prior information for all the units and $\underline{Y}$ the matrix of values of the measurement variables. A subset s of n labels is to be selected and the units with labels in s form the sample. For each unit in the sample, the values y_i are observed without error and an estimator, say $t_{(s)}$, is used to make a descriptive inference about population characteristics such as mean total and ratio.

Definition 2.1

The units of a population are said to be identifiable if they can be uniquely labelled from 1 to N and the label of each unit is known.

Definition 2.2

The vector $\underline{Y} = (Y_1, Y_2, \ldots, Y_N)$ is called a parameter of the finite population.

2.2.1 *Descriptive and Analytic Inferences*

Descriptive inferences may be characterised by the property that the parameter of interest is a known function of the values attached to the N units in the population.

The object is to estimate a given property of a given population at the time the sample was drawn. If all the units were evaluated, then in the absence of measurement errors, there would be no uncertainty in a descriptive inference. However, if the parameters of interest can not be expressed as a function of the values attached to the N units then the inference is analytic. The objective in this case is to estimate a parameter in another population related in some way to the population being sampled. This other population may often be a super-population proposed by the statistician. For example, estimation of a finite population total is a descriptive inference, whereas the estimation of a coefficient in an economic model is usually an analytic inference (Smith, 1981a).

There are currently two conflicting and well developed philosophies regarding the theory of descriptive inferences for samples drawn from finite populations. The first theory based on the distribution generated by random sampling, the P-distribution, is called randomisation inference. Smith (1976) gives an excellent review of the inferential problems relating to the randomisation distribution. The second theory employs stochastic models to represent the population structure and inferences based on the probability distribution specified in the model, the so called ξ-distribution, is called model based inference. Royall (1976a) gives an excellent description of this approach. The two theories of inference are discussed hereunder:

2.3 *Randomisation Based Approach to Inference*

Randomisation based inference following in the tradition of Neyman (1934) relies on the design based approach to survey sampling. This approach asserts that the primary source of randomness is the probability ascribed by the sampling design to various subsets of the finite population $\{1, 2, 3, \ldots, N\}$. According to this approach, the randomisation distribution created by the random choice of the units to be observed for a given sampling plan, forms the basis of randomisation inference. The aim is to provide robust reliable estimation of simple parameters such as means, totals, ratios, etc., in large scale multipurpose surveys, see for instance, Cochran (1976) The emphasis is on procedures that are randomisation consistent and work well regardless of population shape (Hansen et al., 1983).
In this section we consider the implications of the randomisation approach on estimation, variance, asymptotic normality and inference.

2.3.1 *Estimation*

The problem of estimation in survey sampling relates to finding good estimates of the population parameter. By a good estimator, we mean an estimator with its distribution concentrated near the population parameter. For example, Smith (1976, p.193) says: "The best strategy is to choose an estimator that reflects $\underline{Y}$ as closely as possible".

Under the randomisation approach the criteria for a good estimator are:
(1) Unbiasedness, (2) Consistency, (3) Efficiency, and (4) Sufficiency. Two main methods of estimation are those of least squares and the maximum likelihood. However, these methods have not been explicitly used in sample surveys. For example, Cochran (1977, p.150) observes that very little use has been made of the traditional

theory of statistical inference in survey sampling in order to make good estimates from data. Cochran gives two main reasons for this attitude. "Firstly, in surveys that contain a large number of items, there is a great advantage, even with computers, in estimation procedures that require little more than simple addition, whereas the superior methods of estimation, such as maximum likelihood, may necessitate a series of successive approximations before the estimate can be obtained. Secondly, most of the estimation methods in theoretical statistics assume that we know the functional form of the frequency distribution followed by the data in the sample and the method of estimation is carefully geared to this type of distribution. The preference in survey sampling has been to make only limited assumptions about this frequency distribution." This implies that the estimation techniques in sampling has been "restricted in scope". The usual practice has been of finding unbiased estimates and measuring their reliability in terms of interval estimation.

Before discussing the ideas such as unbiasedness, consistency, efficiency, sufficiency and likelihood, it appears necessary to state the formal framework for inference for finite population provided by Godambe and others from the Indian School of Statistics during the 1960's. Recall that a finite population consists of N labels, with label set $U = \{1, 2, \ldots, N\}$. The population parameter is the vector $\underline{Y} = \{Y_1, Y_2, \ldots, Y_N\}$. $\underline{Y}$ is a member of the parameter space $\mathcal{Y}$, which contains all possible populations and $\mathcal{Y} \in R_N$. Prior information exists on which the survey design can be based. A sample, s, is a subset of U, usually of n units, so that $s = (i_1, i_2, \ldots, i_n)$. A sampling scheme assigns a probability P_s to each s. S is the set of all possible samples and $\sum_S P_s = 1$. The data comprise the labels, s, selected by the scheme P_s and the associated values

Y_i, $i \in s$. Thus data = $(i, Y_i : i \in s)$. An estimator e_s, is a function of the data and any relevant prior information (Smith, 1976).

2.3.1.1 *Unbiasedness*

One property of a good estimator is unbiasedness. The randomisation approach defines the bias with respect to the sampling plan P.

Definition 2.3 (P-Unbiasedness)

An estimator t_s of the population parameter T is said to be unbiased with respect to a sampling plan P , if whatever $y_1, y_2, \ldots, y_N$, we have

$$E_p(t_s) = \sum_S P_s t_s = T \ . \qquad (2.1)$$

Such an estimator will be called P-unbiased. The bias is given by

$$\text{P-Bias} = \sum_S P_s(t_s - T) = B \ . \qquad (2.2)$$

Thus for a given sample, whether or not an estimator is P-unbiased depends on the probabilities with which the sampler might have chosen to observe other sets of units and not on the sample actually selected and the probabilistic relationship between the observed and the unobserved y's.

In the randomisation based approach P-bias is considered as an important performance characteristic. The reason for this is that under this approach the main criterion for evaluating an estimator is its mean square error determined by the randomness imposed by the sampling plan P . This quantity can be expressed as

$$MSE(t_s) = E_p(t_s - T)^2 = V(t_s) + B^2 \ , \qquad (2.3)$$

where $V(t_s) = E_p\{t_s - E_p(t_s)\}^2$, is the variance of the estimator.

Now supposing that $V(t_s)$ is relatively small for large samples, it appears quite natural with the randomisation based inference to search for an estimator for which B^2 is small. Thus the main emphasis in the randomisation approach is upon P-unbiased or P-consistent estimators.

Regarding unbiasedness, Smith (1984, 8.2) argues that "In the randomisation approach the values $\underline{y}$ are assumed to be unknown constants and the only probabilities are those introduced by the surveyor where a random selection scheme is used. If $P(s/\underline{z})$ is not a random selection scheme then no randomisation inference is possible. With random sampling the probability distribution is known exactly and does not depend on any unknown parameters, nor on the unknown constants $\underline{y}$. Since the values $\underline{y}$ do not index the distribution $P(s/\underline{z})$ they cannot be interpreted as parameters in the usual statistical sense of that term and so interpretation of $\underline{y}$ as a set of unknown constants which are neither random variables nor parameters seems justified".

"Since $\underline{z}$ and $P(s/\underline{z})$ are known, inferences about this distribution are irrelevant. The only way to employ $P(s/\underline{z})$ is to take averages or expectations. If for a statistic t_s, $E(t_s) = T$, the finite population function of interest, we can say that t_s is unbiased but if $E(t_s) \neq T$, then one can say nothing for the bias and any measure of efficiency will depend on the unknown population values y_i, $i \notin s$. *The concentration on unbiasedness in randomisation inference now appears natural because there really is no other property that can be determined given the sampling scheme* $P(s/\underline{z})$. Any other comparisons will depend on the unknown constants and hence can not be determined".

It may, however, be pointed out that it is not necessarily an advantage to use unbiased estimators because they may result in much larger mean squared error than biased but consistent estimators. Thus instead of unbiasedness, the randomisation approach emphasizes the consistency of the estimators with the minimisation of the mean squared error being the main criterion for choice among the estimators, see for instance, Hansen and Hurwitz (1943); Hansen, et al.,(Vol. 2 (1953)) and Kalton (1983).

2.3.1.2 Consistency

A method of estimation is called consistent if the estimate approaches the population parameter to be estimated as the sample size increases. For example, in case of simple random sampling, the sample mean $\bar{y}$ and the sample total $N\bar{y}$ are consistent estimators of the population mean and total respectively.

Definition 2.4 (Consistent Estimator)

An estimate of $\hat{\theta}$ of the population parameter θ is said to be consistent estimator if,

$$P(\hat{\theta} \to \theta) \to 1 \ , \quad \text{as} \quad n \to N \ . \qquad (2.4)$$

Equation (2.4) means the probability that $\hat{\theta}$ approaches θ as n becomes larger and larger is 1 , see for instance (Cochran (1977 p.21). However, Hansen, Hurwitz and Madow (1953) and Murthy (1967, p.39) give a different definition of consistency. According to these authors an estimator is consistent if the probability of the difference $(\hat{\theta}_n - \theta)$ being less than any specified small quantity tends to unity as n is indefinitely increased. Consider a sequence of estimators $\hat{\theta}_1, \ldots, \hat{\theta}_n$, then $\hat{\theta}_n$ is said to be a consistent estimator of θ , if $\hat{\theta}_n \to \theta$ as $n \to \infty$, where the index denotes the sample size. Formally,

if $\lim_{n\to\infty} P\{|\hat{\theta}_n - \theta| < \varepsilon\}_\theta = 1$, then $\hat{\theta}_n$ is a consistent estimator of θ.

Randomisation approach have comprised of

(i) selection procedures, such that each member of the population has a known non-zero probability of selection, and

(ii) estimators that are randomisation consistent, that is, for reasonably large enough samples the correctness of inference does not depend on any assumed model.

Thus, if unbiasedness is not required then it is possible to find biased but consistent estimators that will give smaller mean squared error, for some populations than those given by unbiased estimators. The familiar examples are ratio, regression and post-stratified estimators. To illustrate this point, we consider the usual ratio estimator for the population total $\hat{T}_R = \left(\frac{\sum_s y_i}{\sum_s x_i}\right) \sum_{i=1}^{N} x_i$, where x_i are known auxiliary variable values. This estimator is biased but consistent. The justification to use this estimator, however, does not depend on any principle of statistical inference related to the randomisation distribution but on practical intuitive reasons, see for instance Särndal (1978) and Smith (1984). These reasons are:

(a) an auxiliary variable x is available and $\sum_{i=1}^{N} x_i$ is known, and

(b) increased precision can be obtained because $\hat{T}_R$ takes advantage of the positive correlation between x and y . For example, Cochran (1963) on the basis of computation from an actual data (y_i = acres in corn for farm i , x_i = size in acres of farm i), concludes "that the choice between using farm size in the design (through stratification), or in the method of estimation (through ratio estimator), depends on a number of considerations, some purely practical".

Consistency is a desirable property of estimators. For example, Hansen, Madow and Tepping (1983, p.779) state that, "A necessary

condition for a probability sampling design is that it includes randomisation consistent estimators. The purpose is to help ensure that the biases of estimation for samples of finite size are small relative to the standard error". However, consistency is not a sufficient condition for the biases to be small relative to standard error for a given sample size n. The reason is that consistency is a property of a sequence of estimators, and single estimator may be a member of many different sequences. Thus, for a given sample, an estimator can be interpreted either as consistent or inconsistent depending on how the estimator and the sampling plan are assumed to change as the sample and the population size increase. This implies that the consistency has only limited usefulness and one needs to go beyond simply achieving consistency.

2.3.1.3 Efficiency

An estimator with smaller variance is preferable because the smaller the variance, the more concentrated the sampling distribution around the population parameter, assuming that we have consistent estimators.

Definition 2.5 (Efficiency)

If we have two estimators, $\hat{T}_1$ and $\hat{T}_2$, and $V(\hat{T}_1) < V(\hat{T}_2)$ then the efficiency of $\hat{T}_1$ relative to $\hat{T}_2$ is given by

$$E_f = \frac{V(\hat{T}_1)}{V(\hat{T}_2)} . \qquad (2.5)$$

We have defined the efficiency in relative terms, but if we could find an estimator with a variance that is smaller than that of any other estimator, we could use it as the basis to measure efficiency; and, in

terms of efficiency, we could say that an estimator with the smaller variance is an efficient estimator.

Neyman (1934) introduced the idea of efficiency through the concept of a shortest confidence interval in a class of intervals. Neyman considered only the estimators which were linear in the sample values. Invoking the Gauss Markov theorem, he derived minimum variance unbiased estimators. Because sample sizes would usually be large the distribution of these estimators would be approximately normal and hence the minimum variance estimators should give the narrowest confidence intervals. Using this method, he demonstrated that stratified random sampling would usually be more efficient than simple random sampling. A nice and detailed description of Neyman approach appears in Smith (1976), and O'Muircheartough and Wong (1981).

Ratio and regression estimation were introduced during the 1930's in order to incorporate into the estimation procedure an auxiliary variable which was assumed to be linearly related to the survey variable. The efficiency of the ratio and regression estimation depended on the relationship between the auxiliary variable and the survey variable. Cochran (1942) gives a comprehensive account of the theory. Hansen and Hurwitz (1943) introduced sampling with probability proportional to size as a convenient and efficient method for carrying out multistage sampling.

Horvitz and Thomson (1952) provided the theory for sampling with unequal probabilities without replacement. They also noted that the concept of a linear estimator as applied by Neyman was not as convenient as it appeared, since there were many classes of linear estimators being proposed in the literature. These authors identified three subclasses of linear estimators and concluded that, although a best

linear estimator may be found in each subclass this does not imply that any of the estimators is best for all classes. Thus the Neyman's concept of efficiency is limited in scope. They derived the "best" estimator in a particular class which is now known as Horvitz-Thompson estimator. This paper is of particular importance because it pointed towards serious problems that existed in defining the concepts like "linear" and "best" for finite populations, see, for instance, Smith (1976).

Godambe (1955) demonstrated that for sampling from a finite population there does not exist any unique best linear estimator for the population total, which means that no minimum variance unbiased estimator can exist uniformly for all possible populations. Basu (1971) criticises various randomisation based ideas including the P-unbiasedness, unequal probability estimator $\hat{T}_{HT}$ and the attempts made by Horvitz and Thompson (1952) to single out various classes of label-dependent linear P-unbiased estimators. Basu also provides a proof of Godambe's theorem. Lanke (1973, 1975) also gives a nice review and an entirely satisfactory proof. Smith (1976, p.187) says: "One consequence of this (Godambe's) non-existence theorem is that no empirical comparison can ever be conclusive, for in any particular case somebody may be able to construct a better estimator..... The problem of a lack of best estimators arises because of the generality of Neyman's formulation of the solution to inference problem. Inferences are made with respect to the P-distribution for any population γ, regardless of its structure. But this is too much freedom for a satisfactory theory of inference and no optimum properties can be found for all populations".

Since we know that there is no UMV estimator, extensive research work has been carried out to compare different estimators under different conditions in order to identify certain types of populations

for which the performance of the estimators is satisfactory. Let t_s be an estimator and P_s a design, then the pair (t_s, P_s) is called a strategy. Considerable empirical work has been done to compare different strategies. Recall that a strategy is efficient if it is related closely to the distribution of the population values $\underline{Y}$. Smith (1976, p.192) has tabulated the possible situations and sampling strategies which is reproduced here for illustration as Table 2.1.

Table 2.1 The Efficiency of Survey Plans (t_s, P_s) Relative to $\underline{Y}$

	t_s is related to $\underline{Y}$	t_s is not related to $\underline{Y}$
P_s is related to $\underline{Y}$	Very Efficient	Inefficient
P_s is not related to $\underline{Y}$	Efficient	Very Inefficient

The best strategy is to choose an estimator that reflects $\underline{Y}$ as closely as possible. A good estimator can overcome a poor design but a good design can be ruined by the use of a poor estimator.

Regarding efficiency, Smith (1976, p.193) says: ...."Neyman's global approach breaks down when only a single survey is considered. Here the user of a survey requires an efficient estimate together with a measure of its accuracy for the particular survey, not an estimate that has a property based on hypothetical repetitions of survey. When efficiency is considered then it has been shown that no best estimator can be found within Neyman's framework, primarily because that framework is too general and makes no assumptions about the distribution of the values in the population. Empirical studies show that estimators that reflect the population distribution but do not necessarily depend

on the survey deseign can be much more efficient than estimators that depend on the design through some criterion such as unbiasedness. Efficiency must be related to the population distribution and the fact that most surveys give satisfactory results is due to the survey design reflecting the population distribution rather than the fact that the framework for inference is satisfactory".

Since Godambe's theorem shows that there does not exist a best estimator within the randomisation model, therefore choosing estimators that are reasonable on other statistical grounds seems to be attractive. The framework of the superpopulation or ξ distribution (to be discussed in Section 2.4) appears to be an appropriate choice for generating such estimators.

2.3.1.4 Sufficiency

In many estimation problems, we may be able to summarise the information in the sample $y_1, y_2,\ldots,y_n$. This means that we can find some function of the sample that tells us just as much about θ as the sample itself. Such a function would be sufficient for estimation purposes and is accordingly called a sufficient statistic.

Definition 2.6 (Sufficient Statistic)

Let $y_1,y_2,\ldots,y_n$ be a random sample from the density $f(y; \underline{\theta})$, where θ may be a vector. A statistic $S = s(y_1, y_2,\ldots,y_n)$ is defined to be a sufficient statistic if and only if the conditional distribution of T given S does not depend on θ for any given statistic $T = t(y_1, y_2,\ldots,y_n)$.

The concept of sufficiency implies that if we know the value of sufficient statistic, then the sample values themselves are not needed and can tell us nothing more about θ. This is true since the distribution of the sample given the sufficient statistic does not depend on θ .

In the search for best estimators within the formal framework (Section 2.3.1) decision theoretic ideas such as admissibility have been applied but these efforts have resulted only in limited theoretical and negligible practical success. More restrictive criteria such as hyperadmissibility and necessary bestness have also been developed but proved to be totally unapplicable and hence no real progress could be made along these lines, see for instance, Smith (1976) and O'Muircheartaigh and Wong (1980).

Basu (1969) showed that statistic D, comprising the set of distinct labels in the sample s together with the corresponding observations y_i, is not only sufficient but minimal sufficient for the population parameter $\underline{Y}$. This implies that in sampling with replacement the repetition of a unit does not provide any extra information. Consequently the mean of the distinct sample units $\bar{y}_d$ can be shown to be superior to the sample mean $\bar{y}_s$ by the application of Rao-Blackwell theorem. Thus, even the Rao-Blackwellisation also does not help the construction of a unique best estimator. Since D is not a complete statistic, therefore, no "best" estimator can exist. Thus the idea of sufficiency leads to the conclusion that only an adequate summary of a survey is the complete set of data, see for instance, Smith (1976). Basu (1971) ascribed the incompleteness of D to the fact that for a non-informative design, the label part of D is an ancillary statistic, i.e., its distribution is independent of the parameter $\underline{Y}$. It is, therefore, possible to find non-trivial functions of D whose expectations are identically zero.

Definition 2.7 (Non-informative Design)

A design P is said to be non-informative if P_s does not depend, for any s, on the realised values y_i for $i \in s$. (Särndal, 1978, p.33)

Definition 2.8 (Ancillary Statistic)

When a quantity which is actually a random variable is treated as a fixed variable and the appropirate distribution for evaluating the standard error, etc., is the conditional distribution, then the conditioning variable of this kind is called an ancillary statistic (Royall, 1976a).

Since the Fisherian concept of sufficiency did not work, Godambe (1966) applied the concept of linear sufficiency to finite population sampling. Using this concept he was able to show that the sample mean $\bar{y}_s$ is a linear sufficient estimator for the population mean under any sampling design. Further, by applying a restrictive form of censoring, he obtained a unique linear sufficient estimator for the population mean in a class of all linear unbiased estimators, see for example, O'Muircheartaigh and Wong (1981).

However, Smith (1983a) says: "Basically the sufficient statistic is the observed data together with the associated labels which we can write as

$$d_s = \{(i, y_i): i \in s\} .$$

By including the labels all we can say is that the units in the sample take their observed values and that the unobserved units can take any set of values in the appropriate parameter space. This is rather limited inference".

2.3.1.5 Likelihood

The application of likelihood in the finite population context also involves difficulties in the randomisation based inference. The basic result is that the likelihood takes the value P_s for all

populations which would have generated the sample data and zero for all other populations. This implies that after selecting and observing the sample, we know everything about the sampled units $(y_i, i \in s)$ and nothing about the non-sampled units $(y_i, i \notin s)$.

Under the traditional theory the probabilities relate to hypothetical infinite population $f(y, \underline{\theta})$ which depends on a vector of parameters $\underline{\theta}$ and is usually of small dimensions. In such applications of likelihood, this model is provided by the common probability distribution $f(y, \underline{\theta})$ and the sample provides the estimate $\hat{\underline{\theta}}$ of $\underline{\theta}$ from which future values of y can be predicted. In survey sampling the population, $\underline{Y}$, is real and the parameter has N dimensions, where N is usually very large. Since the parameter space in this case is N dimensional whereas the sample data is only n dimensional with $n < N$, the the conventional analysis does not work. Thus the only way to get information on $(N-n)$ unobserved units is to relate them by a mathematical model to the n observed units, see Smith (1976) for further details.

Several alternative approaches have been suggested to overcome these problems. One method is to ignore the labels, i.e., to base the inference only on unlabelled samples. Utilising this idea Hartley and Rao (1968) and Royall (1968) developed a likelihood function of the form given by the hypergeometric distribution. This new likelihood function permits the construction of a unique maximum likelihood estimator, see for instance, Smith (1976) and O'Muircheartaigh and Wong (1981).

Other main approaches to compare the estimators are:

(i) Bayesian (Ericson, 1969) based on the idea of replacing the un-informative labels by that of exchangeable prior information.

(ii) To assume that the finite population has itself been generated

as a random sample from an infinite superpopulation model; and (iii) Linear least squares prediction approach (Royall, 1970).

The linear least squares approach,using the linear model and the Gauss Markov theorem,has produced some interesting results. We shall discuss this approach in Section 2.4.

2.3.2 *Variance*

The precision of an estimator is determined by the variance if the estimator is unbiased and by the mean square error if the estimator is biased. Variances are therefore required for the choice of estimators. However, variance depends on the framework for inference. Both the randomisation based and the model based approaches are well developed in terms of variances. The main difference between these approaches lies in using different probability frameworks for defining the variances. The randomisation approach defines the variance with respect to the sampling distribution.

Definition 2.9 (P-Variance)

The P-variance of an estimator $\hat{T}$ with respect to the sampling plan P, is defined as

$$V(\hat{T}) = E_p\left(\hat{T}-E(\hat{T})\right)^2 = \sum_S P_s\left(\hat{T}-E(\hat{T})\right)^2 . \qquad (2.6)$$

Similarly, MSE of an estimator $\hat{T}$ is defined as

$$MSE_p(\hat{T}) = E_p(\hat{T}-T)^2 = \sum_S P_s(\hat{T}-T)^2 . \qquad (2.7)$$

$$= V_p(\hat{T}) + B^2 .$$

Thus under the randomisation appraoch, for a fixed size sampling plan, the variance and the standard error of an estimator are expressed

as averages over all possible samples and are therefore constants which do not depend on the sample actually observed. These variances are therefore unconditional. Consider for example the regression estimate. The variance of a regression estimator is constant when it is calculated with respect to the randomisation approach, that is, the same standard error applies for every sample. On the other hand, the regression model yields a different formula for the standard error, a formula which depends on the distribution of x in the sample and that in the population. Thus, for a given sample selected at random, the two approaches lead to different estimated standard errors. They also lead to different strategies for selecting the sample. The randomisation approach to the regression estimator requires that a random sampling plan be used with every possible sample of the desired size having the same probability of being chosen. The prediction approach, on the other hand, suggests that a sample whose mean is not close to that of the population mean should be rejected, especially if the x's in the sample are not well spread.

Royall and Cumberland (1981a) state that "Under the randomisation approach there is no general mechanism of studying or even defining the variance in relation to the particular sample". Smith (1976) says: "Should not the statement of the accuracy of an estimator depend on the sample that has been drawn rather than on all possible samples that might be drawn. Regarding the appropriateness of standard errors, Royall (1976a) says: "Of two standard errors, the one whose mathematical model is least doubtful is not necessarily the most appropriate for inference and the other which is most appropriate for inference is not necessarily the one whose mathematical model requires fewest assumptions or even the most plausible assumptions.....The appropriately conditioned standard error which reflects the accuracy of the estimate

calculated from the sample is that derived under the regression model with sample units held fixed". Lahiri (1968) has also expressed similar views.

We note from the above discussion that Lahiri (1968), Royall 1976a), Royall and Cumberland (1981a, b) and Smith (1976; 1983a) have convincingly argued that P-distribution is often not relevant as a frame of reference for assessing the precision of an observed estimate after the sample is drawn. These authors have advocated a conditional measure of accuracy based on the sample actually observed.

2.3.3 *Asymptotic Distribution*

We have discussed concepts such as unbiasedness and standard error under the randomisation approach. Inference, however, depends not only on bias and variance but also on the asymptotic normality, see for example, Smith (1984, 8.7). This implies that inference requires a relationship between these statistical characteristics. If the sample is large then, for most of statistics, such a relationship is established by an appeal to the central limit thereom.

The usual practice is that having selected an estimator, its randomisation variance is calculated and inferences are made by an appeal to a form of a central limit theorem, see for example, Madow (1948) and Hájek (1960). Thus in randomisation approach, for samples large enough, the validity of randomisation inferences does not depend on the assumptions concerning the distribution of characteristics in the finite population from which the sample was drawn.

Hansen, Madow and Tepping (1983) argue that inferences should be based on the criteria of consistency and asymptotic normality. However, for finite populations such an appeal to asymptotic results is not possible directly but it is hoped that if the sample sizes are large enough then approximate normality will hold.

For finite populations, the asymptotic justification of this result in the randomisation approach is provided by considering that the given finite population is embedded in a sequence of finite populations indexed by ν and of increasing size N_ν with increasing sample size n_ν . Properties such as asymptotic unbiasedness, consistency and asymptotic normality are determined by assuming that $n_\nu \to \infty$ and $N_\nu \to \infty$, often holding $\frac{n_\nu}{N_\nu}$ fixed, see for instance, Scott and Wu (1981), and Smith (1984).

Smith (1983b) argues that although randomisation is well defined, the method of making inferences is not assumption free. He says: "Neither of these criteria (consistency and asymptotic normality as suggested by HMT, 1983) is defined for the given finite population, both require the construction of a hypothetical sequence of finite populations of increasing size. This sequence is just as much a statistical model as a regression model relating the two variables X and Y , the main distinction being that the construction of the sequence is purely arbitrary and can never be tested against data".

The arbitrary nature of the sequences has also been recognised by Hansen, Madow and Tepping and discussed in Sub-Section 2.3.1.2. The same argument could be applied to show that a given estimator will be asymptotically normally distributed or not, depends on the

sequence selected. Besides, it is possible to construct finite populations for which the central limit theorem does not hold, see for instance, Smith (1979).

From this discussion we conclude that the randomisation inference is not free of assumptions and the usefulness of the asymptotic results under the randomisation approach is restricted in scope. In view of this situation, as suggested by Smith (1983a), there is a need for an approximate theory based on the given finite population rather than a hypothetical sequence of finite populations.

2.3.4 *Unconditional Inference*

The randomisation approach to finite population sampling is believed to provide mathematically rigorous, objective and probabilistic inferences, using the concepts such as unbiasedness, consistency and asymptotic normality, which are claimed to be free from distributional assumptions. These inferences are based on the distribution created by the artificial randomisation which is induced by the investigator when he uses a random sampling plan to choose the units for observation. For example, Hansen, Hurwitz and Madow (1953b), Kish (1965, p.23-24), and Stuart (1976, p.32) state that the use of probability sampling plan creates the only probability distribution on which reliable statistical inferences can be based.

Fisher (1935) suggested that artificial randomisation might provide an "absolute guarantee of the validity of the calculations" used in inference. This idea developed into the doctrine called randomisation based inference. His approach has a strong influence on the finite population sampling theory. For example, Godambe (1969)

says "Presumably randomisation is adopted in practice, with a view to utilizing the sampling distribution generated by randomisation". Hansen, Madow and Tepping (1983, p.776) state that: Design decisions may be guided and evaluated by models, but inferences concerning population characteristics should be made on the basis of the induced randomisation, at least when samples are reasonably large". These authors, while comparing inferences from prediction theory and the randomisation theory, argue: "Probability sampling methods provide a confidence interval for the population characteristics being estimated, and for large enough samples, the confidence interval is valid and short enough to provide as precise statements as desired about the value being estimated". These and other similar arguments suggest that in the absence of randomisation, valid probabilistic inferences are impossible.

However, the arguments against the randomisation inferences are also very strong. For example, Neyman and Pearson (1937) rejected the randomisation approach and used a regression model in the analysis of agricultural data. Their main argument was that the real world relationships are more complicated than those to be represented in a mathematically tractable model. They stated: "Mathematics deals with mathematical conceptions not with real things and we can expect no more than a certain amount of correspondence between the two". Royall (1976a) observes: "The probability distribution determined by artificial randomisation is not necessarily appropriate even when it is available. To claim that, in general, probabilistic inferences are not valid when the randomisation distribution is not available is simply wrong. To claim that, in general, a meaningful standard error estimate can not be calculated in the absence of mechanical randomisation is simply wrong. This is not to deny that randomisation is valuable, but only to deny that it represents the basis of all, valid, rigorous, probabilistic inference".

Smith (1976) argues that: "Although randomisation can be justified, this does not mean that inferences have to be based on the randomisation distribution. Fisher's position was that the randomisation was relevant before the data were collected but not in the analysis of data and this has been the position of most statisticians in the experimental sciences. Why should survey statisticians adopt a different standpoint?" Royall and Cumberland (1981a) also express a similar viewpoint. They state "Probability sampling distribution is not a prerequisite for rigorous statistical inference, valid inferences can be made without it and when randomisation is present, the distribution it creates is not necessarily appropriate for inference".

Little (1983), states that the randomisation based theory of inference, being a large sample theory, is more restricted in applications than the model based approach. Moreover, the implied coverage properties of the design based estimators and standard errors are often a hope rather than a reality. For example, Royall and Cumberland (1981a) and Holt and Smith (1979) show that even when randomisation confidence intervals have the correct coverage probabilities over all samples, the coverage properties of these intervals for samples from identifiable subgroups of populations such as post-strata, may be quite unacceptable. Royall and Cumberland also show that for some populations the overall unconditional coverage properties are unsatisfactory.

Smith (1983b) notes that the randomisation distribution is defined on a given sampling frame for a fixed set of measurements variable $\underline{Y}$. Thus only descriptive inferences can be made to this frame and analytic inferences to some wider superpopulations require additional model type assumptions. Besides, in social surveys non-response and missing values add a non-random selection to even the best

random sampling schemes and this destroys the strict basis for randomisation inference. He concludes that the randomisation inference is fraught with conceptual difficulties and is usually advocated on the grounds of robustness, not on those of the statistical theory. However, Smith (1984, 8.2) says: "Although randomisation distribution is assumption free, and hence must be robust, randomisation inferences require assumptions which may not be satisfied and hence the inferences may not be robust".

We note that the randomisation inferences are unconditional and average overall possible samples that might have been drawn. As noted by Smith (1984) there are no principles which could lead to conditional inferences in the randomisation approach and this is the major problem with these inferences.

2.4 *Model Based Inference*

The model based theory of inference for sample surveys has emerged as an alternative to the randomisation based approach. In the model based approach to inference for a finite population, the measurement values $y_1, y_2, \ldots, y_N$ associated with N units of the population are assumed to be the realised outcome of the random variables $Y_1, Y_2, \ldots, Y_N$ having N dimensional joint distribution ξ, where the superpopulation ξ is modelled to reflect the available background knowledge. Usually ξ is indexed by an unknown parameter θ, the estimation of which is required as a preliminary to making inferences about the finite population itself. The model ξ provides the relationship between y_i and its parameter θ and also the probability distribution through which the estimators may be evaluated. Thus the observed values $y_i, i \in s$ provide both estimates of θ for analytic inferences and prediction of $y_i, i \notin s$ for descriptive inferences. The prediction

approach has been developed by Royall et al., in a series of papers during the seventies. The main references are: Royall (1970, 1971, 1976a, b, 1982), Royall and Herson (1973a, b), Royall and Eberhardt (1975) and Royall and Cumberland (1978, 1981a,b, 1982 and 1983).

It may, however, be pointed out that the model based approach to inference from finite population is not restricted to any particular statistical philosophy. Alternative models can be constructed using the ideas of exchangeability or the Bayesian framework. For example, Kalbfleisch and Sprott (1969) approached the finite population inference problem from the fiducial prediction theory viewpoint whereas Ericson (1969) and Sugden, (1979) proposed models based on exchangeability assumptions. Scott and Smith (1969) make the classical normal distribution assumptions and they also employ the closely related linear models. Our discussion is concentrated on the non-Bayesian superpopulation approach developed by Royall (1970) using a linear model. We are mainly concerned with the problems such as estimation, variance, asymptotic distribution and conditional inference.

2.4.1 Estimation

The non-existence of optimal estimators (Godambe, 1955) and the non-informativeness of the likelihood (Godambe, 1966), as discussed earlier in Section 2.3, stimulated a great deal of research and have brought more clearly the limitations of the randomisation model. The basic problem pointed out by this research, however, concerns the relationship of the observed units to the unobserved ones. The randomisation model discussed earlier in Sction 2.3.1 expresses no relationship between the observed and the unobserved units except that the unobservedones could have been in the model. Royall (1976a) regarding this aspect says: "If we are to learn about some units (those not in the sample) from the

observations we have made on others (those in the sample), then the two groups must have a stronger logical connection than this.

Just such a connection was provided by the superpopulation models which had been waiting in the wings of finite population sampling theory since the 1940's. These models characterise the actual population values, both those observed in the sample and the unobserved ones as realisation of random variables. The joint probability law of these random variables supplies the link between sample and non-sample units which is missing from the randomisation model. From the sample we learn about the probability law, which is then used to predict the values of the unobserved variables. Within this framework, finite population estimation problems are essentially problems in statistical prediction".

Under the model based approach the bias and variance of an estimator are defined with respect to the superpopulation model ξ .

Definition 2.10 *(ξ-unbiasedness)*

Under the model ξ an estimator $\hat{T}$ of T will be called ξ-unbiased if for every s ,

$$E_{\xi}(\hat{T}-T) = 0 \ . \qquad (2.8)$$

Thus ξ-unbiasedness is defined with reference to ξ only and does not depend on the sampling plan P used to decide which units should be observed.

The model based approach treats the sample units as fixed and averages over the different values of y_i that the model might have generated. This approach presents a different outlook for the estimation of population total than that provided by the randomisation approach. Under the prediction approach, after the sample is observed, the population total, T , can be written as the sum of two parts

$$T = \sum_{i=1}^{N} y_i = \sum_{s} y_i + \sum_{\tilde{s}} y_i \ , \qquad (2.9)$$

where $\tilde{s}$ denotes the set of non sample units. Since $\sum_{s} y_i$ is known without error, the problem is to predict $\sum_{\tilde{s}} y_i$ from the sample. If $\hat{T}$ is any estimator of T, then $\hat{T} - \sum_{s} y_i$ is the implied estimator of $\sum_{\tilde{s}} y_i$. Smith (1976) writes some conventional estimators in the predictive form, we give two of these estimators here in order to illustrate this approach.

Example 2.1 (Simple Random Sampling Without Replacement)

$$\hat{T} = N\bar{y} \ ,$$

where

$$\bar{y} = \sum_{s} y_i/n \ .$$

This can be written as

$$\hat{T} = \sum_{s} y_i + (N-n)\bar{y} \ ,$$

which implies,

$$\hat{Y}_i = \bar{y} \ , \quad i \notin s \qquad (2.10)$$

This appears to be a sensible estimator.

Example 2.2 (Simple Random Sampling With Replacement)

$$\hat{T} = N\bar{y} \ , \ \bar{y} = \sum_{i=1}^{n} y_i/n \ \text{ including any repetitions.}$$

Let $\bar{y}_d$ denote the mean of the distinct units, then

$$\bar{y}_d = \frac{1}{d} \sum_{i \epsilon s} y_i$$

Now

$$\hat{T} = \sum_{i \epsilon s} y_i + (N-d)\left\{\bar{y} + \left(\frac{d}{N-d}\right)(\bar{y}-\bar{y}_d)\right\} \ .$$

This implies that

$$\hat{Y}_i = \bar{y} + \frac{d}{N-d}(\bar{y}-\bar{y}_d), \quad i \in s \qquad (2.11)$$

This is not an intuitive estimator of Y_i and explains why $\bar{y}$ is not a sufficient statistic as discussed earlier in subsection 2.3.1.4.

2.4.1.1 Estimation Strategies Under the Model ξ .

Consider the linear regression model,

$$Y_i = \beta x_i + e_i \ , \qquad (2.12)$$

where e's are independent random variables with $E(e_i/x_i) = 0$, $E(e_i^2/x_i) = \sigma^2 v(x_i)$ and $E(e_i, e_j/x_i, x_j) = 0$ for $i \neq j$. This model provides a relationship between y_i and x_i . The appropriateness of $\hat{T}$ as an estimator of T is evaluated in terms of the properties of the difference $\hat{T} - \sum_s y_i$ as a predictor of the sum of the random variables $\sum_{\bar{s}} y_i$. For example, $\hat{T}$ is described in the prediction approach as being unbiased for T , if for every sample, $\hat{T} - \sum_s y_i$ is an unbiased predictor of the unobserved sum $\sum_{\bar{s}} y_i$, that is if $E_\xi(\hat{T} - \sum_s y_i) = E_\xi(\sum_{\bar{s}} y_i)$. Thus, whether an estimator is unbiased or not in a particular application depends on the form of the regression function $E(y_i)$ and the sample s . For example, for the model (2.12), the ratio estimator $\hat{T}_R$ is the best linear ξ-unbiased estimator with $v(x_i) = x_i$.

Definition 2.11 (Pξ-unbiasedness)

An estimator $\hat{T}$ will be Pξ-unbiased predictor of T , if and only if, for a given P and ξ , $E_p E_\xi(\hat{T}-T) = 0$.

Definition 2.12 (Estimation Strategy)

A sampling and estimation strategy $P:\hat{T}$ will be considered better than another strategy $P':\hat{T}'$, if the former has a smaller mean square error than the latter, that is , if

$$MSE(P:\hat{T}) < MSE(P':\hat{T}')$$

where $MSE(P:\hat{T})$ is defined as,

$$MSE(P:\hat{T}) = E_\xi[\sum_s P_s(\hat{T}-T)^2] = E_\xi E_p(\hat{T}-T)^2 \ .$$

This implies that a strategy with smaller expected mean square error with respect to the distribution of $Y_1, Y_2 \ldots, Y_N$ is a better one.

2.4.1.2 Choice of Estimators Under the Regression Model.

An estimator $\hat{T}$ for T can be uniquely expressed as,

$$\hat{T} = \sum_s y_i + \hat{\beta} \sum_{\tilde{s}} x_i \ , \qquad (2.14)$$

where $\hat{\beta}$ is the implied estimator of β and does not depend upon the unobserved y's . If we have two statistics T^* and T^{**} , then the one whose implied estimator β is better is the better estimator for T . This result is expressed as Lemma 2.1.

Lemma 2.1 (Royall, 1970).

For any sampling plan P , if T^* and T^{**} have implied estimators β^* and β^{**} for β which satisfy,

$$E_\xi(\beta^* - \beta)^2 \leq E_\xi(\beta^{**} - \beta)^2 \ , \qquad (2.15)$$

for each s such that $P_s > 0$, then

$$MSE(P:T^*) \leq MSE(P':T^{**}) \ . \qquad (2.16)$$

This lemma implies that a sampling and estimation strategy $P:T^*$ is better than other strategy $P':T^{**}$ if the expected (with respect to ξ) mean square error of the former is less than the expected sampling mean square error for the latter.

It will be interesting to give an important result in linear regression and prediction theory. This result is expressed in the form of Theorem 2.1.

Theorem 2.1 (Royall, 1970)

For any sampling plan P, if $\hat{T}$ is the linear estimator satisfying either,

(i) $E_\xi(\hat{T}-T) = 0$, for every s such that $P_s > 0$, or

(ii) for every σ^2, $MSE(P:\hat{T})$ is a bounded function of β, then

$$MSE(P:\hat{T}^*) \leq MSE(P:\hat{T}) , \tag{2.17}$$

with strict inequality unless $\hat{T} = \hat{T}^*$ with probability one under model ξ for all s such that $P_s > 0$.

Proof

Define

$$\hat{\beta}^* = \sum_s \frac{x_i y_i}{v(x_i)} \Bigg/ \sum_s \frac{x_i^2}{v(x_i)} , \tag{2.18}$$

the weighted least squares estimator of β, and

$$\hat{T}^* = \sum_s y_i + \hat{\beta}^* \sum_{\tilde{s}} x_i \tag{2.19}$$

From (2.14), it is clear that T is a linear function of the sample y's if and only if $\hat{\beta}$ is a linear function of these variables, that is, $\hat{T} = \sum_s y_i + \hat{\beta} \sum_{\tilde{s}} x_i$.

Now

$$E_\xi(\hat{T}-T) = E_\xi\left[\sum_s y_i + \hat{\beta} \sum_{\tilde{s}} x_i - \sum_s y_i - \sum_{\tilde{s}} y_i\right] .$$

$$= E_\xi\left[(\hat{\beta}-\beta) \sum_{\tilde{s}} x_i\right] . \tag{2.20}$$

$E_\xi(\hat{T}-T) = 0 \Rightarrow E_\xi\left[(\hat{\beta}-\beta) \sum_{\tilde{s}} x_i\right] = 0$, which is true only if $E_\xi(\hat{\beta}-\beta) = 0$.

$$MSE(P:\hat{T}^*) = E_\xi E_p (\hat{T}^* - T)^2$$

$$= E_\xi E_p \left[\sum_s y_i + \hat{\beta}^* \sum_{\tilde{s}} x_i - \sum_s y_i - \sum_{\tilde{s}} y_i \right]^2 ,$$

$$= E_p E_\xi \left[(\hat{\beta}^* - \beta)^2 \left(\sum_{\tilde{s}} x_i \right)^2 + \left(\sum_{\tilde{s}} e_i \right)^2 \right] \qquad (2.21)$$

Similarly,

$$MSE(P:\hat{T}) = E_p E_\xi \left[(\hat{\beta} - \beta) \sum_{\tilde{s}} x_i - \sum_{\tilde{s}} e_i \right]^2 . \qquad (2.22)$$

Using Lemma 2.1, we have

$$MSE(P:\hat{T}^*) \leq MSE(P:\hat{T}) \quad , \quad \text{if} \quad E_\xi(\hat{\beta}^* - \beta) \leq E_\xi(\hat{\beta} - \beta) .$$

This theorem shows that under the model ξ , if $\hat{\beta}^*$ is the weighted least squares estimator of β , then for any sampling plan P , a best linear ξ-unbiased estimator of T is $\hat{T}^*$. A result similar to this theorem is given by Scott and Smith (1969) in the Bayesian context.

This theorem leads to three best linear ξ-unbiased estimators for T , when $v(x) = 1$, x and x^2 respectively. These are:

Case I : $v(x) = 1$, gives

$$\hat{T}_0 = \sum_s y_i + \frac{\sum_s x_i y_i}{\sum_s x_i^2} \sum_{\tilde{s}} x_i . \qquad (2.23)$$

Case II : $v(x) = x$, gives

$$\hat{T}_1 = \sum_s y_i + \left[\sum_s \frac{x_i y_i}{x_i} \Big/ \sum_s \frac{x_i^2}{x_i} \right] \sum_{\tilde{s}} x_i \ ,$$

$$= \sum_s y_i \left[1 + \frac{1}{\sum_s x_i} \sum_{\tilde{s}} x_i \right] \ ,$$

$$= \frac{\sum_s y_i}{\sum_s x_i} \sum_{i=1}^{N} x_i \ . \qquad (2.24)$$

which is the usual ratio estimator.

Case III : $v(x) = x^2$, gives

$$\hat{T}_2 = \sum_s y_i + \left[\sum_s \frac{x_i y_i}{x_i^2} \Big/ \sum_s \frac{x_i^2}{x_i^2} \right] \sum_{\tilde{s}} x_i \ ,$$

$$= \sum_s y_i + \frac{1}{n_{(s)}} \sum_s \left(\frac{y_i}{x_i} \right) \sum_{\tilde{s}} x_i \ . \qquad (2.25)$$

These estimators are best for any sampling plan.

The Horvitz-Thompson estimator with probability proportional to size is PPS unbiased. Godambe (1955) and Godambe and Joshi (1965) have shown that if only fixed size sampling plans are considered, then under model ξ with $v(x) = x^2$, $MSE(PPS:\hat{T}_{HT}) \leq MSE(P:\hat{T})$ for any P and any $\hat{T}$ which is P-unbiased. However, if P-unbiasedness is not demanded then this inequality can be reversed. Besides, PPS sampling plans with $v(x) = x^2$, this inequality holds for any fixed size sampling plan and a wide class of variance functions. This can be expressed as Theorem 2.2.

Theorem 2.2 (Royall, 1970)

If (i) $\max(nx_i) \leq \sum_{j=1}^{N} x_j$ for all $i=1,2,\ldots,N$ and (ii) $v(x)/x^2$ is

a non-increasing function, then for any sampling plan P for which $P_s > 0$ only if $n(s) = n$

$$MSE(P:\hat{T}_{HT}) \gtrsim MSE(P:\hat{T}_2) \ . \tag{2.26}$$

In practice, the variance function v is not known exactly. It is, therefore, desirable not only to derive the best estimator for a particular v but also to compare estimators under various assumptions for v. This theorem facilitates such a comparison between $\hat{T}_{HT}$ and $\hat{T}_2$. We can also have similar comparisons between estimators which are optimal for simple variance functions.

Example 2.3 (Royall, 1970)

Let g be any non-negative real number, $0 \lesssim g \lesssim h$ and $\hat{T}_g = \sum_s y_i + \hat{\beta}_g \sum_{\bar{s}} x_i$ be the estimator which is optimal for $v(x) = x^g$, then

$$MSE(P:\hat{T}_g) \lesssim MSE(P:\hat{T}_h) \ . \tag{2.27}$$

for every P. However, if $v(x)/x^h$ is non-decreasing, then above inequality can be reversed.

From this example it follows that if $v(x)/x^2$ is non-decreasing, then the estimators $\hat{T}_0$, $\hat{T}_1$ and $\hat{T}_2$ satisfy the relation

$$MSE(P:\hat{T}_2) \lesssim MSE(P:\hat{T}_1) \lesssim MSE(P:\hat{T}_0) \ .$$

2.4.1.3 Optimal Sampling Plan:

We note that Theorem 2.1 says that under the given model with variance function v, the estimator $\hat{T}^*$ is optimal for P. It is thus obvious that for a fixed size sample an optimal sampling plan is

one for which minimum value of $E_\xi(\hat{T}-T)^2$ is achieved. This implies that for the given estimator $\hat{T}$ and sample size n an optimal sampling plan is clearly one that selects s in S_n (collection of all sets s in S which contain exactly n labels) with certainty so that $MSE(P:\hat{T})$ is minimised. In order to get a clear idea about what is achieved by the choice of such a plan we express $E_\xi(\hat{T}-T)^2$ in another form, that is

$$E_\xi(\hat{T}-T)^2 = E_\xi\left[\sum_s y_i + \hat{\beta}\sum_{\tilde{s}} x_i - \sum_s y_i - \sum_{\tilde{s}} y_i\right]^2 ,$$

$$= E_\xi\left[(\hat{\beta}-\beta)\sum_{\tilde{s}} x_i - \sum_{\tilde{s}} e_i\right]^2 ,$$

$$= E_\xi\left[(\hat{\beta}-\beta)^2\left(\sum_{\tilde{s}} x_i\right)^2\right] + \sigma^2 \sum_{\tilde{s}} v(x_i) . \qquad (2.28)$$

From expression (2.28) we observe that a sampler has two choices. First is to choose a sample that will provide a good estimate of the expected values of the total of the non-sample units. This implies that s should be selected such that $\left(\sum_{\tilde{s}} x_i\right) E_\xi(\hat{\beta}-\beta)$ is small. The second being to observe those units whose y values have greatest variance, so that only the sum of the least variable values must be predicted. This means to choose s so that $\sum_{\tilde{s}} v(x_i)$ is small.

If the optimal estimator $\hat{T}^*$ is to be used, then for a wide class of variance functions, the best sample would be to observe those units whose x values are the largest. This can be shown through Theorem 2.3.

Theorem 2.3 (Royall, 1970)

If $v(x)$ is non-decreasing, $v(x)/x^2$ is non-increasing and the optimal estimator $\hat{T}^*$ is to be used, then

$$MSE(P^*:\hat{T}^*) \leq MSE(P:\hat{T}) \quad , \tag{2.29}$$

for any sampling plan P in P_n and any ξ-unbiased linear estimator $\hat{T}$, where P_n denotes the collection of all sampling plans P with fixed size n and P^* is a sampling plan which selects s^* with certainty, i.e., $P^*(s^*) = 1$ and s^* is any set of n labels for which $\max_{S_n} \sum_s x_i = \sum_{s^*} x_i$.

This theorem shows that for $v(x) = x$, the ratio estimator is optimal under a model which is adequate for many actual populations and the optimal sampling plan to use with this statistic is the purposive sampling plan P^*. The sampling plan P^* is also optimal for use with estimators $\hat{T}_{HT}$ and $\hat{T}_2$, see for example, Royall (1970).

The purposive sampling scheme is not rubust to departures from assumptions. Royall and his co-workers have concentrated their efforts to make model based methods more robust by increasing constraints on the sample. Royall and Herson (1973a, b) introduced balanced sampling. A balanced sample matches the properties of the known prior design variables $\underline{z}$ in the sample to those in the population as a whole. An estimator can be protected against failure of linear regression model by balancing on the higher moments of the auxiliary varaible. Royall and Cumberland (1981a) proposed a best fit sample. The best fit sample is a purposively selected sample which balances the sample c.d.f $F_s(z)$ and the population c.d.f $F(z)$. The sample is selected such that $\max|F_s(z) - F(z)|$ is minimized.

2.4.2 Variance

The randomisation based variances as discussed earlier in Section 2.3.2 are unconditional, that is, they do not depend on the sample

actually observed and therefore are not suitable for making the inferences from the unique sample actually observed. A fundamentally different approach to the problem of variance estimation which depends on the sample actually observed, is the model based approach.

Definition 2.11 (ξ-Variance)

Under the model ξ, the variance of an estimator $\hat{T}$, for any given s, is defined as

$$V_\xi(\hat{T}) = E_\xi(\hat{T}-E_\xi\hat{T})^2 \ . \tag{2.30}$$

In the model based approach the standard error of an estimated value is made conditional on the observed sample and thus follows the conditionality principle (Fisher, 1955; Cox, 1958; Cox and Hinkley, 1974) of conditioning on an ancillary statistic s, see for instance, Royall (1976a), Royall and Cumberland (1981a) and Smith (1976).

The model based approach provides reliable estimates of precision which do not share the poor properties of the randomisation based standard errors when the sample happens to contain units with extreme values. Both the randomisation based and the model based approaches have been applied to ratio and regression estimators and the model based approach has provided quite promising results about variances and variance estimators as compared to the standard results under the randomisation approach, see for instance, Royall and Cumberland (1981a, b, 1982, 1983).

However, the model based approach has also been criticised. For example, Hansen, Madow and Tepping (1983) state: "If the assumed model does not accurately represent the state of nature, estimates of

population parameters may be substantially biased and the statements about the sampling errors of those estimates may be very misleading". These authors, however, recognised that if the model is acceptable or sufficiently so, the model based approach results in substantial simplification and improvement in variance estimates, see for instance, HMT(1983, p.780).

However, Royall and his fellow workers recognised this problem and they have considered sampling plans and estimators so as to make the models more robust to misspecifications. For example, Royall and Cumberland (1981a) have recommended the use of robust model based variance estimators based either on regression or residual or jack-knife methods in order to handle such shortcomings. Regarding this issue, Smith (1984) says:"...If the population is stratified then models are fitted within strata. If there is relationship between a survey variable y and a prior (design) variable z then a ratio or regression estimator is employed. This estimator is protected against misspecification of the mean by using a balanced sample and against misspecification of the variance structure by using robust estimators of variance".

Cumberland and Royall (1983) assert that "For the purpose of inference, quantities such as bias and variance should be defined in terms of prediction models conditioned on the ancillary statistics s, rather than in terms of the probability distribution of s. The unconditional variances, which do not depend on the sample actually selected, are not suitable for making inferences about T from the unique sample actually observed".

2.4.3 *Asymptotic Distribution*

Statistical inference requires more than simply an unbiased estimator and its standard error. It necessitates a relationship between these characteristics. The prediction approach is based on the probability models for the population values. These models, not the random sampling distribution, provide the basic probabilities. The asymptotic properties are determined by making an appeal to the Lindeberg-Feller conditions as $N \to \infty$, $n \to \infty$ and $f \to 0$, where $f = \frac{n}{N}$.

2.4.4 *Conditional Inference*

The model based approach to inference is concerned with the prediction of the actual outcome of the population characteristics and is made conditional on the outcome of the sampling. The possibility of conditioning may be desirable when sampling units are known to differ in relative importance. Royall and Cumberland (1981a, 1983) assert that inferences must be made conditionally on observable characteristics of the sample drawn, and they require assumptions regarding the population structures which are most naturally expressed through prediction models.

So far as making the inferences after the data have been observed, there is a clear difference between the model based inference, which conditions on the achieved sample and the randomisation based inference , which averages over all samples that might have been drawn and this is the most important distinction between these two approaches to inference, see for instance, Smith (1983b). The likelihood principle also implies that inferences should not be based on what Basu (1971) termed as "artifact of randomisation". Smith (1984) says: "Even when the sampling scheme and estimator are similar under both approaches, inferences from

a model will often differ from randomisation inferences. Under a model, when design variables $\underline{z}$ are known the selection rule $P(s/\underline{z})$ can be ignored and inferences made conditional on the units in the sample. Randomisation inferences are unconditional and average overall samples that might have been drawn. There are no principles which lead to conditional inference in the randomisation approach and this is the one of the major problems which must be faced".

The model based approach to inference has the attraction that it embraces both descriptive and analytic inferences and can also be extended to accommodate missing values and non-response, see for instance, Little (1982). It should, however, be noted that the model based procedures can be highly effective when the model is correct but must be used with caution because small deviations from an assumed model may lead to serious errors in inferences, see for example, Hansen et al., (1983). However, if the model is true, then conditional intervals will have the correct conditional coverage and also the correct average coverage under the repeated sampling.

Additional discussion of model based versus randomisation based inference from finite populations can be found, for instance, in Smith (1976, 1983a,b, 1984), Royall (1976), Cassel et al., (1977), Särndal (1978, 1981), O'Muircheartaigh and Wong (1981), Royall and Cumberland (1981a, b, 1982, 1983), Hansen, Madow and Tepping (1983), Kalton (1983) and Sundberg (1983).

2.5 Summary and Conclusions

We have discussed the randomisation and the model based approaches to inference for finite population. Our evaluation of these approaches is based on the criteria for estimation, variance, asymptotic distribution and inferences. Both these approaches are well developed in these aspects.

We observe that there is now general agreement that randomisation has an important role in the selection of samples, the only conflict is whether the randomisation is appropriate for inferences after the sample has been observed. The randomisation inferences are made according to the randomness generated by the sampling design. The randomisation inferences are unconditional and the validity of these inferences is based on the large sample approximations. These inferences are not free from assumptions and their usefulness is limited in scope. Besides, the randomisation approach does not address the question of efficiency and has failed to provide minimum variance unbiased estimators.

The model based approach considers the structure of the population and provides a relationship between the observed and the unobserved units. The quantitites such as bias and variances are conditioned on the observed sample rather than on all possible samples that might have been drawn. The inferences concerning the prediction of the actual outcome of the population characteristics are made conditional on the sample actually observed. This approach has provided robust variance estimators which have proved to be superior to the design-based estimators. Our observation is that the model based approach has proved to be a useful and meaningful tool for the analysis of finite population sampling problems. This approach is heavily relied on in this study.

CHAPTER 3

ESTIMATION OF VARIANCE FROM COMPLEX SURVEYS

3.1 Introduction

The main objective in any sample survey is to draw inferences from the characteristics of a sample of elements about the corresponding population characteristics such as mean, total and ratios, etc. To this end it is also necessary to estimate the precision of the estimator used. The precision of an estimator is determined by the variance if the estimator is unbiased and by the mean square error if the estimator is biased. The estimation of variance is, therefore, an important aspect of the analysis of survey data.

The variance of an estimator, however, depends on the framework for inference. There are currently two well developed theories of inference for finite populations, i.e., the randomisation based theory and the model based theory. The main difference between these approaches lies in the estimation of variance because they use different probability frameworks for defining variance. We have discussed both these approaches in greater detail in Chapter 2.

In this chapter, we consider variance estimation procedures under both the randomisation based and the model based approaches. In Section 3.2, we review the methods of variance estimation under the randomisation based approach. Our main concern is with the jack-knife procedures, however, we introduce some useful alternative competing techniques in order to facilitate comparison. We also introduce a useful technique, known as the weighted jack-knife, developed recently as an alternative

to the ordinary jack-knife for variance estimation. During recent years the prediction approach has posed a big challenge to the randomisation approach. In Section 3.3, we discuss methods of variance estimation under this approach. Finally, in Section 3.4 we give a summary.

3.2 Randomisation Based Methods of Variance Estimation

In the randomisation approach to survey sampling the concept of sampling design is the central one. The design specifies the way in which a sample is drawn from the finite population. Most of the available literature under this approach is confined to the estimators of the variances of descriptive statistics such as totals, means and the difference between means, etc. For simple random sampling, estimates of the variances of more complex statistics such as correlation and regression are readily available. Literature also exists for complex analytic statistics like simple, partial and multiple correlation and regression coefficients, etc., dealing with relationships among the variables and other complicated functions of sample observations for complex designs, see for instance Deming (1960), Tepping (1968), Woodruff (1971), Kish and Frankel (1974) and Woodruff and Causey (1976).

There are, however, some situations where no estimates of variance could be made from a single sample unless certain assumptions are made about the surveyed populations. Such situations arise with probability sampling where no replication is used, for example, in systematic sampling with a single random start or in sampling with one selection per stratum.

3.2.1 Variances of the Estimates

Consider a finite population $Y_1, Y_2, \ldots, Y_N$. Let $\hat{T} = N\bar{y}$ be the estimate of the population total. $T = \sum_{i=1}^{N} y_i$. Then, under simple random sampling, the variance of $\hat{T}$ is given by:

$$V(\hat{T}) = E_p(\hat{T}-T)^2 = \frac{N^2S^2}{n}(1-f) , \qquad (3.1)$$

where

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i-\bar{Y})^2$$

is the population variance and $f = \frac{n}{N}$ is the finite population correction factor.

The expression (3.1) for $V(\hat{T})$ is used for three main purposes,

(i) to assess the precision of the estimate $\hat{T}$.

(ii) to facilitate a comparison of $\hat{T}$ with other estimators of T, and

(iii) to determine the size of the sample needed to yield a desired precision.

However, in practice, S^2 is not known and in order to make use of the sampling variance (3.1) we must estimate S^2 from the sample data. Using a simple random sample $y_1, y_2, \ldots, y_n$, we can obtain an unbiased estimate $s^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2/(n-1)$ of S^2, which when substituted in (3.1) gives an unbiased estimate of the variance of $\hat{T}^2$, that is

$$\hat{V}(\hat{T}) = N^2(1-f)\frac{s^2}{n} . \qquad (3.2)$$

From the above formulation we note that there are two variances for an estimator $\hat{T}$ of T, namely

(a) $V(\hat{T})$, the true variance, and

(b) $\hat{V}(\hat{T})$, the estimator of the true variance.

In extension of this simple situation, survey samplers have developed various devices to exploit their background knowledge of the sampling situation. Some approaches incorporated this knowledge into the design whereas others into the estimator. Consequently, most of the surveys today are based on complex designs which include stratification, clustering and multistage selection of elements and the estimation procedures that involve ratio, correlation and regression coefficients, and discrimination functions. The usual practice has been to adopt the following procedures for obtaining expressions for true variances for a given estimate.

(1) Text Book Variances, (2) Taylor Series Method. (3) Design Effects.

3.2.1.1 Text Book Variances

This means to take an appropriate formula derived directly from the randomisation distribution for $\mathrm{Var}(\hat{T})$. We illustrate this point through examples 3.1 and 3.2.

Examples 3.1

Consider a stratified simple random sampling design. For the population mean per unit, the estimate used in stratified simple random sampling is $\bar{y}_{st}$,

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} = \sum_{h=1}^{L} W_h \bar{y}_h \quad , \tag{3.3}$$

where

$\bar{y}_h = \sum_{i=1}^{n_h} y_{hi}/n_h$ is the sample mean for the hth stratum,

$W_h = \frac{N_h}{N}$ is the stratum weight,

y_{hi} is the value obtained for the ith unit in the hth stratum and n_h is the number of units in the sample for stratum h . From Cochran (1977), the variance of the estimate $\bar{y}_{st}$ is

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{n_h} (1-f_h) \quad , \tag{3.4}$$

where

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi}-\bar{Y}_h)^2}{(N_h-1)} \quad ; \text{ is the true variance,}$$

$$f_h = \frac{n_h}{N_h} \quad , \text{ is the sampling fraction in the stratum, and}$$

$$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h} \quad , \text{ is the true mean.}$$

Example 3.2

As another example of the text book variance we consider probability proportional to size sampling plan. A well known general estimate of the population total for unequal probability sampling without replacement is

$$\hat{Y} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \quad , \tag{3.5}$$

where

π_i = probability that the ith unit is in the sample, and

π_{ij} = probability that the ith and jth units are both in the sample.

If $\pi_i > 0$, $(i = 1,2,\ldots,N)$, then from Cochran (1977) the variance of $\hat{Y}$ is

$$V(\hat{Y}) = \sum_{i=1}^{N} \frac{(1-\pi_i)}{\pi_i} y_i^2 + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j} y_i y_j \ , \qquad (3.6)$$

where π_{ij} is the probability that units i and j are both in the sample.

For linear estimators the text books provide general formulae for variances for complex designs based on the assumptions of simple random sampling. However, we note that for probability proportional to size sampling plan (example 3.2) the variance depends solely on the quantities π_i and π_{ij}. For complex designs, it may be very difficult to compute the joint inclusion probabilities and hence the variance.

Survey designs are generally based on complex multistage design features that include unequal probabilities of selections from the population, stratification at one or more levels and clustering and we also desire to obtain the variances of such statistics as correlation and regression coefficients and other complicated functions of sample observations. Such situations may cause concern because text books do not provide formulae for the variance of complex statistics, involving relationships from complex surveys. Besides, for ratio and regression estimates the simplest design (SRS), in any event, requires Taylor Series approximation for securing variance formulae. Thus for non-linear estimators we need Taylor Series methods. In addition to this, measures of variability of sample results depend on the sample design and are subject to design effects, see for instance, Kish (1965, p.585).

3.2.1.2 The Taylor Series Method

The Taylor expansion method for approximating the variances of ratio means has been in use for many years. Deming (1960), Kish (1965) and Woodruff (1971) have discussed its use for approximating the variances of other non-linear functions of sample totals. The method is also known as the linearisation technique or the δ method. Tepping (1968) gives a detailed description of this method for approximating the variances of more complex statistics from survey data. Tukey (1964) describes that the method gives an approximate formula for the variance of complex statistic and the approximate estimate of the variance by taking into consideration the varaince of the linear terms of Taylor Series expansion of the statistic.

Suppose that $\underline{Y} = (Y_1, Y_2, \ldots, Y_t)$ is a vector of population parameters for certain variables and $\hat{\underline{Y}} = (\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_t)$ be the corresponding vector of unbiased estimators. Let $f(\underline{Y})$ be the function of the population parameters whose estimate is required and $f(\hat{\underline{Y}})$ be the estimate of $f(\underline{Y})$. We would like to obtain the variance of $f(\hat{\underline{Y}})$. By first order Taylor approximation, we get

$$f(\hat{\underline{Y}}) = f(\underline{Y}) + \sum_{i=1}^{t} (\hat{Y}_i - Y_i) \left. \frac{\partial f(\hat{\underline{Y}})}{\partial \hat{Y}_i} \right|_{\hat{Y}_i = Y_i} + O(n^{-1}) \qquad (3.7)$$

It is assumed that $f(\hat{\underline{Y}})$ and the first order partial derivatives of $f(\hat{\underline{Y}})$ are continuous. Now taking expectations, we have $E\{f(\hat{\underline{Y}})\} \doteq f(\underline{Y})$ and

$$\begin{aligned} V(f(\hat{\underline{Y}})) &= E\left(f(\hat{\underline{Y}}) - E\{f(\hat{\underline{Y}})\}\right)^2 , \\ &\doteq E\left[\sum_{i=1}^{t} (\hat{Y}_i - Y_i) \frac{\partial f(\underline{Y})}{\partial Y_i}\right]^2 \\ &\doteq \mathrm{Var}\left[\sum_{i=1}^{t} (\hat{Y}_i - Y_i) \frac{\partial f(\underline{Y})}{\partial Y_i}\right] . \end{aligned} \qquad (3.8)$$

Hence

$$V\left(f(\hat{\underline{Y}})\right) \doteq \text{Var}\left(\sum_{i=1}^{t} \hat{Y}_i \frac{\partial f(\underline{Y})}{\partial Y_i}\right) \tag{3.9}$$

With complex estimators this method may become cumbersome because, first the statistic has to be expressed in a tractable form and then partial derivatives have to be obtained. Finally numerous variances and covariances have to be computed which depend on the complexity of the design.

A simplification of equation (3.9) in terms of the weighted values of the primary selection has been developed by Woodruff (1971). This simplification is outlined as under:

Let $\hat{Y}_i = \sum_j W_{ij} X_{ij}$, where $\hat{Y}_i$ (the estimate of Y_i) is a weighted average of the original observations, the X_{ij}'s . Usually $W_{ij} = 1/p_j$, where p_j is the probability (prior to draw) of including the unit j in the sample. Now, proceeding as before, we estimate $f(\underline{Y})$ by $f(\hat{\underline{Y}})$ which hereafter for simplicity we denote as f and $\hat{f}$ respectively. Let $D_i = \frac{\partial f}{\partial Y_i}$ evaluated at $\underline{Y}$. Then using the first order Taylor series expansion of $\hat{f}$ about f , we get:

$$\hat{f} \doteq f + \sum_{i=1}^{t} D_i (\hat{Y}_i - Y_i) , \tag{3.10}$$

and

$$V(\hat{f}) \doteq \text{Var}\left(\sum_{i=1}^{t} D_i \hat{Y}_i\right) ,$$

$$\doteq \text{Var} \sum_{i=1}^{t} \left(D_i \sum_{j=1}^{n} W_{ij} X_{ij}\right) ,$$

$$\doteq \text{Var}\left(\sum_{j=1}^{n} z_j\right) ,$$

where $$z_j = \sum_{i=1}^{t} D_i W_{ij} X_{ij} . \tag{3.11}$$

According to this simplified version $V(f(\hat{\underline{Y}}))$ can be estimated from $\mathrm{Var}\left(\sum_{j=1}^{n} z_j\right)$ by applying the same variance formula to the z_j's that would be applied to $w_{ij}x_{ij}$'s to compute $\mathrm{Var}(\hat{Y}_i)$ for each variable i. This simplification avoids the computation of t variances of $\hat{Y}_i$ and $t(t-1)/2$ covariances of $\hat{Y}_i$ and Y_i. The variance estimate is obtained by estimating the z_j's from the sample and then applying the variance estimation formula for a single variate. The Taylor expansion method is general and applicable to any design provided the variance estimation formula for estimated total is known, see for instance, Rao (1975). In order to illustrate the method, we give below an example of variance formula for the regression coefficient from Tepping (1968).

Example 3.3. (Variance for Regression Coefficient from a Complex Design)

We consider the variance of the regression coefficient from a self-weighting stratified multistage design. Let x_{hi}, y_{hi} be the values of the variables x and y associated with the ith elementary sampling unit selected for the sample in stratum h. We consider the statistic

$$b = \frac{\frac{1}{n}\sum_{h=1}^{L}\sum_{i=1}^{n_h} x_{hi}y_{hi} - \left(\frac{1}{n}\sum_{h=1}^{L}\sum_{i=1}^{n_h} x_{hi}\right)\left(\frac{1}{n}\sum_{h=1}^{L}\sum_{i=1}^{n_h} y_{hi}\right)}{\left(\frac{1}{n}\sum_{h=1}^{L}\sum_{i=1}^{n_h} x_{hi}^2\right) - \left(\frac{1}{n}\sum_{h=1}^{L}\sum_{i=1}^{n_h} x_{hi}\right)^2}, \qquad (3.12)$$

where $n = \sum_{h=1}^{L} n_h$ denotes the number of elementary units in the sample. Our main concern here is to approximate the variance of b which is achieved by first expressing the statistic given by (3.12) in a tractable form, obtaining the partial derivatives and finally applying the formula (3.9).

Let $\underline{u} = (u_1, u_2, \ldots, u_t)$ be a vector of statistics such that the expected value of $u_k (k = 1,2,\ldots,t)$ is U_k. Let $(U_1, \ldots, U_t)$ be the corresponding vector of population parameters. Suppose that the population parameter of interest is a function $f(\underline{U})$ and is to be estimated by $f(\underline{u})$, then

$$V(f(\underline{u})) \doteq \mathrm{Var}\left(\sum_{i=1}^{t} u_i \frac{\partial f(\underline{u})}{\partial u_i} \right) . \tag{3.13}$$

Now in order to obtain the variance of b we express(3.12) in a suitable form so as to derive the partial derivatives. To achieve this we introduce the following variables and notations.

$$w_h = \sum_{i=1}^{n_h} x_{hi}\, y_{hi} , \qquad w = \sum_{h=1}^{L} w_h ,$$

$$u_h = \sum_{i=1}^{n_h} x_{hi}^2 , \qquad u = \sum_{h=1}^{L} u_h ,$$

$$x_h = \sum_{i=1}^{n_h} x_{hi} \qquad x = \sum_{h=1}^{L} x_h ,$$

$$y_h = \sum_{i=1}^{n_h} y_{hi} \qquad y = \sum_{h=1}^{L} y_h . \tag{3.14}$$

The regression coefficient b may be written as

$$b = \frac{nw - xy}{nu - x^2} \tag{3.15}$$

Thus b now represents $f(\underline{u})$ in expression (3.13). Denoting N, W, U, X, Y as the expectations of n, w, u, x, y respectively, the

required partial derivatives for b are

$$\frac{\partial b}{\partial N} = \frac{X(UY-WX)}{(NU-X^2)^2}, \qquad \frac{\partial b}{\partial W} = \frac{N}{NU-X^2},$$

$$\frac{\partial b}{\partial U} = -\frac{N(NW-XY)}{(NU-X^2)^2}, \qquad \frac{\partial b}{\partial X} = \frac{-NUY+2NWX-X^2Y}{(NU-X^2)^2}$$

$$\frac{\partial b}{\partial Y} = -\frac{X}{NU-X^2}, \tag{3.16}$$

Using (3.13), the variance of b is now approximated by

$$\mathrm{Var}(b) \doteq \mathrm{Var}\left(n\frac{\partial b}{\partial N} + w\frac{\partial b}{\partial W} + u\frac{\partial b}{\partial U} + x\frac{\partial b}{\partial X} + y\frac{\partial b}{\partial Y}\right),$$

$$\doteq \sum_{h=1}^{L} \mathrm{Var}\left(\frac{\partial b}{\partial N}n_h + \frac{\partial b}{\partial W}w_h + \frac{\partial b}{\partial U}u_h + \frac{\partial b}{\partial X}x_h + \frac{\partial b}{\partial Y}y_h\right). \tag{3.17}$$

The main drawback of the Taylor Series method is that the evaluation of the partial derivatives may be difficult for certain parameters (e.g. multiple and partial correlation coefficients). A useful approximation to the required partial derivatives could be obtained by using the numerical method developed by Woodruff and Causey (1976). However, the results obtained for the partial and multiple correlation coefficients by using this numerical method are not very promising.

The first order Taylor Series expansion for a multiple regression coefficient has been derived by Folsom (1974) and Fuller (1974). Programmes for Taylorised Standard errors are available from Hidiroglou et al., (1975), Kish et al., (1972), Holt (1977) and Woodruff and Causey (1976).

3.2.1.3 Design Effects:

In order to deal with the complex designs mentioned earlier in this Section, Kish (1865) proposed a useful device known as the design effect (Deff). He describes this as the ratio of the variance of the estimate obtained from the complex sample to the variance of the estimate obtained from a simple random sample of the same number of units. Thus,

$$\text{Design Effect (Deff)} = \frac{\text{Var}(\hat{T}) \text{ from a Complex Sample}}{\text{Var}(\hat{T}) \text{ from Simple Random Sample}} \qquad (3.18)$$

Stratification tends to reduce the variance and the design effect may be less than one. On the other hand, clustering tends to increase the variance and the design effect is usually greater than one. Consider, for instance, a simple random sample of n clusters, each containing M elements, drawn from N clusters in the population. Then, the variance of the sample mean per element is

$$V(\bar{y}) \doteq \frac{S^2}{nM}\left[1 + (M-1)\rho\right] \qquad (3.19)$$

where M is the constant cluster size, ρ is the intracluster correlation coefficient and $\frac{S^2}{n}$ is the simple random sample variance. It is apparent from (3.19) that the variance is inflated by the factor $(M-1)\rho$ by using clusters instead of elements as sampling units. The factor $1 + (M-1)\rho$ is called the design effect or Kish's Deff for clusters of size M.

The Deff attempts to summarise the complexities of the sample design, especially those of clustering and stratification. It may even include the effects of ratio and regression estimation, of double sampling and of varied sampling fractions (Kish, 1965).

3.2.1.3(a) Modified Deff:

In order to adjust for changes in structure, a modified Deff could be used following Kish (1965, p.162-163). For a complex design, suppose that the true variance is

$$\mathrm{Var}(\bar{y}) = \mathrm{Var}_{SRS}(\bar{y})\ [1+(M-1)\mathrm{roh}] \tag{3.20}$$

where roh measures the homogeneity and is estimated from the past survey

$$\mathrm{roh} = \left[\frac{\mathrm{Var}(\bar{y})}{\mathrm{Var}_{SRS}(\bar{y})} - 1\right] \frac{1}{(M-1)} \tag{3.21}$$

Thus for a new survey with cluster size ℓ, we can use

$$\mathrm{Var}(\bar{y}) = \mathrm{Var}_{SRS}(\bar{y})[1+(\ell-1)\mathrm{roh}]\,. \tag{3.22}$$

The purpose of this modification is that "roh" has distinct values for diverse variables and for different populations. For a specified variable and population, its value depends on the nature and the size of clusters. It is a characteristic of the clusters and affects the variances when the clusters serve as sampling units. Thus by studying "roh", the aim is to remove the effect of average cluster size when comparing results across different variables and populations, see for instance O'Muircheartaigh et al., (1980, p.444-45).

3.2.2 Methods for Estimating the Variances of Estimates

When variance estimates are required for statistical inference, the randomisation approach offers several choices. However, the usual practice is to use the methods discussed under:

3.2.2.1 To Use Text Book Unbiased Estimators of the Variances if Available

The adoption of this procedure with complex surveys gives rise to the following problems:

(a) An unbiased estimator may not be available for the variance of certain classes of statistics such as ratio estimators, correlation and regression coefficients.

(b) The use of available unbiased estimators may, in some cases, involve complexities. Consider for example, the Horvitz-Thompson (1952) estimator for the population total

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \tag{3.23}$$

which is an unbiased estimator of Y with variance (Cochran, 1977).

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{N} \frac{(1-\pi_i)}{\pi_i} y_i^2 + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} y_i y_j \ , \tag{3.24}$$

where π_{ij} is the probability that units i and j both are in the sample. An unbiased sample variance estimator of $V(\hat{Y}_{HT})$ is given by

$$\hat{V}(\hat{Y}_{HT}) = \sum_{i=1}^{n} \frac{(1-\pi_i)y_i^2}{\pi_i^2} + 2 \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j\pi_{ij}} y_i y_j \ , \tag{3.25}$$

provided that none of the π_{ij} in the population vanishes, where π_i = probability that the ith unit is in the sample. A more elegant form of the variance estimator of $V(\hat{Y}_{HT})$ given by Yates and Grundy (1953) and Sen (1953) is

$$\hat{V}_{YG}(\hat{Y}_{HT}) = \sum_{i=1}^{n} \sum_{j>i} \frac{(\pi_i\pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 , \qquad (3.26)$$

with the same restriction on π_{ij} .

The drawback with these variance estimators is that the terms $(\pi_i\pi_j - \pi_{ij})$ often vary widely and both $\hat{V}_{HT}$ and $\hat{V}_{YG}$ can take negative values for some samples and some sets of π_{ij}'s although $\hat{V}_{YG}$ is generally less likely to do so, see, for example, Yates and Grundy (1953), Raj (1956), Rao and Singh (1973), Lanke (1974), Vijayan (1975), Cochran (1977) and Rao (1979). There are, however, many inclusion probability proportional to size sampling plans corresponding to different possible sets of joint inclusion probabilities $\{\pi_{ij}\}$. For a given sample, the expected value of $\hat{V}_{HT}$ and $\hat{V}_{YG}$ depends on the nature of the joint inclusion probabilities corresponding to units that happen to be included in the sample. It has been noted that these sets of π_{ij}'s are not uniquely determined, for instance, Brewer and Hanif (1969) have catalogued about 34 exact and approximate methods for implementing the inclusion probability proportional to size sampling plan. This diversity of possibilities as well as the mathematical intractibility of many of these methods leads to complexities.

3.2.2.2 Substitution Method

According to this method population parameters in the variance formula are replaced by their corresponding sample estimates. In the simplest cases the variance of the statistic, for a given design, is a known function of certain population parameters which are themselves estimated from the sample and substituted for the parameters in the variance function so as to obtain the estimate of the variance. This is illustrated through a simple example hereunder:

Example 3.4 (Variance Estimation for the Sample Mean).

For a simple random sample of n units from a population of N units, without replacement, the variance of the sample mean is

$$\sigma^2_{\bar{x}} = \frac{N-n}{N}\frac{\hat{S}^2}{n} ,$$

where

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2 \qquad (3.27)$$

The population variance S^2 may be estimated by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad (3.28)$$

Thus, in the case of this example, if the variance function is linear in the parameters and the estimates of these parameters are unbiased, we get an unbiased estimator of the variance. If the function is rational, we get at least a consistent estimator of the variance. This is illustrated through example 3.5.

Example 3.5 (Variance Estimation for the Ratio Estimator)

For simple random sampling, the variance of the ratio estimator, $\hat{T}_R = \frac{\bar{y}_s}{\bar{x}_s}\sum_{i=1}^{N} x_i$, for the population total T is approximated by

$$V(\hat{T}_R) = \frac{N^2(1-f)}{n}(S_Y^2+R^2S_X^2 - 2RS_{XY}) \quad , \qquad (3.29)$$

where, $R = \frac{\bar{Y}}{\bar{X}}$, $\bar{X}$ and $\bar{Y}$ being the expected values of $\bar{x}_s$ and $\bar{y}_s$ respectively, and $S_{XY} = \rho S_X S_Y$ is the covariance between y_i and x_i (Cochran, 1977). The usual practice is to estimate $V(\hat{T}_R)$ by

$$\hat{V}_1(\hat{T}_R) = \frac{N^2(1-f)}{n}\left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}\right) , \qquad (3.30)$$

where s_x^2 , s_y^2 and s_{xy} are the sample estimates for the population variances and covariance. There is an alternative sample variance estimator of $V(\hat{T}_R)$ which is given by

$$\hat{V}_2(\hat{T}_R) = \frac{N^2\bar{x}^2(1-f)}{n\bar{x}_s^2}\left(s_y^2+\hat{R}^2 s_x - 2\hat{R}s_{xy}\right) . \qquad (3.31)$$

The variance estimator $\hat{V}_2(\hat{T}_R)$ appeared in the first edition of Yates (1949, p.213). Since the merits of $\hat{V}_1(\hat{T}_R)$ and $\hat{V}_2(\hat{T}_R)$ were not quite obvious under probability sampling theory, Yates preferred $\hat{V}_2(\hat{T}_R)$ for his book. Cochran (1963, p.163) chose $\hat{V}_1(\hat{T}_R)$ for his well known book. However, Cochran (1977, p.155) remarks that it is still not clear whether $\hat{V}_1(\hat{T}_R)$ is preferable to $\hat{V}_2(\hat{T}_R)$. Kish (1965, p.204) and Murthy (1967, p.373) obtained $\hat{V}_2(\hat{T}_R)$ in their well known books by noting that $\hat{T}_s = \hat{R}_s(N\bar{x})$ and $\hat{V}(\hat{T}_s) = N\bar{x}^2\hat{V}(\hat{R}_s)$ using the customary formula for $\hat{V}(\hat{R}_s)$. Sukhatme and Sukhatme (1970, p.144) recommended the analogue of $\hat{V}_2(\hat{T}_R)$ for estimating the variance of the ratio means but used the analogue of $\hat{V}_1(\hat{T}_R)$ for a ratio adjusted estimate of the population total. Rao and Rao (1971b) wrote that when $\bar{X}$ is known the commonly used varinace estimator is $\hat{V}_1(\hat{R}_s)$ and when $\bar{X}$ is not known the variance estimator is $\hat{V}_2(\hat{R}_s)$. Thus to Rao and Rao, it appears that $\hat{V}_1$ is the estimator of choice with $\hat{V}_2$ considered to be not a competitor but a substitute to be used only when $\hat{V}_1$ was not available. Konijn (1973, p.48) in his book considered both $\hat{V}_1$ and $\hat{V}_2$ and remarked that if $\bar{X}$ is known $\hat{V}_2$ is preferable to $\hat{V}_1$. Konijn (1973, p.353) also remarks that Hájek (1958) showed $\hat{V}_2$ to be a good estimator of the conditional variance.

For more complex analytic statistics like correlation and regression coefficients or for more complex designs the variance may not be a known function of the parameters that are easily estimated for substitution into the variance formula, see for instance, Cochran (1977, 315-318), Sukhatme and Sukhatme (1970, p.180-83) and Tepping (1968). In order to illustrate this point, we give below an example of the variance estimation formula for the regression coefficient (3.12).

Example 3.6 (Variance Estimation for Regression Coefficient From a Complex Design).

From (3.17) we have

$$\mathrm{Var}(b) = \sum_{h=1}^{L} \mathrm{Var}(\ell_h) \ , \tag{3.32}$$

where the linear combination ℓ_h is ,

$$\ell_h = \frac{\partial b}{\partial N} n_h + \frac{\partial b}{\partial W} w_h + \frac{\partial b}{\partial U} u_h + \frac{\partial b}{\partial X} x_h + \frac{\partial b}{\partial Y} y_h \ .$$

The coefficients in the linear form ℓ_h are unknown population parameters. The usual practice, as explained earlier in this section, is to substitute sample estimates in the expressions for the derivatives just as we substituted in example 3.5 . For large samples, this procedure yields satisfactory results.

For estimating the variance of the regression coefficient b as given by (3.12), these substitutions will be

$$\frac{\partial b}{\partial N} \doteq \frac{\bar{x}(\bar{y}-b\bar{x})}{n s_x^2} \ , \qquad \frac{\partial b}{\partial W} \doteq \frac{1}{n s_x^2} \ ,$$

$$\frac{\partial b}{\partial U} \doteq \frac{b}{n s_x^2} \ , \quad \frac{\partial b}{\partial X} \doteq \frac{-(\bar{y}-2b\bar{x})}{n s_x^2} \ \text{ and } \ \frac{\partial b}{\partial Y} = -\frac{\bar{x}}{n s_x^2} \tag{3.33}$$

where $\bar{x} = \frac{x}{n}$, $\bar{y} = \frac{y}{n}$, $\bar{u} = \frac{u}{n}$, $s_x^2 = \bar{u}^{-2} - \bar{x}^{-2}$.

This method of substitution gives the variance estimate for the desired statistic but appears to be a bit cumbersome.

Example 3.7 (Variance Estimation for the Ratio Estimator in Stratified SRS)

Consider stratified simple random sampling without replacement. The ratio estimate is given by,

$$\hat{R} = \frac{y}{x} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{n_h} y^*_{hi}}{\sum_{h=1}^{L} \sum_{i=1}^{n_h} x^*_{hi}} \tag{3.34}$$

where $y^*_{hi} = \frac{N_h}{n_h} y_{hi}$ and $x^*_{hi} = \frac{N_h}{n_h} x_{hi}$.

In this case the derivatives are

$$\frac{\partial \hat{R}}{\partial y^*_{hi}} = \frac{1}{x} , \quad \text{and} \quad \frac{\partial \hat{R}}{\partial x^*_{hi}} = \frac{-y}{x^2} .$$

Now (3.34) can be written as

$$\sum_{h=1}^{L} \sum_{i=1}^{n_h} \left(\frac{\partial \hat{R}}{\partial y^*_{hi}} y^*_{hi} + \frac{\partial \hat{R}}{\partial x^*_{hi}} x^*_{hi} \right) = \sum_{h=1}^{L} \sum_{i=1}^{n_h} \frac{N_h}{n_h} \left(\frac{y_{hi}}{x} - \frac{y}{x^2} x_{hi} \right) .$$

Letting $Z_{hi} = \frac{N_h}{n_h} \left(\frac{y_{hi}}{x} - \frac{y}{x^2} x_{hi} \right)$ and $\bar{Z}_h = \left(\frac{\bar{y}_h}{x} - \frac{y}{x^2} \bar{x}_h \right) \frac{N_h}{n_h}$, the variance estimator for $V(\hat{Y}_{RC}) = \sum_{h=1}^{L} \frac{N_h^2 (1-f_h)}{n_h} (S_{yh}^2 + R^2 S_{xh}^2 - 2R\rho_h S_{yh} S_{xh})$ can be obtained as

$$\hat{V}(\hat{Y}_{RC}) = \sum_{h=1}^{L} \frac{N_h^2 (1-f_h)}{n_h} (s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2\hat{R}\hat{\rho}_h s_{yh} s_{xh}) . \tag{3.35}$$

However, with more complex designs and more complex statistics this procedure is laborious and hence the need for other more simple and convenient methods.

3.2.2.3 The Keyfitz Method

This method, also known as the "paired selection" method, requires that two primary selections be drawn from each stratum, see for instance, Keyfitz (1957) and Kish (1965, 4.3, 6.4B). Regarding the applicability of this method Kish and Frankel (1970) pointed out that "a model of two independent primary selections from each stratum is probably the most basic design that conforms adequately, if not perfectly to actual design of many surveys". We outline this method for obtaining the variance estimate for the ratio estimate as example 3.8.

Example 3.8 (Variance Estimation for Ratio Estimate).

Consider a single stage stratified cluster design,

$$\hat{R} = \frac{\hat{Y}_{st}}{\hat{X}_{st}} = \frac{\sum_{h}^{L} \hat{Y}_h}{\sum_{h=1}^{L} \hat{X}_h} = \frac{\sum_{h=1}^{L} \frac{N_h}{2}(y_{h1}+y_{h2})}{\sum_{h=1}^{L} \frac{N_h}{2}(x_{h1}+x_{h2})} \quad . \tag{3.36}$$

where y_{h1} and y_{h2} represent the totals for the y_i variable from the two primary selections in the hth stratum. Similarly x_{h1} and x_{h2} represent the corresponding totals for the variable x_i, $\hat{Y}_{st} = \sum_{h}^{L} N_h \bar{y}_h$ and $\hat{X}_{st} = \sum_{h}^{L} N_h \bar{x}_h$ are the standard estimates of the population total for Y and X respectively, N_h is the number of units in the hth stratum of the population and $\bar{y}_h = \sum_{i=1}^{n_h} y_{hi}/n_h$ and $\bar{x}_h = \sum_{i=1}^{n_h} x_{hi}/n_h$ are the sample means for the h^{th} stratum.

In order to obtain the variance estimate of $\hat{R}$, we define

$$\bar{y}_h = \frac{1}{2}(y_{h1}+y_{h2}) \ .$$

Also,

$$s^2_{yh} = \frac{1}{(n_h-1)} \sum_{i=1}^{n_h} (y_{hi}-\bar{y}_h)^2 = \frac{1}{2}(dy_h)^2 \ , \tag{3.37}$$

where

$$dy_h = (y_{h1}-y_{h2})$$

Now if $\hat{Y}_h = N_h\bar{y}_h$, then

$$\hat{V}(\hat{Y}_h) = \left(\frac{N_h}{2}\right)^2 (1-f_h)(dy_h)^2 = (1-f_h)(d^*_{yh})^2 \ ,$$

where

$d^*_{yh} = \frac{N_h}{2} dy_h$ and $(1-f_h)$ is the finite population correction factor for the h^{th} stratum. Similarly, $\hat{V}(\hat{X}_h) = (1-f_h)(d^*_{xh})^2$ and

$$\hat{Cov}(\hat{Y}_h,\hat{X}_h) = (1-f_h)(d^*_{yh})(dx^*_h) \ . \tag{3.38}$$

Since sampling is independent in different strata, the variance estimate of the stratified ratio mean is given by

$$\hat{V}(\hat{R}) = \frac{1}{\hat{\bar{X}}^2}\left[\sum_h \mathrm{Var}(\hat{Y}_h) + \hat{R}^2 \sum_h \mathrm{Var}(\hat{X}_h) - 2\hat{R}\sum_h \mathrm{Cov}(\hat{Y}_h,\hat{X}_h)\right] \ . \tag{3.39}$$

By substituting from (3.38) and simplifying we get

$$\begin{aligned}\hat{V}(\hat{R}) &= \frac{1}{\hat{\bar{X}}^2}\left[\sum_h^L (1-f_h)d^*_{yh}+\hat{R}^2\sum_h^L (1-f_h)d^*_{xh} - 2\hat{R}\sum_h (1-f_h)(d^*_{yh})(d^*_{xh})\right], \\ &= \frac{1}{\hat{\bar{X}}^2}\sum_{h=1}^L (1-f_h)(d^*_{yh}-\hat{R}d^*_{xh})^2 \ .\end{aligned} \tag{3.40}$$

This could be further simplified as

$$\hat{V}(\hat{R}) = \frac{1}{(\hat{X})^2} \sum_{h=1}^{L} (1-f_h)\left[(y_{h1}-y_{h2}) - \hat{R}(x_{h1}-x_{h2})\right]^2 . \qquad (3.41)$$

This is the Keyfitz variance estimate for the $V(\hat{R}_c)$ of the combined ratio estimator, as given in Example 3.7. The Woodruff simplification, as discussed in Subsection 3.2.1.2, can also be applied here as well.

The design facilitates stratified multistage selection with PPS from unequal clusters. For the linear case, it has the advantage that for a fixed number of selections, it is possible to carry the stratification as far as possible and still compute the variance from the sample. The design also permits simple formulae of computing variance estimates. The application of this method to more complex situations has been discussed by Keyfitz (1957), Kish and Hess (1959) and Kish (1965). Woodruff (1971) generalised the method to handle non linear estimators, unequal probabilities of selection and for sample size n_h in the strata. Bean (1975) and Chanġurde (1981) applied this method to compute the variance estimates for the Canadian Labour Force Survey.

3.2.2.4 Estimation of Variance Using Deffs

In situtations where we can not estimate the variance directly, it is generally assumed that simple random sample variance computed from the large probability sample would yield a good approximation of the true variance. For example, Kish and Frankel (1974) treated proportionate stratified sampling as simple random sampling for computation of complicated estimates.

We have pointed out earlier (Subsection 3.2.1.3) that stratification tends to reduce the variance and Deff may be less than one whereas

clustering inflates the variance and Deff may be greater than one. Under these circumstances one can use $\hat{V}(\bar{x}) = \hat{V}_{SRS}(\bar{x}) \times deff$, where $\hat{V}(\bar{x})$ is the variance incorporating all the complexities of the design, see, for instance, Kish (1965). However, the problem with this expression is how to compute deff. The solution to this problem is to compute the design effect from an old survey with similar structure and for which the variances have been computed.

From the above discussion, we note that for complex surveys it would be incorrect to assume simple random selection and to use the simple formulae available in the standard literature. Moreover, it is well recognised that in complex surveys the assumption of two independent selections is seldom met. Under these circumstances the alternative is to use one of the approximate methods, to be discussed in the next few subsections for obtaining the estimates of true variance from complex surveys. These methods recognise the lack of independence among sample units.

3.2.2.5 Replicated Sampling:

One approach to the problem of variance estimation from complex surveys is to interweave within a sample design a small number of replications. Each of these replications provides an estimate of the parameter. Therefore, their comparison provides an estimate of the variances from the sample. As an example, suppose that an overall sample is selected in the form of r independent subsamples from the same population using the same probability design. Let $\hat{\theta}_i$ denote the sample estimate calculated from the ith sample using the same estimation procedure, then $\bar{\hat{\theta}} = \frac{1}{r} \sum_{i=1}^{r} \hat{\theta}_i$. From r independent estimates, the sample variance of $\hat{\theta}$ is

$$\hat{V}(\bar{\hat{\theta}}) = \sum_{i=1}^{r} (\hat{\theta}_i - \bar{\hat{\theta}})^2 / r(r-1) \quad . \tag{3.42}$$

The main advantage of the replicated sampling is that $\hat{\theta}_i$ can be any statistic, it may be a linear statistic like mean and total or a non-linear statistic such as ratio or regression coefficient. This method provides a simple and convenient way of estimating the variance and the evaluation of bias regardless of the complexity of the sample design within replicates and the complicated form of the estimator.

The replicated technique was first introduced by Mahalnobis (1938-41, 1946) and was later developed more fully by Deming (1950, 1960, 1963). Koop (1960, 1962, 1967), Lahiri (1958), McCarthy (1966, 1969), Mokashi (1950) and Srikantan (1963) Madow and Madow (1944) and Tukey (1949), also used similar procedures in connection with selection of systematic samples with multiple random starts. Other references are Bailar and Dalenius (1969) and Richard and Freeman (1979).

The replicated sampling has the limitations that fewer replicates means smaller numbers of degrees of freedom for the variance estimator which could lead to unstable estimation. On the other hand, larger numbers of replications will increase the cost and decrease the efficiency.

3.2.2.6 Random Group Method

Sometimes when the sample is large, it is convenient to use a slight variation of the replicated method known as random group method. Suppose that a simple random sample is drawn with replacement. Let these units be divided randomly into g groups consisting of n/g observations each. The variance estimate is given by

$$\hat{V}(\bar{\hat{\theta}}) = \sum_{j=1}^{g} (\hat{\theta}_j - \bar{\hat{\theta}})^2/g(g-1) \ , \tag{3.43}$$

where

$\hat{\theta}_j$ = sample estimate computed from the jth random group and

$$\bar{\hat{\theta}} = \sum_{j=1}^{g} \hat{\theta}_j/g \ .$$

This method is also suitable for multistage survey samples. However, if there are few primary selections (PSU's), then this procedure may also not be very useful. Moreover, there is a loss of information. Murthy (1967) shows that the random group variance estimator $\hat{V}_g(\bar{y})$ provides an unbiased estimate of $V(\bar{y})$ but is less efficient than $\hat{V}(\bar{y})$.

3.2.2.7 Pseudo-Replication Techniques

In order to overcome the difficulties described above, McCarthy (1966) proposed the method of half-sample replication called the "pseudo-replication" technique for estimating variances from stratified samples in which two primary selections are drawn from each stratum. A half-sample is formed by selecting one of the two primary selections in each stratum and employing the estimates of interest from the L selected PSU's. There are in all 2^L such half samples.

Consider a stratified sampling design with two independent selections per stratum. We assume that the selections are made with replacement. Let y_{h_1} and y_{h_2} be the observations from stratum $h(h = 1,2,\ldots,L)$, then the unbiased estimate of the population mean $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{3.44}$$

where $W_h = N_h/N$ and $\bar{y}_h = \frac{y_{h1}+y_{h2}}{2}$. The usual variance estimator is

$$\hat{V}(\bar{y}_{st}) = \frac{1}{4}\sum_{h=1}^{L} W_h^2 d_h^2 , \tag{3.45}$$

where $d_h = (y_{h1}-y_{h2})$. The variance estimate based on independent replications $(y_{11},\ldots,y_{L1})$ and $(y_{12},\ldots,y_{L2})$ is

$$\hat{V}_R(\bar{y}_{st}) = \frac{1}{4}(\bar{y}_{st1}-\bar{y}_{st2})^2 . \tag{3.46}$$

Now suppose a half-sample replicate is obtained by choosing one of (y_{11},y_{12}), one of $(y_{21},y_{22}),\ldots,$ and one of (y_{L1},y_{L2}) . Let us define the indicator function δ such that

$$\delta_{h1} = \begin{cases} 1 , & \text{if } y_{h1} \text{ is selected for the half-sample} \\ 0 , & \text{otherwise} \end{cases}$$

$$\delta_{h2} = 1 - \delta_{h1} \tag{3.47}$$

The half sample estimate of $\bar{Y}$ is given by

$$\bar{y}_{hs} = \sum_{h=1}^{L} W_h(\delta_{h1}y_{h1}+\delta_{h2}y_{h2}) . \tag{3.48}$$

There are 2^L possible half-samples for a given sample and if $\bar{y}_{hs,i}$ denotes the estimate for the ith half-sample, $i = 1,2,\ldots,2^L$, then

$$\bar{y}_{st} = \frac{1}{2^L}\sum_{i=1}^{2^L} \bar{y}_{hs,i} . \tag{3.49}$$

Let,

$$\delta_h^{(i)} = 2\delta_{h1}^{(i)} - 1 = \begin{cases} +1 & \text{if } y_{h1} \text{ is in the ith half sample} \\ -1 & \text{if } y_{h2} \text{ is in the ith half sample} \end{cases}$$

then $\frac{1}{2^L}\sum_{h=1}^{L}\delta_h^{(i)} = 0$ and

$$(\bar{y}_{hs,i}-\bar{y}_{st})^2 = \frac{1}{4}\sum_{h=1}^{L} W_h^2(\delta_h^{(i)})^2 d_h^2 + \frac{1}{2}\sum_{h<j}\sum \delta_h^{(i)}\delta_j^{(i)} W_h W_j d_h d_j \ . \qquad (3.50)$$

Now

$$\frac{1}{2^L}\sum_{i=1}^{2^L}(\bar{y}_{hs,i}-\bar{y}_{st})^2 = \frac{1}{4}\frac{1}{2^L}\sum_{i=1}^{2^L} W_h^2(\delta_h^{(i)})^2 d_h^2 + \frac{1}{2}\sum_{h<j}\sum d_h d_j W_h W_j\left[\frac{1}{2^L}\sum_{i=1}^{2^L}\delta_h^{(i)}\delta_j^{(i)}\right] \qquad (3.51)$$

Since, $(\delta_h^{(i)})^2 = 1$ and $\frac{1}{2^L}\sum_{i=1}^{2^L}\delta_h^{(i)}\delta_j^{(i)} = 0$, therefore

$$\frac{1}{2^L}\sum_{i=1}^{2^L}(\bar{y}_{hs,i}-\bar{y}_{st})^2 = \frac{1}{4}\sum_{h=1}^{L} W_h^2 d_h^2 = \hat{V}(\bar{y}_{st}) \ . \qquad (3.52)$$

Thus from (3.45) and (3.52), we note that there is no loss of information if all 2^L half sample replicates are used. However, when L is large 2^L will be very large. This reflects the limitation of the half-sample procedure because the precision of variance estimate depends on the number of replications and sufficient number of replications may be costly. The current population survey conducted by the U.S. Bureau of Census provides one such example.

If the primary selections within a stratum are made by simple random sampling without replacement, then (3.48) can be written as

$$\bar{y}_{hs} = \sum_{h=1}^{L} W_h\left(1-\frac{2}{N_h}\right)^{\frac{1}{2}}\left\{(\delta_{h1}y_{h1}+\delta_{h2}y_{h2}) - \bar{y}_h\right\} + \bar{y}_{st} \ , \qquad (3.53)$$

and

$$(\bar{y}_{hs,i}-\bar{y}_{st})^2 = \hat{V}(\bar{y}_{st}) + \frac{1}{4}\sum_{h<j}\sum\left(1-\frac{2}{N_h}\right)\delta_h^{(i)}\delta_j^{(i)} W_h W_j d_h d_j \ . \qquad (3.54)$$

Several modification of the half-sample method such as balanced half-sample and partially balanced half-sample replications have been developed. These are described as under:

3.2.2.7.A Balanced Repeated Replications (BRR)

For a stratified random sample with L strata and two primary selections per stratum, it is shown that the variability among the half samples arises from the between strata contribution to these estimates. This variability comes from the crossproduct terms of $d_h d_j$ which cancel one another over the entire set of 2^L half samples or if we use an infinite number k of replications. McCarthy (1966) suggested a way of controlling the selection of half samples so that the variability of the variance estimator is reduced.

McCarthy suggested that if an orthogonally balanced set of half samples is selected using Placket and Burman (1946) matrices, then it would give the same precise estimate of variance as would have been obtained from a complete set. The method was later more fully developed by Kish and Frankel (1968, 1970), who in their repeated replication scheme for the varinace estimation used both the principal set of selected half-samples and the associated complementary set. These authors believe that the principal set of half-samples in the repeated replication scheme is best chosen to be "balanced" and the scheme is termed as Balanced Repeated Replication (BRR) method.

A Placket and Burman matrix is defined as K x K matrix of +1 and -1's. The elements of final column are all -1 . The dimensions of a matrix to be used in a given problem would depend upon the number of strata. Let K be the number of half samples selected, then K has the restriction that it must be a multiple of 4 and $L \leq K \leq L + 3$, where in all cases $L > 2$ and $K << 2^L$. Let δ_{ij} denote the element

in the ith row and the jth column. The first (K-1) columns of this matrix possess the property of balance, i.e.,

$$\sum_{i=1}^{K} \delta_{ij} = 0 \ , \ j = 1,2,\ldots,K-1. \tag{3.55}$$

The last column is not balanced, since

$$\sum_{i=1}^{K} \delta_{ij} = -K \quad \text{for} \quad j = K \ . \tag{3.56}$$

These matrices have the additional property of orthogonality, i.e.,

$$\sum_{i=1}^{K} \delta_{ij}\delta_{ij'} = 0 \quad \forall \quad j \neq j' \tag{3.57}$$

For L = 5, 6, 7 or 8, the smallest orthogonal matrix that can be used is 8 x 8. In order to illustrate this point we give here a matrix M for L = 5 constructed such that the columns (strata) of the matrix are orthogonal to each other. In a 8 x 8 matrix below, the rows identify the half samples and the columns refer to strata.

Matrix M Used to Form 5 Strata Design:-

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

For L = 5, 6, or 7 any set of 5, 6 or 7 columns can be selected whereas for L = 8 the full 8 x 8 matrix must be used. The variance estimator is, $\frac{1}{K}\left\{\sum_{i=1}^{K}(\bar{y}_{hs,i}-\bar{y}_{st})^2\right\}$, where for the number of strata between 5 and 8, K = 8.

The BRR technique is particularly useful for complicated non-linear stratistics such as regression coefficients and the estimation involving complex adjustments. Frankel (1971) and Kish and Frankel (1974) used this technique for estimating the variances of correlation and regression coefficients. Properties of the BRR technique have been investigated by various authors for a variety of estimates, for instance, Bean (1975), Campbell and Meyer (1978), Fenech (1983), Kalton (1977), Krewski (1978a) Krewski and Rao (1978, 1981), Lemeshow and Epp (1977), Lemeshow and Levy (1979), Stanek III and Lemeshow (1977), McCarthy (1966, 1969a, b), Maurer, Jones and Brynt (1977), Rao, J.N.K. (1975), Shah, Holt and Folsom (1977), Simmons and Baired (1968) and Tepping (1968). Koch and Lemeshow, (1972) and Freeman, et al., (1976) discussed the multivariate applications of BRR with complex sampling. Casady, R.J. (1975) extended the BRR technique for estimating the components of variance from a multistage design. Nathan, G. (1973, 1975) discussed tests of independence in contingency tables. Gurney and Jewett (1975) demonstrated a general method for generating orthogonal replications and illustrated for a design with three primary selection per stratum.

The method could be generalised to multistage sampling, where PSU's are selected with probability proportional to size with replacement within strata, see for instance, Rao (1975).

3.2.2.7B. Partially Balanced Half Samples

We have seen earlier in Subsection 3.2.2.7A that the minimum number K of balanced half samples required to obtain cross-product balance is $L + 3$. Thus, when the number of strata is large, full orthogonal balancing may be too expensive to process the results. In such situations, in order to reduce the number of half samples, one could generate a smaller number of partially balanced half-samples, see, for instance, McCarthy (1966, 1969a, b).

Consider a stratified simple random sample consisting of L strata with two selections from each stratum and suppose that a balanced set of $K \geq L/m$ balanced half samples is selected, where L/m is assumed to be an integer for convenience sake. Now $\delta_h^{(i)}\delta_j^{(i)} = 0$ when h and j belong to the same group or belong to different groups but do not correspond to the same column in the two balanced sets. The variance estimator is given by

$$\hat{V}_{PB,K}(\bar{y}_{st}) = \frac{1}{K}\sum_{i=1}^{K}(\bar{y}_{hs,i}-\bar{y}_{st})^2 = \frac{1}{4}\sum_{h=1}^{L} W_h^2 d_h^2 + \frac{1}{2}\sum_{h<j}{}' W_h W_j d_h d_j \ , \qquad (3.58)$$

where Σ' denotes the summation over all pairs (h, j) such that

(i) $h < j$,

(ii) h is from one set of K strata and j is from another set of K strata,

(iii) h and j represent corresponding columns from the L/m orthogonal columns that make up a balanced set.

The number of terms in the summation Σ' are

$$K(L/m)C_2 = \frac{L(m-1)}{2} \ , \qquad (3.59)$$

which increases as the number of component samples increases. The efficiency of a partially balanced variance estimator will depend on the number of component samples (m) as well as the arrangement of strata into groups, for a given m .

We note that there are $\frac{L(m-1)}{2}$ different non-vanishing cross-product terms which contribute substantially to the total variance of $\hat{V}_{PB,K}(\bar{y}_{st})$. Their total contribution, however, depends upon the pattern by which the parent sample is split into m component samples. Lee (1972-73) examined the possibility of several strata arrangements into groups so that the sum of the cross-product terms could be reduced by suitably splitting the parent sample into m component samples. Through the theoretical and empirical results it is demonstrated that in order to increase the precision of a partially balanced design, the parent sample should be divided into component samples such that each component sample has a greater degree of homogeneity in the associated quantities $W_h^2 S_h^2$. The procedures considered by Lee include: AOA (Ascending order arrangement), HTA (Heterogeneity arrangement) and SAOA (Semi-ascending order arrangement). The SAOA procedure consists in arranging the strata in the ascending order of the magnitudes of $W_h^2 S_h^2$ first, then rearranging the last $\frac{L}{2}$ (or $\frac{L-1}{2}$ if L is odd) strata in the reverse order and finally taking the first L/m strata for the first group, second L/m strata for the second group and so on. The SAOA procedure gives a better pattern of arrangement and also provides a more precise estimator of $\hat{V}_{PB,K}(\bar{y}_{st})$ as compared to AOA or HTA procedures. Lee also considered modifications of the repeated replication and balanced repeated replication methods. These modifications are summarised in the next subsection.

3.2.2.8 Repeated Partial Sample (RPS), Balanced Repeated Partial Sample (BRPS) and the Generalised Repeated Partial Sample(GRPS) Schemes.

In practice, most of the complex surveys do not generally conform to the condition of two primary selections per stratum. In order to resolve such situations, a more general version of the pseudo-replication techniques was proposed by Lee (1973), where the restriction is that at least two selections are drawn from each stratum. Lee's method is also based on the concept of partial samples where a partial sample is formed by removing one selection per stratum from the parent sample. The variance of the statistic is estimated from the variabilities among the pseudo-values of the statistic. The pseudovalues, one from each partial sample, are computed by the same estimation procedure as is used to compute the statistic.

We consider briefly the RPS, BRPS and GRPS schemes hereunder.

(a) RPS Scheme: This scheme consists in forming a number of partial samples from the parent sample by deleting one sample unit from each stratum. So far as the construction is concerned it is analogous to the ordinary jackknife method. Each partial sample contains $\left(\sum_{h=1}^{L} n_h - L\right)$ sample units. Let y_{hc} be the value of y characteristic for the sampled unit in stratum h, which is excluded from the partial sample c. The cth pseudo value is then given by

$$y_{(c)} = \sum_{h=1}^{L} W_h \sum_{j \neq c}^{n_h} y_{hj} \,/\, (n_h - 1) = \bar{y}_{st} + \sum_{h=1}^{L} W_h (\bar{y}_h - y_{hc})/n_h - 1 \,. \qquad (3.60)$$

Since, there are $\prod_{h=1}^{L} n_h$ different partial samples with $\prod_{h=1}^{L} n_h$ associated different pseudo values of $y_{(c)}$, therefore for a complete set of $\prod_{h=1}^{L} n_h$ different partial samples the average of the pseudo values is

$$\bar{y}_c = \frac{1}{\prod_{h=1}^{L} n_h} \sum_{c=1}^{\prod_{h=1}^{L} n_h} y_{(c)} ,$$

$$= \bar{y}_{st} + \frac{1}{\prod_{h=1}^{L} n_h} \sum_{h=1}^{L} \frac{W_h}{n_h - 1} \sum_{c=1}^{\prod_{p=1}^{L} n_p} (\bar{y}_h - y_{hc}) . \qquad (3.61)$$

Thus, each of the sampled units in stratum h is drawn for omission from an equal number, $\prod_{p=1}^{L} n_p / n_h$, of the total number $\prod_{p=1}^{L} n_p$ partial samples. Taking the sum of squared deviations $(y_{(c)} - \bar{y}_{st})$ and averaging over all, the different partial samples, the variance estimate is given by

$$\hat{V}_{RPS}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{s_h^2}{n_h(n_h - 1)} . \qquad (3.62)$$

For $n_h = 2$, $\hat{V}_{RPS}$ provides an unbiased estimate of the variance of $\bar{y}_{st}$.

(b) Balanced Repeated Partial Samples: Since the computation of $\hat{V}_{RPS}(\bar{y}_{st})$ involves a large number $\prod_{p=1}^{L} n_p$ partial samples which will be laborious and expensive if the parent sample is large. To overcome this difficulty a smaller number K of partial samples are selected so that they are balanced. Then

$$\hat{V}_{RPSB,K}(\bar{y}_{st}) = \sum_{c=1}^{K} (\bar{y}_{ct} - \bar{y}_{st})^2 / K , \qquad (3.63)$$

where

$$\bar{y}_{ct} = \bar{y}_{st} + \sum_{h=1}^{L} W_h (\bar{y}_h - y_{hc}) / \sqrt{n_h - 1}$$

The balanced set of K partial samples can be obtained either by using Placket-Burman matrices or the balanced designs devised by

factorial experiments available in literature.

(c) Generalised Repeated Partial Sample Scheme (GRPS): In the simplest form the GRPS scheme consists in splitting the L strata into L groups, each group consisting of one stratum only. Then, a partial sample of (2L-1) primary selections is formed by selecting one primary selection from the rth group (stratum) and adding it to all the primary selections of the remaining (L-1) groups. Thus a total of 2L different partial samples of (2L-1) primary selections each can be formed from the parent sample. All the 2L different partial samples are then used for estimating the variance. If the number of strata is large, the L strata can be grouped into $g \leq L$ groups such that the ℓth group consists of r_ℓ strata where the g numbers $r_\ell (\ell = 1,2,\ldots,g)$ need not be equal and $L = \sum_{\ell=1}^{g} r_\ell$. One primary selection is then drawn from each of the r_ℓ strata of the ℓth group to form a combination of r_ℓ primary selections. Thus 2^{r_ℓ} such different combinations can be created from the ℓth group. By adding one such combination to all the primary selections of the remaining (g-1) groups a partial sample of $(2L-r_\ell)$ primary selections may be formed. Thus, a total of 2^{r_ℓ} different partial samples can be formed in association with the ℓth group and by repeating the procedure for each of the g groups a total $t = \sum_{\ell=1}^{g} 2^{r_\ell}$ different partial samples can be formed from the parent sample.

Let $y_{(\ell p)}$ be the pseudovalues obtained from the pth partial sample associated with the ℓth group, $\ell = 1,2,\ldots,g$ and estimated by,

$$y_{(\ell p)} = \sum^{(\ell)} W_h \bar{y}_h + \sum^{\ell} W_h y_{hp} ,$$

$$= \bar{y}_{st} + \sum^{\ell} W_k (y_{kp} - \bar{y}_k) . \qquad (3.64)$$

where $\sum^{(\ell)}$ denotes the summation over r_ℓ strata of the ℓth group, $\sum^{(\ell)}$ denotes the summation over the remaining $(L-r_\ell)$ groups and y_{kp} is the y value of the sampled unit (from the kth stratum of the ℓth group) which is included in the Pth partial sample. In general, the average of $y_{(\ell p)}$ over all the 2^{r_ℓ} different partial samples associated with the ℓth group reduces to $\bar{y}_{st}$. This implies that the averaging of $y_{(\ell p)}$ over all the t different partial samples is also equal to $\bar{y}_{st}$. The average of the squared deviations over all the 2^{r_ℓ} different partial samples associated with the ℓth group reduces to

$$\sum^{(\ell)} W_k^2 (y_{kp}-\bar{y}_k)^2/4 \ . \tag{3.65}$$

The average of the squared deviations $(y_{(\ell p)} - \bar{y}_{st})^2$ over the t different partial samples is

$$\sum_{\ell=1}^{g} \sum^{(\ell)} W_k (y_{k1}-y_{k2})^2/4 = \sum_{h=1}^{L} W_h^2 (y_{h1}-y_{h2})^2/4 = \hat{V}(\bar{y}_{st}) \ . \tag{3.66}$$

Thus,

$$\hat{V}_{GRPS}(\bar{y}_{st}) = \sum_{\ell=1}^{g} \sum_{p=1}^{K_\ell} (y_{(\ell p)}-\bar{y}_{st})^2/K_\ell \tag{3.67}$$

where, $k_\ell \leq 2^{r_\ell}$ is the number of different partial samples associated with the ℓth group. The expression (3.66) gives an unbiased estimator of the variance of $\bar{y}_{st}$. When $g = 1$, it reduces to BRR variance estimator. Further details on these schemes could be found in Lee, (1972, 1973).

The pseudo-replication techniques have the limiation that these are applicable only to the situations where two selections are drawn from each stratum, except the general methods proposed by Lee. However, Lee's

methods have not found wide applicability. The other difficulty with BRR techniques relates to the approximate degrees of freedom for the estimated variance. Thus, it appears worthwhile to consider other methods with wide applicability and more attractive features such as the jackknife, the bootstrap and the weighted jackknife.

3.2.2.9 The Jackknife Method:

We have seen that the BRR technique has the disadvantage that the appropriate matrices must be derived for each problem of interest and the half samples must be constructed in accordance with the matrix specifications. The Taylor Series method, on the other hand, requires the evaluation of thepartial derivatives which may be difficult in certain cases. A technique which provides a simple variance estimator and does not require the evaluation of partial derivatives, as in the case Linearization variance estimator, or the construction of Placket and Burman matrices, as with the BRR variance estimator, is the jackknife method.

The jackknife, introduced by Quenouille (1949, 1956) as a method of bias reduction, has been shown to have robust qualities in a wide variety of situations. Tukey (1958) extended Quenouille's procedure to an inference technique that could be appropriate either when distributional assumptions are in question or alternatively when a complete distributional theory is impossible to derive analytically. The name "jackknife" suggested by Tukey refers to a generally applicable tool like the boy scout's jackknife that has great utility in many situations even though most of its uses could be carried out better by applying specialised tools if such tools are available.

The simple idea of the jackknife technique can be described as follows: Consider $\hat{\theta}$ as an estimate of a parameter θ based on a sample of size n. Now delete one of the observations, say the jth, from the sample and recompute the estimate. Call it $\hat{\theta}_{-j}$. Continue the process until each observation has been deleted exactly once. Define the pseudovalues

$$\tilde{\theta}_j = n\hat{\theta} - (n-1)\hat{\theta}_{-j} \quad , \quad j = 1,2,\ldots,n \qquad (3.68)$$

Then the jackknife estimate of θ is the average of these pseudovalues,

$$\tilde{\theta} = \frac{1}{n} \sum_{j=1}^{n} \tilde{\theta}_j \quad , \qquad (3.69)$$

and jackknife variance estimate is

$$\hat{V}(\tilde{\theta}) = \frac{1}{n(n-1)} \sum_{j=1}^{n} (\tilde{\theta}_j - \tilde{\theta})^2 \quad . \qquad (3.70)$$

Properties of the jackknife in the case of simple random sampling have been extensively studied. Miller (1974b) gives an excellent review of the method and its wide range of applicability. Kish and Frankel (1970) have proposed four variants of the jackknife method for estimating the variance from a stratified sample with two PSU's per stratum.

Besides, being not always advantageous, the indiscriminate use of the jackknife may have disasterous consequences in certain situations. For example, the jackknife does not perform well in the general area of order statistics such as estimation of median and range. Another area in which the jackknife has little or no success is in time series analysis. It may, however, be pointed out that the jackknife is a general procedure not necessarily intended to be the best procedure in

specific situations. Since our main interest lies in the jackknife method and its applications, further details regarding the method and related issues will appear in Chapter 5.

3.2.2.10 The Bootstrap Method:

Efron (1979a) proposed an alternative procedure for variance estimation in complicated situations. This procedure is based on repeated samples generated randomly according to the empirical distribution function of the sample. He has coined the term "bootstrap" for this procedure. The following introduction to bootstrap method is given by Efron (1981a).

"The bootstrap is a simple and straightforward method for calculating approximated biases, standard deviations, confidence intervals, and so forth, in almost any nonparameteric estimation problem. Method is keyword here, since little is known about the bootstrap's theoretical basis, except that,(a) it is closely related to the jackknife; (b) under reasonable conditions it gives asymptotically correct results; and (c) for some simple problems which can be analysed completely, for example, ordinary linear regression, the bootstrap automatically produces standard solutions". It may, however, be noted that it is not still known how one could bootstrap in complex surveys.

The most interesting feature of the bootstrap is that it provides a connection between statistical theory and computational power. Regarding this aspect Efron (1979b) says, "The purpose of mathematical theory, and in fact all scientific theory, is to reduce complicated situations to simple ones. Just what a scientist means by "simple" is determined by experience, training, convention and the limitations of human reasoning faculties. A Taylor Series expansion is a classic example of this process, a given function is expressed as a sum of multiples of powers. Since we are taught a lot about sums, multiples

and powers, the explanation may be a good deal easier to understand than the function as originally stated.

The advent of high speed computers has redefined "simple" in the mathematical sciences. For example, an optimization problem which can be reduced to a problem in linear programming, is in most instances, now considered solved. Since the simplex method is so efficient in numerically solving linear programmes".

The purpose of Efron's article is to show this same process at work in mathematical statistics . "A theory which enables a scientist to understand his data with the help of a high speed computer may now be as useful as a theory which only requires a table of exponential function, particularly if the latter theory does not exist. Computer assisted theory is no less "mathematical" than the theory of the past, it is just less constrained by the limitations of human brain".

The basic principle of the bootstrap is that the variance of T, where T is a functional defined on the distribution of F, can be estimated by the variance of T on F^*, the empirical cumulative density function for the sample. Following Efron (1979a, 1979b) the underlying idea of bootstrap can be described as follows:

Consider a random sample of size n observed from a completely unspecified probability distribution F

$$X_i = x_i,\ X_i \overset{\text{ind}}{\sim} F\ ,\quad i = 1,2,\ldots,n\ . \tag{3.71}$$

where F denotes the distribution on either the real line or the plane but does not play any role in the theory. Let $\underline{X} = (X_1,X_2,\ldots,X_n)$ and $\underline{x} = (x_1,x_2,\ldots,x_n)$ denote the random sample and its observed realisation respectively. The problem we wish to solve is as follows: Given a specified random variable $R(\underline{X},F)$ possibly depending on both $\underline{X}$ and

the unknown distribution F, estimate the sampling distribution of R on the basis of observed data $\underline{x}$. Now the bootstrap method can be described as follows:

(1) Construct the sample probability distribution $\hat{F}$, putting mass $1/n$ at each point $x_1, x_2, \ldots, x_n$.

(2) With $\hat{F}$ fixed, draw a random sample of size n from $\hat{F}$, say

$$X_i^* = x_i^* , \quad X_i^* \underset{\text{ind}}{\sim} \hat{F} , \quad i = 1,2,\ldots,n \tag{3.72}$$

Call this the bootstrap sample,

$$\underline{X}^* = (X_1^*, X_2^* \ldots, X_n^*) , \quad \underline{x}^* = (x_1^*, x_2^*, \ldots, x^*_n)$$

The values of $\underline{X}^*$ are selected with replacement from the set $\{x_1, \ldots, x_n\}$.

(3) Approximate the sampling distribution of $R(\underline{X}, F)$ by the bootstrap distribution of

$$R^* = R(\underline{X}^*, \hat{F}) \tag{3.73}$$

i.e., the distribution of R^* induced by the random mechanism (3.72) with $\hat{F}$ held fixed at its observed value. The point is that the distribution of R^*, which in theory can be calculated exactly once the data is observed, equals the desired distribution of R if $F = \hat{F}$.

3.2.2.10A. Calculation of Bootstrap Distribution:

The difficult part of the bootstrap procedure is the actual calculation of the bootstrap distribution. The possible methods of calculation are:

(i) Direct theoretical calculation as illustrated in Examples 3.9 and 3.10 below:

Example 3.9:

Consider a probability distribution F putting all its mass at 0 or one, and let the parameter of interest be $\theta(F) = \text{Prob}_F(X=1)$ The most obvious random variable of interest is

$$R(\underline{X}, F) = \bar{X} - \theta(F) \quad , \quad \text{where} \quad \bar{x} = \sum_{i=1}^{n} x_i/n \tag{3.74}$$

Having observed $\underline{X} = \underline{x}$, the bootstrap sample $\underline{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)$ has each component independently equal to one with probability $\bar{x} = \theta(\hat{F})$ and zero with probability $1 - \bar{x}$. Now from the standard binomial result

$$R^* = R(X^*, \hat{F}) = \bar{X}^* - \bar{x} \tag{3.75}$$

has mean and variance $E_*(\bar{X}^* - \bar{x}) = 0$, $\text{Var}_*(\bar{X}^* - \bar{x}) = \bar{x}(1-\bar{x})/n$, where E_* and Var_* indicate probability calculations with respect to the bootstrap distribution of $\underline{X}^*$, with $\underline{x}$ and $\hat{F}$ fixed.

Example 3.10:

Consider the estimation of $\theta(F) = \text{Var}_F(X)$, the variance of an arbitrary distribution on the real line, using the estimator

$$t(\underline{X}) = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1) \tag{3.76}$$

We wish to know the sampling distribution of

$$R(\underline{X}, F) = t(\underline{X}) - \theta(F) \tag{3.77}$$

Let $\mu_k(F)$ indicate the kth central moment of F, then

$$\mu_k(F) = E_F(X - E_F X)^k \tag{3.78}$$

and

$$\hat{\mu}_k = \mu_k(\hat{F}) \quad ,$$

the kth central moment of $\hat{F}$. Standard sampling theory results (Cramer (1946), Section 27.4) show that

$$R^* = R(\underline{X}^*, \hat{F}) = t(\underline{X}^*) - \theta(\hat{F})$$

has

$$E_* R^* = 0 \quad \text{and} \quad \text{Var}_* R^* = \frac{\hat{\mu}_4 - [(n-3)/(n-1)]\hat{\mu}_2^2}{n} \tag{3.79}$$

(ii) Monte Carlo Approximation to the Bootstrap Distribution:

According to this method repeated replications of $\underline{X}^*$ are generated by taking random samples of size n from $\hat{F}$, say $x^{*1}, x^{*2}, \ldots, x^{*N}$ and the histogram of the corresponding values $R(x^{*1}, \hat{F})$, $R(x^{*2}, \hat{F}), \ldots,$ $R(x^{*N}, \hat{F})$ is taken as an approximation to the actual bootstrap distribution. This approach is illustrated by example 3.11 for the estimation of correlation coefficient from Efron (1979b).

Example 3.11: (Standard Error of Correlation Coefficient)

Table 3.1 gi the average LSAT and GPA scores for entering classes of 15 American Law Schools in 1973. For each school two numbers are given:

x_i = average LSAT score of entering students at School i ,

y_i = average undergraduate GPA score of entering students at School i .

Table 3.1 The Average LSAT and GPA for 15 American Law Schools, 1973

School	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LSAT	576	635	558	578	666	580	555	661	651	605	653	575	545	572	594
GPA	3.39	3.30	2.81	3.03	3.44	3.07	3.00	3.43	3.36	3.13	3.12	2.74	2.76	2.88	2.96

Note: LSAT is a national test for prospective Lawyers and GPA the undergraduate grade point average.

The observed Pearson correlation coefficient between n pairs of numbers (x_i,y_i), $i=1,2,\ldots,15$ is

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}, \text{ where } \bar{x} = \sum_{i=1}^{n} x_i/n, \bar{y} = \sum_{i=1}^{n} y_i/n .$$

$$= 0.776$$

Scatter plot for these 15 pairs appears as Fig. 3.1. We wish to estimate the standard error for $\hat{\rho}$. The bootstrap procedure for this is as follows:

(A) Suppose that the data points $x_1,x_2,\ldots,x_{15}$ are independent observations from some bivariate distribution F on the plane. Then the true standard error of $\hat{\rho}$ is a function of F, denoted by $\sigma(F)$,

$$\sigma(F) = \left[\mathrm{Var}_F \hat{\rho}(X_1,X_2,\ldots,X_n)\right]^{\frac{1}{2}} \tag{3.80}$$

It is also a function of the sample size n, and the functional form of the statistic $\hat{\rho}$, but both of these are known to the statistician.

(B) We do not know F, but we can estimate it by the empirical probability distribution $\hat{F}$,

$\hat{F}$: mass $\frac{1}{n}$ on each observed data point x_i, $i=1,2,\ldots,n$.

(C) The bootstrap estimate of $\sigma(\hat{F})$ is

$$\hat{\sigma}_B = \sigma(\hat{F}) \tag{3.81}$$

It may, however, be pointed out that for the correlation coefficient and most statistics, even very simple ones, the function $\sigma(F)$ is impossible to express in closed form. This is the reason why bootstrap is not in common use. However, $\hat{\sigma}_B$ can easily be approximated by Monte Carlo methods. The main steps for example 3.11 are:-

Fig. 3.1

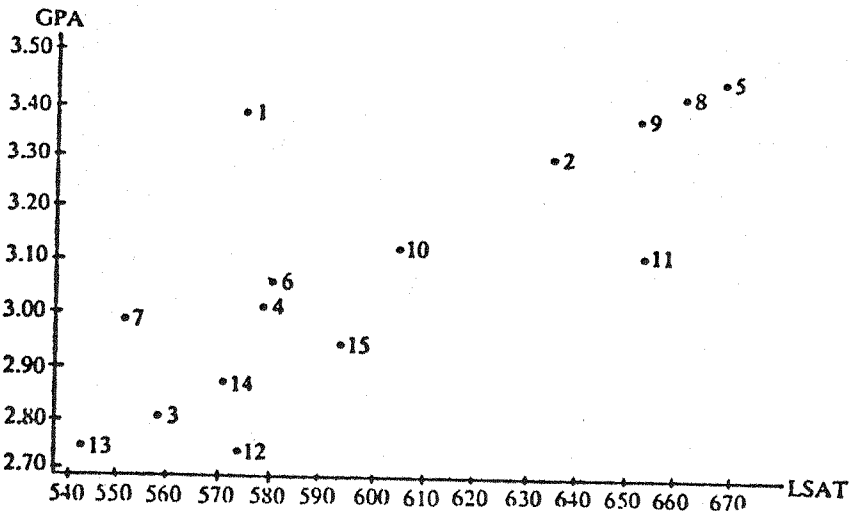

Fig. 3.1 A plot of the law school data given in Table 3.1.

Fig. 3.2

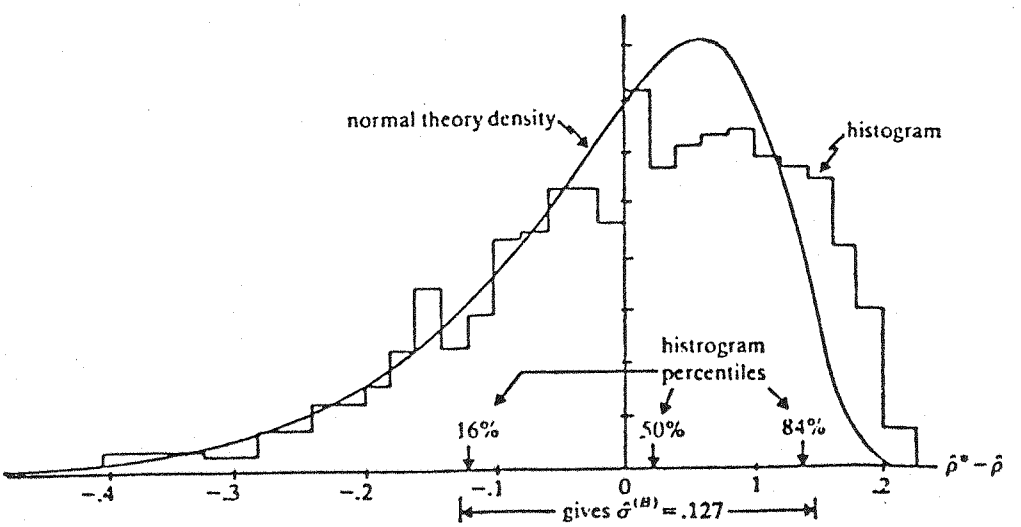

Fig. 3.2 Histogram 1000 bootstrap replications of $\hat{\rho}^*-\hat{\rho}$ gives a bootstrap estimate of accuracy $\hat{\sigma}^{(B)}=.127$ for the correlation coefficient $\hat{\rho}$ = .776 of the law school data. The normal theory density of $\hat{\rho}^*-\hat{\rho}$ has a similar shape, but falls off more quickly at higher values of $\hat{\rho}^*-\hat{\rho}$.

(a) Let $\hat{F}$ be the empirical distribution of the 15 observed data points i.e. the probability distribution which puts mass 1/15 at each observed point (x_i, y_i) as described above.

(b) Use a random number generator to draw 15 new points (x_i^*, y_i^*) independently with replacement from $\hat{F}$, so that each new point is an independent random selection of one of the 15 original data points. We call these new points as "bootstrap sample". They are the subset of original points and it is possible that some of the original points might have been selected zero times, once, twice, etc.

(c) Compute $\hat{\rho}^* = \hat{\rho}(x_i^*, y_i^*)$, the correlation coefficient for the bootstrap sample.

(d) Repeat steps (b) and (c) a large number of times, say K, each time using an independent set of new random numbers to generate the new bootstrap sample. Call the resulting sequence of bootstrap correlation coefficients $\hat{\rho}^{*1}, \hat{\rho}^{*2}, \ldots, \hat{\rho}^{*K}$

(e) Approximate $\hat{\sigma}_B$ by

$$\hat{\sigma}_B = \left(\sum_{j=1}^{K} (\hat{\rho}^{*j} - \tilde{\rho})^2 / (K-1) \right)^{\frac{1}{2}} \tag{3.82}$$

where

$$\tilde{\rho} = \sum_{j=1}^{K} \hat{\rho}^{*j} / K$$

As $K \to \infty$, (3.82) approaches the original definition (3.81). For this example $K = 1000$ and Fig. 3.2 shows bootstrap replications $\hat{\rho}^{*1}, \hat{\rho}^{*2}, \ldots, \hat{\rho}^{*1000}$. The abscissa is plotted in terms of $(\hat{\rho}^{*j} - \hat{\rho})$.

From (3.82) we get the estimate of standard error of $\hat{\rho}$.

$$\hat{\sigma}_B = 0.127$$

The normal theory estimate of standard error of $\hat{\rho}$ (Johnson and Kotz, 1970, p.229) for this case is

$$\hat{\sigma}_N = \frac{1-\hat{\rho}^2}{\sqrt{n-3}} = 0.115$$

It is apparent that $\hat{\sigma}_B$ and $\hat{\sigma}_N$ compare favourably well.

(iii) Taylor Series Approximation Method.

The Taylor Series approximation method can also be used to obtain the mean and variance of the bootstrap distribution of R*. This, however, turns out to be the same as the usual jackknife theory (see Efron, 1979a, 1979b, 1981a, 1982).

The important point about the bootstrap is that it can be applied just as well to any statistic, simple or complicated, as to the correlation coefficient. The great power of the bootstrap like jackknife is that it can be used in extremely complicated situations. The bootstrap methods, however, have not found wide applicability and recognition like the jackknife methods. Efron (1979a, 1979b, 1981a, 1981b, 1981c, 1982), Efron and Stein (1981) and Efron and Gong (1983) have discussed several other applications such as ratio estimation, regression models, cross validation, censored data and the error rates of prediction rule. Rocke and Down (1981) discuss the application of the bootstrap for estimating the variances of location estimators. Parr (1983) establishes the connection between the jackknife, the bootstrap and the delta method. Through simulation results and asymptotic expansion,

he shows that the three methods coincide at least to first order terms. The absence of skewness leads to the second order coincidence of the jackknife and the delta methods.

It may, however, be pointed out that the bootstrap is a generalisation of the jackknife. It is also true that the bootstrap like jackknife does not always win the ground, see, for instance, Rocke and Down (1981). In certain cases, such as order statistics, it performs better than the ordinary jackknife but also requires more computational time. In fact, the two methods are very closely related. The ordinary jackknife is the method of choice if one does not want to do bootstrap computations due to cost factor. The bootstrap method has limited applications, particularly in connection with complex survey situations it is not known how bootstrap could be defined. In view of this situation, we will not discuss the bootstrap method any further in this study rather we will concentrate on the jackknife method which is simple and has wider utility.

3.2.2.11 The Weighted Jackknife:

The standard jackknife procedure, outlined earlier in subsection 3.2.2.9, defines the pseudovalues symmetrically with respect to the observations. It may, however, be pointed out that certain observations are influential and may cause a wide variation in pseudovalues. A case may be judged influential if the important features of the analysis are changed substantially when it is deleted from the data, in particular, if the pseudo-sample contains an outlier. Thus an estimate that depends heavily on such data points or observations is prone to instability. The standard jackknife procedure is sensitive to aberrant values and may lead to erroneous results in the presence of such observations or data points. This point has been noted,

among others, by Efron (1979b), Efron and Gong (1983), Hinkley (1977a, 1978), Hobbs, et al., (1980) and Miller (1974a). It was in an attempt to find a suitable version of the jackknife that could be used with non-symmetric data that Hinkley (1977b) proposed a weighted jacknife procedure for robust regression analysis. In the weighted jackknife the pseudovalues are defined in terms of the distances "w_i" of the observations from the centre of the data whereas in the standard jackknife "w_i" is constant.

The concept of the weighted jackknife is based on the empirical influence function. In order to grasp the development of the weighted jackknife procedure it is essential to understand the connection between the jackknife and the influence function (Huber, 1972). The story starts from 1947 when von Mises introduced the idea of influence function in his study of differentiable statistical functions, but it remained relatively unnoticed for about two decades until investigators (Hampel, 1974) interested in robust estimation discovered its usefulness. The idea of influence functions is as follows: Let the unknown parameter θ be considered to be a function $\theta = T(F)$ of the underlying distribution F, and its estimator $\hat{\theta}$ to be the same function of the sample distribution F_n. Then the influence function $I(y, \theta)$ is defined as

$$I(y,\theta) = \lim_{\varepsilon \to 0} \frac{T\left[(1-\varepsilon)F+\varepsilon\delta_y\right]-T(F)}{\varepsilon} \qquad (3.83)$$

where δy is the distribution function which places mass one at y. The influence function $I(y,\theta)$ measures the amount of change in $T(F)$ for an infinitesimal change in the weight assigned by F to y. It is like a partial derivative of T with respect to a change in F at ordinate y.

The connection between the jackknife and the influence function is that the pseudo-values give finite difference sample estimates of the influence function. For

$$\varepsilon = -\frac{1}{n-1}, \quad 1-\varepsilon = \frac{n}{n-1} \quad \text{and} \quad F = F_n, \tag{3.84}$$

the quantity $(1-\varepsilon)F + \varepsilon\delta_y$ at $y = Y_i$ becomes

$$(1-\varepsilon)F_n + \varepsilon\delta_{Y_i} = \frac{n}{n-1}F_n - \frac{1}{n-1}\delta_{Y_i} = F_{n-1,-i},$$

where $F_{n-1,-i}$ is the sample distribution function based on $(n-1)$ observations with the ith observation deleted. Now if the finite difference sample estimate of $I(y,\theta)$ at Y_i for $\theta = \hat{\theta} = T(F_n)$ is defined by,

$$\hat{I}(Y_i,\hat{\theta}) = \left.\frac{T\left[(1-\varepsilon)F+\varepsilon\delta_{Y_i}\right]-T(F)}{\varepsilon}\right|_{F=F_n,\ \varepsilon=-\frac{1}{(n-1)}} \tag{3.85}$$

then it follows that

$$\begin{aligned}\tilde{\theta}_i &= n\hat{\theta} - (n-1)\hat{\theta}_{-i} \\ &= \hat{\theta} + (n-1)(\hat{\theta}-\hat{\theta}_{-i}) \\ &= \hat{\theta} + \hat{I}(Y_i,\hat{\theta})\end{aligned} \tag{3.86}$$

because

$$\hat{\theta}_{-i} = T(1-\varepsilon)F_n + \varepsilon\delta_{Y_i}).$$

If the influence functions is sufficiently smooth so that $\hat{I}(y,\hat{\theta})$ converges to $I(y,\theta)$ for all y as $n \to \infty$, then each pseudovalue $\tilde{\theta}_i$ is approximately $\theta + I(Y_i,\theta)$. This implies that the jackknife

will be behaving correctly asymptotically because $\tilde{\theta}$ will be asymptotically normally distributed with mean θ and variance $\int I^2(y,\theta)dF(y)/n$, which is correct distribution of $\hat{\theta}$ for any underlying distribution function F.

Huber (1972, p.1052) notes that "Hampel's Influence Function is the most important heuristic tool for constructing robust estimates with specific properties. One will strive for influence functions which are bounded (to limit the influence of any single "bad" observation), which are reasonably continuous in y (to achieve insensitivity against round-off and grouping effects), and which are reasonably continuous as functions of F (to stabilize asymptotic variance of the estimate under small changes of F)".

Exploiting the connection between the jackknife and the influence function as explained in the preceding paragraph, Hinkely proposed the weighted pseudo-values and the weighted jackknife variance estimator for the general linear model $\underline{Y} = \underline{X}\beta+\underline{e}$ with rank $(\underline{X}) = P$ and $\text{Var}(\underline{e}) = \sigma^2\underline{I}$. The formulation is briefly outlined as follows:

Let $\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$ be the least squares estimator of $\underline{\beta}$ based on full data and $\hat{\underline{\beta}}_{-i}$ be the alternative least squares estimator obtained by deleting the ith row, that is, $\hat{\underline{\beta}}_{-i} = (\underline{X}_{-i}^T\underline{X}_{-i})^{-1}\underline{X}_{-i}^T\underline{Y}_{-i}$. where the subscript -i attached to a quantity means that the ith case has been deleted. Thus $\underline{X}_{-i}$ means that the ith row $\underline{x}_i^T$ has been deleted from $\underline{X}$ matrix. Similarly $\underline{Y}_{-i}$ indicates that the ith value is deleted from $\underline{Y}$ vector.

Now the contribution of the ith observation to the jackknifed estimator is

$$\hat{\underline{\beta}} - \hat{\underline{\beta}}_{-i} = \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i}{1-w_i} \quad . \tag{3.87}$$

where $w_i = \underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i$ provides a measure of how far the ith case is from the centre of the data and $R_i = y_i - \underline{x}_i^T\underline{\beta}$ is the ith residual. The empirical influence function is

$$\hat{I}(\underline{\beta}; \underline{x}^T_i, Y_i) = n(1-w_i)(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i}) \quad . \tag{3.88}$$

The weighted jackknifed pseudovalues are now defined as

$$\tilde{\beta}_{W_i} = \hat{\underline{\beta}} + \hat{I}(\underline{\beta}; \underline{x}_i^T, Y_i) \quad , \tag{3.89}$$

and the weighted jackknifed variance estimate is given by

$$\hat{V}_{WJ} = \frac{1}{n(n-p)} \sum_{i=1}^{n} (\tilde{\beta}_{Wi}-\tilde{\beta}_W)(\tilde{\beta}_{Wi}-\tilde{\beta}_W)^T \quad , \tag{3.90}$$

where

$$\tilde{\beta}_W = \frac{1}{n} \sum_{i=1}^{n} \tilde{\beta}_{Wi} \quad .$$

is the weighted jackknifed estimator.

The weighted jackknife possesses nice properties and provides a variance estimator which is robust against the non-homogeneity of the error variances. The method appears to be promising as compared with the standard jackknife and other methods in the sense that the residuals are weighted by their real effect on the estimation. Thus, it is hoped that the estimates based on the weighted pseudovalues will be less sensitive to extreme data points and will give reasonably stable variance estimates. However, still much remains to be investigated regarding the applications of this method to other non-linear estimators such as ratio estimator and correlation coefficient as well as in the

area of order statistics. We discuss the weighted jackknife method, its properties and its further applications in greater detail in Chapters 8 - 9.

3.3 Model Based Approach to Variance Estimation:

In the previous section we have discussed the randomisation based approach to variance estimation from complex surveys. This approach, as we have seen, offers several choices but it has been recognised that the randomisation based variances may not provide a suitable framework for inference from an observed sample. In comparison to this, under the prediction approach to finite population sampling, the variance of an estimated value is determined conditionally from the set s of sample units actually observed. By conditioning on the sample actually chosen, the prediction approach often reveals relationships that are important for inference but that are concealed in the randomisation based theory. The prediction approach has proved to be a useful and meaningful tool for analysing finite population sampling problems.

In this section, we examine variance estimation under the prediction approach. In particular, we are concerned with a non linear estimator, the ratio estimator, for which the prediction approach has suggested alternatives to the randomisation based variance estimator. The prediction approach has indicated serious drawbacks in the variance estimator (3.30) for the ratio estimate. For example, Royall (1971) suggested that this estimator can be badly biased with respect to a superpopulation model unless the sample is balanced on size i.e. the average size of the sample units is approximately equal to the population average. Royall and Eberhardt (1975), and Royall and Cumberland (1978) have suggested alternative robust variance estimators. We review these and other variance estimators under the model ξ. The preliminaries

have been discussed in Chapter 2 and we will refer to these when the need arises.

3:3.1 Variance Estimates for the Ratio Estimator:

Consider a population of N units labelled $1,2,\ldots,N$. Associated with the unit i is a pair (x_i,y_i) with x_i known and y_i fixed but unknown. The variable y_i represents the value of some characteristics of interest while the auxiliary variable x_i is frequently a measure of size of the ith unit. Some practical examples are:

(i) y_i may be the volume of wheat produced by the ith farm and x_i the acreage under wheat, and

(ii) y_i may be the number of patients in a hospital during a particular period and x_i the number of beds at that hospital.

Let $y_1,y_2,\ldots,y_N$ denote the realised values of the independent random variables $Y_1,Y_2,\ldots,Y_N$. Then we assume that the relationship between x_i and y_i can be expressed by a regression model,

$$Y_i = \beta x_i + e_i \ , \tag{3.91}$$

where, $E(e_i/x_i) = 0$, $E(e_i^2/x_i) = \sigma x_i^2$ and $E(e_i e_j/x_i,x_j) = 0$; for $i \neq j$. This model provides a simple mathematical description of a population in which the pair (x_i,y_i) are concentrated about a straight line through the origin with scatter about the line increasing as x increases. This model reflects the structure of many populations.

The ratio estimator for the population total $T = \sum_{i=1}^{N} y_i$, is defined as $\hat{T}_R = \frac{\sum_s y_i}{\sum_s x_i} \sum_{i=1}^{N} x_i$. Under the randomisation approach the

ratio estimators $\hat{T}_R$ is biased. However, the bias becomes small as the sample size increases and is negligible when n is large. The randomisation based analysis requires that the samples be selected via simple random sampling. It is this artificial randomisation, employed in the selection of the sample, which generates the probability distribution from which the properties of the ratio estimator $\hat{T}_R$ are derived. Under the randomisation based approach the precision of $\hat{T}_R$ is measured by

$$V_c = \frac{N^2}{n}(1 - \frac{n}{N}) \frac{1}{N-1} \sum_{i=1}^{N} (y_i - Rx_i)^2 ,$$

where

$$R = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} \tag{3.92}$$

which does not depend on the selected sample s.

The prediction approach suggests markedly different results for $\hat{T}_R$ than those discussed above under the randomisation based approach. Under the prediction appraoch, after the sample is observed, the population total, T, can be written as the sum of two parts,

$$T = \sum_{s} y_i + \sum_{\tilde{s}} y_i \tag{3.93}$$

where s is a set of n units in the sample and $\tilde{s}$ denotes the set of non-sample units. The ratio estimator for population total $\hat{T}_R$ is unbiased, infact BLUE, under the prediction approach and a measure of precision attained by $\hat{T}_R$ as an estimator of T is given by $E_\xi(\hat{T}_R - T)^2$ This result from Royall (1970) is presented in Theorem 3.1.

Theorem 3.1

Under the model (3.91) the ratio estimator $\hat{T}_R$ is unbiased, that is, $E_\xi(\hat{T}_R - T) = 0$ and its error variance is given by

$$E_\xi(\hat{T}_R - T)^2 = \frac{N}{n}\left(1 - \frac{n}{N}\right)\frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s}\sigma^2 , \qquad (3.94)$$

where $\bar{x}_s$ is the average of n units in the sample s, $\bar{x}_{\tilde{s}}$ is the average of $(N-n)$ units not in the sample and $\bar{x}$, the average of all N units in the population.

From (3.94), we note that $E_\xi(\hat{T}_R - T)^2$ decreases as $\bar{x}_s$ increases which implies that if the model is correct, the ratio estimator is the optimal and the optimal sample consists of the n units having the largest x values. However, when such a sample is used, the ratio estimator can be badly biased under the departures from the model (3.91). For example, if $E(Y_i) = \alpha + \beta x_i$ then the ratio estimator has a bias $\frac{N(\bar{x}-\bar{x}_s)\alpha}{\bar{x}_s}$ which is large if $\bar{x}_s$ is much larger (or smaller) than $\bar{x}$. If $\bar{x} = \bar{x}_s$, that is, if the sample is balanced on x then this bias vanishes. If the sample is balanced on higher powers of x as well, say x^j for $j = 1,2,\ldots,J$, then the ratio estimator is unbiased for any jth degree polynomial regression model, see, for instance, Royall and Herson (1973a), and Royall and Cumberland (1981a).

From (3.94) we also note that the prediction approach provides a markedly different measure of precision in $\hat{T}_R$ compared to that given by the randomisation approach (3.92). An unbiased estimator of (3.94) may be obtained by replacing σ^2 by the weighted sum of squared residual statistic $\hat{\sigma}^2$, (Royall, 1971).

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_s (y_i - R_s x_i)^2/x_i , \qquad (3.95)$$

where $R_s = \frac{\bar{y}_s}{\bar{x}_s}$. This result is expressed as theorem 3.2.

Theorem 3.2

Under the model (3.91), $\hat{\sigma}^2$ given by (3.95) is ξ-unbiased estimator of σ^2 with the resulting unbiased variance estimator

$$\hat{V}_W = \frac{N^2}{n}\left(1-\frac{n}{N}\right)\frac{\bar{x}\bar{x}_s}{\bar{x}_s}\hat{\sigma}^2 \quad . \tag{3.96}$$

This variance estimator can, however, be badly biased if the variance model fails, that is, if the variance function is not linear in x , see, for instance, Royall and Eberhardt (1975). We will discuss this issue later when we compare the variance estimators (Section 3.3.2.1).

These results indicate that V_c given by (3.92), even if known, would not be appropriate as a measure of after sampling uncertainty in $\hat{T}_R$. The reason being that V_c is constant and hence provides the same measure of uncertainty in $\hat{T}_R$ for all samples whereas the error variance given by (3.94) is a decreasing function of $\bar{x}_s$. If we substitute the randomisation based variance estimator $\hat{V}_c$,

$$\hat{V}_c = \frac{N^2}{n}\left(1-\frac{n}{N}\right)\frac{1}{n-1}\sum_s (y_i - R_s x_i)^2 \quad , \tag{3.97}$$

for V_c in (3.92), the situation becomes less satisfactory, see, for instance, Royall (1971). Under the model (3.91), the expected value of $\hat{V}_c$ is

$$E_\xi(\hat{V}_c) = E_\xi\left(\frac{N}{f}(1-f)s_r^2\right) \quad ,$$

$$= \sigma^2 \frac{N}{f}(1-f)\bar{x}_s\left(1-\frac{1}{n}V_x^2(s)\right) \quad , \tag{3.98}$$

where

$$f = \frac{n}{N}\ , \quad s_r^2 = \frac{1}{n-1}\sum_s (y_i - R_s x_i)^2 \quad \text{and} \quad V_x^2(s) = \frac{1}{n-1}\sum_s (x_i - \bar{x}_s)^2/\bar{x}_s^2\ .$$

The relative bias in $\hat{V}_c$ is

$$RB = \frac{E_\xi(\hat{V}_c) - E_\xi(\hat{T}_R - T)^2}{E_\xi(\hat{T}_R - T)^2} = \frac{\left(1 - \frac{1}{n} V_x^2(s)\right)}{\left(\frac{\bar{x}\bar{x}_s}{\bar{x}_s^2}\right)} - 1 \doteq \left(\frac{\bar{x}_s^2}{\bar{x}\bar{x}_s} - 1\right)\ , \quad \text{for large } n \qquad (3.99)$$

(Royall and Eberhardt,1975)

The relative bias given by (3.99) increases as $\bar{x}_s$ increases and is positive or negative according as $\bar{x}_s$ is large or small. The relative bias vanishes only when $\bar{x} = \bar{x}_s$, that is, when the sample is balanced on x.

Royall and Eberhardt (1975) adjusted $\hat{V}_c$ to remove its bias under the model (3.91) and the resulting unbiased estimator is

$$\hat{V}_H = \hat{V}_c \left(\frac{\bar{x}_s \bar{x}}{\bar{x}_s^2}\right) \left(1 - \frac{1}{n} V_x^2(s)\right)^{-1}\ . \qquad (3.100)$$

These authors also showed that $\hat{V}_H$ is asymptotically equivalent to the jackknife variance estimator $\hat{V}_J$, for $g = n$.

The jackknife estimator features both in the randomisation and the model based approaches. In the latter because it conditions in some sense on the sample data.

$$\hat{V}_J = \left(\sum_{i=1}^{N} x_i\right)^2 \left(1 - \frac{n}{N}\right) \frac{g-1}{g} \sum_{j=1}^{g} \left\{\frac{\bar{y}_{(j)}}{\bar{x}_{(j)}} - \frac{1}{g}\sum_{j=1}^{g} \frac{\bar{y}_{(j)}}{\bar{x}_{(j)}}\right\}^2\ , \qquad (3.101)$$

where

$$\bar{y}_{(j)} = \frac{n\bar{y}_s - m\bar{y}_j}{n-m} = \bar{y}_s \left(1 - \frac{\bar{y}_j - \bar{y}_s}{(g-1)\bar{y}_s} \right) ,$$

$$\bar{x}_{(j)} = \frac{n\bar{x}_s - m\bar{x}_j}{n-m} = \bar{x}_s \left(1 - \frac{\bar{x}_j - \bar{x}_s}{(g-1)\bar{x}_s} \right) ,$$

and $\bar{x}_j$ and $\bar{y}_j$ are the averages of x and y, over m units in the jth group $(gm = n)$.

Let

$$D_{(j)} = \frac{\bar{y}_{(j)}}{\bar{x}_{(j)}} - \frac{1}{g} \sum_{j=1}^{g} \frac{\bar{y}_{(j)}}{\bar{x}_{(j)}} = \left(\frac{\bar{y}_{(j)}}{\bar{x}_{(j)}} - R_s \right) - \frac{1}{g} \sum_{j=1}^{g} \left(\frac{\bar{y}_{(j)}}{\bar{x}_{(j)}} - R_s \right) ,$$

$$= -\left(\frac{\bar{y}_j - R_s \bar{x}_j}{(g-1)\bar{x}_s} \left(1 + \frac{d_j}{g-1} \right) - \frac{1}{g} \sum_{j=1}^{g} \left\{ \frac{\bar{y}_j - R_s \bar{x}_j}{(g-1)\bar{x}_s} \left(1 + \frac{d_j}{g-1} \right) \right\} \right)$$

$$\left(\because \frac{1}{\bar{x}_{(j)}} = \frac{1}{\bar{x}_s} \left(1 + \frac{d_j}{g-1} \right) \quad \text{and} \quad d_j = \frac{\bar{x}_j - \bar{x}_s}{\bar{x}_{(j)}} \right) .$$

Now, writing $A_j = \bar{y}_j - R_s \bar{x}_j$, we have

$$D_{(j)} = -\frac{1}{\bar{x}_s} \left(\frac{A_j}{(g-1)} + \frac{A_j d_j}{(g-1)^2} - \frac{1}{g} \sum_{j=1}^{g} \frac{A_j d_j}{(g-1)^2} \right) \tag{3.102}$$

Substituting the version of $D_{(j)}$ from (3.102) in (3.101), the jackknife variance estimator $\hat{V}_J$ can be written as

$$\hat{V}_J = N^2 \left(\frac{\bar{x}}{\bar{x}_s} \right)^2 (1-f) \frac{1}{g} \left(\sum_{j=1}^{g} \frac{A_j^2}{(g-1)} + 2 \sum_{j=1}^{g} \frac{A_j^2 d_j}{(g-1)^2} + \frac{1}{(g-1)^3} \sum_{j=1}^{g} (A_j d_j - \frac{1}{g} \sum_{i=1}^{g} A_i d_i)^2 \right)$$

$$= N^2 (1-f) \left(\frac{\bar{x}}{\bar{x}_s} \right)^2 \frac{1}{g} \left(\sum_{j=1}^{g} \frac{(\bar{y}_j - R_s \bar{x}_j)^2}{(g-1)} + 2 \sum_{j=1}^{g} \frac{(\bar{y}_j - R_s \bar{x}_j)^2 d_j}{(g-1)^2} + \frac{1}{(g-1)^3} \sum_{j=1}^{g} \left\{ (\bar{y}_j - R_s \bar{x}_j) d_j \right.\right.$$

$$\left.\left. - \frac{1}{g} \sum_{i=1}^{g} (\bar{y}_i - R_s \bar{x}_i) d_i \right\}^2 \right)$$

Under the model (3.91), the expected value of $\hat{V}_J$ is given by

$$E_\xi(\hat{V}_J) = N^2(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 E_\xi\left[\sum_{j=1}^{g}\frac{(\bar{y}_j-R_s\bar{x}_j)}{(g-1)} + 2\sum_{j=1}^{g}\frac{(\bar{y}_j-R_s\bar{x}_j)d_j}{(g-1)^2}\right.$$

$$\left. + \frac{1}{(g-1)^3}\sum_{j=1}^{g}\left\{(\bar{y}_j-R_s\bar{x}_j)d_j - \frac{1}{g}\sum_{i=1}^{g}(\bar{y}_i-R_s\bar{x}_i)d_i\right\}^2\right] \qquad (3.103)$$

Now

$$E_\xi\sum_{j=1}^{g}(\bar{y}_j-R_s\bar{x}_j)^2 = E_\xi\left[\beta\bar{x}_j+\bar{e}_j-\hat{\beta}\bar{x}_j\right]^2 = E_\xi\sum_{j=1}^{g}\left[-(\hat{\beta}-\beta)\bar{x}_j+\bar{e}_j\right]^2$$

$$= E_\xi\left[(\hat{\beta}-\beta)^2\bar{x}_j^2-2\bar{e}_j(\hat{\beta}-\beta)\bar{x}_j+\bar{e}_j^2\right] \quad . \qquad (3.104)$$

$$E_\xi(\hat{\beta}-\beta)^2 = E_\xi\left[\frac{\sum_s y_i}{\sum_s x_i} - \beta\right]^2 = \frac{\sigma^2}{\sum_s x_i}$$

$$E_\xi(\bar{e}_j^2) = E_\xi\left[\sum_{i\epsilon j} e_i/n\right] = \frac{1}{n^2}\sum_{i\epsilon j}\sigma^2 x_i = \frac{\sigma^2}{n}\bar{x}_j$$

$$E_\xi(\bar{e}_j\hat{\beta}) = E_\xi\left[\frac{1}{n}\left(\sum_{i\epsilon j} e_i\right)\left(\frac{\sum_s y_i}{\sum_s x_i}\right)\right]$$

$$= E_\xi\left[\frac{1}{n}\sum_{i\epsilon j} e_i\beta + \left(\sum_{i\epsilon j} e_i\right)^2\right]$$

$$= \sigma^2/n \quad .$$

Substituting these expected values in (3.104), we get

$$E_\xi \sum_{j=1}^{g} (\bar{y}_j - R_s \bar{x}_j)^2 = \sum_{j=1}^{g} \left(\frac{\sigma^2 \bar{x}_j^2}{\sum_s x_i} - 2 \frac{\sigma^2}{n} \bar{x}_j + \frac{\sigma^2}{n} \bar{x}_j \right)$$

$$= \frac{\sigma^2}{n} \bar{x}_s \left(\sum_{j=1}^{g} (\bar{x}_j - \bar{x}_s)/\bar{x}_s^2 \right) \qquad (3.105)$$

Substituting (3.105) and the expected values for the other terms in (3.103) after simplification, we get

$$E_\xi(\hat{V}_J) = \sigma^2 \frac{N}{f} (1-f) \left(\frac{\bar{x}}{\bar{x}_s} \right)^2 \left\{ 1 + \sum_{j=1}^{g} \left(\frac{\bar{x}_j - \bar{x}_s}{\bar{x}_s} \right)^2 \Big/ g(g-1) + \sum_{j=1}^{g} \left(\frac{\bar{x}_j}{\bar{x}_s} d_j^2 \right) \Big/ g(g-1)^2 \right.$$

$$\left. - \sum_{j=1}^{g} \left(\left(\frac{\bar{x}_j}{\bar{x}_s} d_j \right) - \frac{1}{g} \sum_{i=1}^{g} \left(\frac{\bar{x}_i}{\bar{x}_s} d_i \right) \right)^2 \Big/ g(g-1)^3 \right\} \qquad (3.106)$$

For large g, the jackknife variance estimator $\hat{V}_J$ can be approximated as

$$\hat{V}_J = N^2 (1-f) \left(\frac{\bar{x}}{\bar{x}_s} \right)^2 \sum_{j=1}^{g} (\bar{y}_j - R_s \bar{x}_j)^2 / g(g-1) \ . \qquad (3.107)$$

when $g = n$ and n is large, $\hat{V}_J$ is closely related to $\hat{V}_C$ and $\hat{V}_H$, that is

$$\hat{V}_J \doteq \hat{V}_C \left(\frac{\bar{x}}{\bar{x}_s} \right)^2 \qquad (3.108)$$

Since from (3.100) we have $\hat{V}_H \doteq \hat{V}_C \dfrac{\bar{x}\tilde{x}_s}{\bar{x}_s^2}$, therefore when n is large and f is small, we get

$$\hat{V}_J \doteq \hat{V}_H \qquad (3.109)$$

Royall and Cumberland (1978b) proposed a variance estimator $\hat{V}_D$,

$$\hat{V}_D = \left(\frac{N}{n^2} \right) (N-n) \frac{\bar{x}\tilde{x}_s}{\bar{x}_s^2} \sum_s (y_i - R_s x_i)^2 \Big/ \left(1 - \frac{x_i}{n\bar{x}_s} \right) \ . \qquad (3.110)$$

This estimator is unbiased under model (3.91) and is very much like $\hat{V}_H$, that is, asymptotically equivalent to $\hat{V}_J$ and approximately unbiased for more general variance models, see for instance, Royall and Cumberland (1981a).

Royall and Eberhardt (1975), assuming that the error distribution is normal, computed the variances of $\hat{V}_J$, $\hat{V}_H$ and $\hat{V}_W$.

$$\mathrm{Var}(\hat{V}_J) \doteq \left(\frac{N^2}{g}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2\right)^2 2\sigma^4 \sum_{j=1}^{g} \bar{x}_j^2 / (g-1)^2 . \qquad (3.111)$$

This implies that for a fixed s and hence fixed n, the variance given by (3.111) is a decreasing function of g which is in agreement with Rao and Beegle's (1967) Monte Carlo results who also observed that g = n gives better results for $\hat{V}_J$.

$$\mathrm{Var}(\hat{V}_W) = \mathrm{Var}\left(\frac{N}{f}(1-f)\frac{\bar{\bar{x}}\tilde{\bar{x}}_s}{\bar{x}_s^2}\sum_s (y_i - R_s x_i)^2 / x_i (n-1)\right) ,$$

$$= \left(\frac{N}{f}(1-f)\frac{\bar{\bar{x}}\tilde{\bar{x}}_s}{\bar{x}_s}\right)^2 \frac{2\sigma^4}{n-1} . \qquad (3.112)$$

and

$$\mathrm{Var}(\hat{V}_H) = \mathrm{Var}\left(\frac{N}{f}(1-f)\frac{\bar{\bar{x}}\tilde{\bar{x}}_s}{\bar{x}_s^2}\sum_s \frac{(y_i - R_s x_i)^2}{(n-1)} \Big/ \left(1 - \frac{1}{n}V_x^2(s)\right)\right) ,$$

$$= \mathrm{Var}(\hat{V}_W)\frac{\left(\frac{\bar{x}_s^2}{\bar{x}_s^2}\right)\left(1 + \left\{\frac{1}{n}V_x^2(s) - \left(\frac{2}{n-1}\right)\left(\frac{\bar{x}_s^3}{\bar{x}_s\bar{x}_s^2} - 1\right)\right\}\right)}{\left(1 - \frac{1}{n}V_x^2(s)\right)^2} , \qquad (3.113)$$

where $\bar{x}_s^{\ell} = \dfrac{\sum_s x_i^{\ell}}{n}$

From (3.112), (3.113) and (3.94), it is apparent that under the assumption of normal distribution both the estimators $\hat{V}_H$ and $\hat{V}_W$ are consistent in the sense that their relative variances approach zero as n increases. The ratio of their variances for large n is approximately given by

$$\frac{\text{Var}(\hat{V}_H)}{\text{Var}(\hat{V}_W)} \doteq 1 + V_x^2(s) \tag{3.114}$$

This implies that under the assumption of normal distribution $\hat{V}_H$ is less efficient than $\hat{V}_W$, see, for instance, Royall and Eberhardt (1975).

3.3.2 Effects of Failures of the Simple Model:

A desirable property of any variance estimator is that it should be robust in its ability to approximate the true mean square error when the correct model is different from the assumed one. A basic model can fail under two situations. Firstly, the variance may not be proportional to x, that is, $\text{Var}(Y_i) = v_i \neq \sigma^2 x_i$. Secondly, the regression function may be $E(Y_i) = h(x_i) \neq \beta x_i$. These two forms of departures from the original model have independent effects on the variance estimators.

Each of the variance estimators considered earlier in Section 3.3.1 is of the form, $\hat{V} = C \sum_i l_i^2(\underline{y})$, where l_i is a linear function of $\underline{y}$ the vector of y's in the sample, and C is a constant which may depend on s but not on y. For example, $\hat{V}_W$ has $C_W = (\frac{N}{f})(1-f)\dfrac{\bar{x}\tilde{x}_s}{\bar{x}_s(n-1)}$ and $l_{i,W}(\underline{y}) = (y_i - R_s x_i)/\sqrt{x_i}$. Since $E\left[l(\underline{Y})^2\right] = \text{Var}\, l(\underline{Y}) + \left[E(l(\underline{Y}))\right]^2$, therefore, under this situation, we have

$$E(\hat{V}) = C \sum_i \{\mathrm{Var}\ell_i(\underline{Y}) + \left[E(\ell_i(\underline{Y}))\right]^2 ,$$

$$= C \sum_i \mathrm{Var}\ell_i(\underline{Y}) + C \sum_i \left[\ell_i(E\underline{Y})\right]^2 . \qquad (3.115)$$

From expression (3.115), we note that the expectation is the sum of two parts, first part depending on the variance function v and the second part obtained by replacing Y_i by its expected value $E(Y_i)$. If the regression function $E(Y_i) = \beta x_i$ is correct, then the second part in (3.115) is zero for all variance estimators under consideration in this section, see, for example, Royall and Eberhardt (1975). We discuss the effects of these departures from the original model separately following Royall and Eberhardt (1975).

3.3.2.1 Variance not Proportional to Size:

We assume that

$$E(Y_i) = \beta x_i \quad \text{and} \quad \mathrm{Var}(Y_i) = v_i \neq \sigma^2 x_i \qquad (3.116)$$

Under this model the error variance of $\hat{T}_R$ is

$$\text{(i)}\quad E_\xi(\hat{T}_R - T)^2 = E_\xi\left(\frac{\sum_s y_i}{\sum_s x_i}\sum_{i=1}^N x_i - \sum_{i=1}^N y_i\right)^2 ,$$

$$= E_\xi\left(\frac{\sum_s e_i}{\sum_s x_i}\sum_{i=1}^N x_i - \sum_{i=1}^N e_i\right)^2 ,$$

$$= E_\xi\left(\frac{\left(\sum_{i=1}^N x_i\right)^2 \bar{e}_s^2}{\bar{x}_s^2} - 2N\left(\sum_{i=1}^N x_i\right)\frac{\bar{e}_s\bar{e}_N}{\bar{x}_s} + N^2\bar{e}_N^2\right) ,$$

$$= \frac{\left(\sum_{i=1}^N x_i\right)^2}{\bar{x}_s^2}\mathrm{Var}(\bar{e}_s) - 2N\frac{\sum_{i=1}^N x_i}{\bar{x}_s}\mathrm{Cov}(\bar{e}_s,\bar{e}_N) + N^2\mathrm{Var}(\bar{e}_N) . \qquad (3.117)$$

Now

$$\mathrm{Var}(\bar{e}_s) = \mathrm{Var}\left(\frac{1}{n}\sum_s e_i\right) = \frac{1}{n^2}\sum_s v_i = \frac{1}{n}\bar{v}_s \quad .$$

$$\mathrm{Var}(\bar{e}_N) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N} e_i\right) = \frac{1}{N^2}\sum_{i=1}^{N} v_i \quad .$$

$$\mathrm{Cov}(\bar{e}_s,\bar{e}_N) = E_\xi(\bar{e}_s,\bar{e}_N) - E_\xi(\bar{e}_s)E(\bar{e}_N) = \frac{1}{nN}\sum_{i\in s} v_i = \frac{1}{N}\bar{v}_s \, .$$

Substituting these expected values in (3.117), we get

$$E_\xi(\hat{T}_R-T)^2 = \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{\bar{x}_s^2}\frac{1}{n}\bar{v}_s - 2N\frac{\sum_{i=1}^{N} x_i}{\bar{x}_s}\frac{1}{N}\bar{v}_s + \sum_{i=1}^{N} v_i \quad ,$$

$$= \frac{\left(\sum_{i=1}^{N} x_i\right)\bar{v}_s}{n\bar{x}_s^2}\left(\sum_{i=1}^{N} x_i - n\bar{x}_s\right) + \sum_{i=1}^{N} v_i - \frac{\sum_{i=1}^{N} x_i\bar{v}_s}{\bar{x}_s} \quad ,$$

$$= \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2}\bar{v}_s\left(1+f\frac{\bar{x}_s}{\bar{x}}\left\{\bar{x}_s\bar{v}_{\tilde{s}} - \bar{x}_{\tilde{s}}\bar{v}_s\right\} \Big/ \bar{x}_{\tilde{s}}\bar{v}_s\right) \quad ,$$

$$= C\left(1+f\left(\frac{\bar{x}_s}{\bar{x}}\right)\left\{\bar{x}_s\bar{v}_{\tilde{s}} - \bar{x}_{\tilde{s}}\bar{v}_s\right\} \Big/ \bar{x}_{\tilde{s}}\bar{v}_s\right) \quad , \qquad (3.118)$$

where

$$C = \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2}\bar{v}_s$$

(ii) Variance Estimator $\hat{v}_c$

Under the model (3.116) the expected value of the randomisation based variance estimator $\hat{V}_c$ is given by

$$E_\xi(\hat{V}_c) = \frac{N}{f}(1-f)E_\xi\left[\sum_s (y_i - R_s x_i)^2/(n-1)\right] ,$$

$$= \frac{N}{f}(1-f)E_\xi \sum_s \left[(\hat{\beta}-\beta)^2 x_i^2 - 2e_i(\hat{\beta}-\beta)x_i + e_i^2\right]/(n-1) ,$$

$$= \frac{N}{f}(1-f)\left[\sum_s x_i^2 \, \mathrm{Var}(\hat{\beta}-\beta) - 2\mathrm{Cov}(e_i,\hat{\beta})\sum_s x_i + \mathrm{Var}\left(\sum_s e_i^2\right)\right] / (n-1) . \qquad (3.119)$$

Now

$$E_\xi \sum_s (e_i^2) = \sum_s v_i$$

$$E_\xi(\hat{\beta}-\beta)^2 = E_\xi\left[\frac{\sum_s y_i}{\sum_s x_i} - \beta\right]^2 = \left(\sum_s v_i\right) / \left(\sum_s x_i\right)^2$$

$$\mathrm{Cov}(e_i,\hat{\beta}) = E_\xi(e_i\hat{\beta}) - E(e_i)E_\xi(\hat{\beta}) = E_\xi(e_i\hat{\beta}) = \frac{v_i}{\sum_s x_i} .$$

Substituting in (3.119) and simplifying we get

$$E_\xi(\hat{V}_c) = C\,\frac{\bar{x}_s^2}{\bar{x}_s\bar{x}}\left\{1 + \frac{1}{n}\left(V_x^2(s) - 2V_{xv}(s)\right)\right\} , \qquad (3.120)$$

where

$$C = \frac{N}{f}(1-f)\,\frac{\bar{x}_s\bar{x}}{\bar{x}_s^2}\,\bar{v}_s \quad \text{and} \quad V_{xv}(s) = \frac{\sum_s (v_i-\bar{v}_s)(x_i-\bar{x}_s)}{(n-1)\bar{x}_s\bar{v}_s} .$$

From (3.118) and (3.120) we note that when f is small and n is large,

$$E_\xi(\hat{T}_R - T)^2 \doteq C \quad \text{and} \quad E_\xi(\hat{V}_c) \doteq C\,\frac{\bar{x}_s^2}{\bar{x}_s\bar{x}}$$

The relative bias in $\hat{V}_c$ is

$$\frac{E_\xi(\hat{V}_c) - E_\xi(\hat{T}_R-T)^2}{E_\xi(\hat{T}_R-T)^2} = \frac{\bar{x}_s^2}{\bar{x}_s\tilde{\bar{x}}_s} - 1 \qquad (3.121)$$

This means that if the sample is balanced on x, the randomisation based variance estimator $\hat{V}_c$ is approximately unbiased under the model (3.116).

(iii) Weighted Least Squares Estimator $\hat{V}_W$

Following the same procedure, the expected value of the weighted least squares variance estimator $\hat{V}_W$ could be obtained and is given by

$$E_\xi(\hat{V}_W) = \frac{N}{f}(1-f)\frac{\bar{\bar{x}}\tilde{\bar{x}}_s}{\bar{x}_s} E_\xi \sum_s (y_i - R_s x_i)^2/x_i(n-1) ,$$

$$= \frac{N}{f}(1-f)\frac{\bar{\bar{x}}\tilde{\bar{x}}_s}{\bar{x}_s^2}\ddot{v}_s\left(1 + \frac{n}{n-1}\left\{\frac{\bar{x}_s}{\bar{v}_s}(\overline{v/x})_s - 1\right\}\right) . \qquad (3.122)$$

For large n, (3.122) reduces to

$$E_\xi(\hat{V}_W) \doteq C\frac{\bar{x}_s}{\bar{v}_s}(\overline{v/x})_s \quad , \quad \text{where} \quad (\overline{v/x})_s = \frac{1}{n}\sum_s (v_i/x_i) .$$

The relative bias in $\hat{V}_W$ is

$$\frac{E_\xi(\hat{V}_W) - E_\xi(\hat{T}_R-T)^2}{E_\xi(\hat{T}_R-T)^2} = \frac{\bar{x}_s}{\bar{v}_s}(\overline{v/x})_s - 1 . \qquad (3.123)$$

Thus the bias in $\hat{V}_W$ can be serious when the model fails and the form of this bias depends on the variance function $v(x)$.

(iv) Variance Estimator $\hat{V}_H$

The expected value of $\hat{V}_H$ under the model (3.116) is simply the product of $\left(\frac{\bar{x}_s\bar{x}}{\bar{x}_s^2}\right) \Big/ \left(1 - \frac{1}{n} V_x^2(s)\right)$ and the expected value of $\hat{V}_c$. Thus, we have

$$E_\xi(\hat{V}_H) = \frac{N}{f}(1-f)\frac{\bar{x}_s\bar{x}_s}{(\bar{x}_s)^2}\bar{v}_s\left(\frac{1+\frac{1}{n}\{V_x^2(s)-2V_{xv}(s)\}}{\left(1-\frac{1}{n}V_x^2(s)\right)}\right),$$

$$= C\left(\frac{1+\frac{1}{n}\{V_x^2(s)-2V_{xv}(s)\}}{1-\frac{1}{n}V_x^2(s)}\right). \qquad (3.124)$$

From (3.124), we note that for large n, $\hat{V}_H$ is approximately unbiased under the models where the variance function is not proportional to x. Also from expressions (3.121) and (3.124), we note that when $\bar{x}_s = \bar{x}$ the estimators $\hat{V}_c$ and $\hat{V}_H$ are both unbiased but when $\bar{x}_s \neq \bar{x}$ the variance estimator $\hat{V}_c$ becomes biased whereas $\hat{V}_H$ retains its unbiasedness.

(v) Jackknife Variance Estimator $\hat{V}_J$

Under the model (3.116) the expected value of $\hat{V}_J$ is given by

$$E_\xi(\hat{V}_J) = C\left(\frac{\bar{x}}{\bar{x}_s}\right)\left[1 + \frac{\sum_{j=1}^{g}\left(\frac{(\bar{x}_j-\bar{x}_s)}{\bar{x}_s}\right)^2}{g(g-1)} + \frac{\sum_{j=1}^{g}\left(\bar{v}_j d_j^2 - 2\sum_{j=1}^{g}\bar{v}_j - \left(\frac{\bar{v}_s}{\bar{x}_s}\right)\bar{x}_j\right)(d_j\bar{x}_j/\bar{x}_s)}{g(g-1)^2\bar{v}_s}\right.$$

$$\left. - \frac{\sum_{j=1}^{g}\left(2\bar{v}_j - \left(\frac{\bar{v}_s}{\bar{x}_s}\right)\bar{x}_j\right)(d_j^2\bar{x}_j/\bar{x}_s) - \frac{1}{g}\sum_{i=1}^{g}\left(2\bar{v}_i - \left(\frac{\bar{v}_s}{\bar{x}_s}\right)\bar{x}_i\right)d_i\sum_{j=1}^{g} d_j\bar{x}_j/\bar{x}_s}{g(g-1)^3\bar{v}_s}\right] \qquad (3.125)$$

Thus when f is small, n is large and g is large as well, the variance estimator $\hat{V}_J$ is approximately unbiased, that is,

$$E_\xi(\hat{V}_J) \doteq C \ . \qquad (3.126)$$

From these results, we conclude that the variance estimators $\hat{V}_H$ and $\hat{V}_J$ are approximately unbiased for all variance functions when n is large, f is small and g is large as well. For $g = n$, the variance estimator $\hat{V}_H$ approximates $\hat{V}_J$. Since $\hat{V}_D$ is like $\hat{V}_H$ similar behaviour is expected from this estimator. The randomisation based variance estimator $\hat{V}_c$ is biased regardless of the variance function. However, when the sample is balanced on x, this estimator is approximately unbiased for any variance function. The weighted least squares variance estimator $\hat{V}_W$ is unbiased for all samples, but only when the variance function is linear in x.

3.3.2.2 Implications of the Failure of Regression Models:

Earlier in Section 3.3.2, we have mentioned that the expectation of each of the variance estimators considered is the sum of two terms, the first depends on the variance function and not on the regression function, whereas the second depends on the regression function and could be obtained by substituting $E(Y_i)$ for Y_i in the estimator. We illustrate this point for the variance estimator $\hat{V}_c$, the others could be handled in exactly the same way. Consider that $E(Y_i) = hx_i$ and $Var(Y_i) = v_i$

$$\hat{V}_c = \frac{N}{f}(1-f)\frac{1}{n-1}\sum_{i=1}^{n}(y_i - R_s x_i)^2 \ .$$

From (3.115), we know that

$$E_\xi(\hat{V}_C) = K \sum_s \mathrm{Var}(y_i - R_s x_i) + K \sum_s \left\{E_\xi(y_i - R_s x_i)\right\}^2$$

where $K = \frac{N}{f}(1-f)\frac{1}{n-1}$,

$$= K \sum_s \mathrm{Var}(y_i - R_s x_i) + K \sum_s \left[E_\xi(y_i) - \frac{\sum_s E_\xi(y_i)}{\sum_s x_i} x_i\right] ,$$

$$= K \sum_s \mathrm{Var}(y_i - R_s x_i) + K \sum_s \left[hx_i - \frac{\sum_s hx_i}{\sum_s x_i} x_i\right] ,$$

$$= C \frac{\bar{x}_s}{\bar{x}_s \tilde{\bar{x}}} \left\{1 + \frac{1}{n}\left(V_x^2(s) - 2V_{xv}(s)\right)\right\} + \frac{N^2}{n}(1-f) \sum_s \left[hx_i - \frac{\bar{h}_s}{\bar{x}_s} x_i\right]/(n-1) ,$$

$$= C \frac{\bar{x}_s^2}{\bar{x}_s \tilde{\bar{x}}} \left\{1 + \frac{1}{n}\left(V_x^2(s) - 2V_{xv}(s)\right)\right\} + b_C . \qquad (3.127)$$

where

$$C = \frac{N}{f}(1-f)\frac{\bar{x}_s \tilde{\bar{x}}}{\bar{x}_s^2} , \quad \bar{h}_s = \frac{1}{n}\sum_s hx_i$$

and

$$b_C = \frac{N^2}{n}(1-f) \sum_s \left(hx_i - \frac{\bar{h}_s}{\bar{x}_s} x_i\right)^2/(n-1) .$$

We note that expression (3.127) equals Expression (3.120) + b_C. Similarly, we can calculate expressions for the other variance estimators. Since we are interested only in second part, therefore we will write only that part of the expression. Thus,

$$b_W = \frac{N}{f}(1-f)\frac{\bar{x}_s \tilde{\bar{x}}}{\bar{x}_s} \sum_s \left[hx_i - \left(\frac{\bar{h}_s}{\bar{x}_s}\right) x_i\right]^2 \Big/ (n-1)x_i \qquad (3.128)$$

$$b_H = \frac{N}{f}(1-f)\bar{x}_s\bar{x}\sum_s\left(hx_i - \left(\frac{\bar{h}_s}{\bar{x}_s^2}\right)x_i\right)^2 \Big/ \left\{(n-1)\bar{x}_s^2\left(1-\left(\frac{1}{n}\right)V_x^2(s)\right)\right\} \tag{3.129}$$

$$b_J = \frac{N}{f}(1-f)\bar{x}^2 n(g-1)\sum_{j=1}^{g}\left(\left(\frac{\bar{h}_{(j)}}{\bar{x}_{(j)}}\right) - \frac{1}{g}\sum_{i=1}^{g}\left(\frac{\bar{h}_{(i)}}{\bar{x}_{(i)}}\right)\right)^2 \Big/ g \tag{3.130}$$

and

$$b_{MSE} = N^2\bar{x}^2\left(\frac{\bar{h}_s}{\bar{x}_s} - \frac{\bar{h}}{\bar{x}}\right)^2 , \tag{3.131}$$

where

$$\bar{h} = \frac{1}{N}\sum_{i=1}^{N} hx_i$$

The expressions b_{MSE}, b_C, b_W, b_H and b_J reflect the implications of departure from the regression model $E(Y_i) = \beta x_i$ on the actual MSE and the variance estimators. We note that the failure of the regression model $E(Y_i) = \beta x_i$ appears to affect each of the estimators under study in approximately the same way and for well balanced samples this effect tends to be conservative (producing positive bias), see, for example, Royall and Eberhardt (1975).

3.4 Summary and Conclusions

Accurate variance estimates, which are trustworthy under broad conditions, are necessary for survey sampling inference. Variance estimation has been a problem in the past while dealing with large scale complex surveys, particularly when the variance estimates were required for non-linear estimators and for analytical statistic.

One can identify two main approaches to variance estimation, that is, the randomisation based approach and the prediction theory approach. The randomisation based approach promotes the use of randomisation methods, that is, the methods justified by the randomisation distribution induced by the sampling design. This approach has provided a variety of methods for variance estimation under complex situations as discussed in Section 3.2. These methods can be classified under four main groups. Firstly, by a careful consideration of the specific function f it is possible that an estimable expression can be derived for $\text{Var } f(Y_1, Y_2, \ldots, Y_N)$. Secondly, it is possible to employ the numerical calculations for the appropriate partial derivatives of f by using Taylor Series method. The third class of methods, which we have discussed, belongs to pseudoreplication techniques developed by McCarthy and others during the 1960's. The fourth class of methods is known as the jackknife procedures which are based on the idea of pseudovalues. The problem with the randomisation based variances is that these are unconditional, that is, they do not depend on the sample actually observed and hence are not suitable for a meaningful inference from an observed sample.

The prediction approach, by utilizing the regression models, provides a suitable framework for studying the properties of the estimators from an observed sample. For example, the prediction approach has pointed out serious flaws in the randomisation based variance estimator $\hat{V}_c$, for the ratio estimate, when the variance function is linear in x . These flaws continue to persist under more general models unless the sample is well balanced on x . The predication approach has also provided robust-variance estimators like $\hat{V}_H$ and $\hat{V}_D$ which perform well under more general models. The ordinary

jackknife variance estimator $\hat{V}_J$ also shares the properties of $\hat{V}_H$ and $\hat{V}_D$. It is hoped that the weighted jackknife procedure may provide the long desired robust estimator that could limit the influence of any one observation and thus lead to stable variance estimates. The weighted jackknife is the main theme of our study.

CHAPTER 4

ESTIMATION OF CORRELATION FROM COMPLEX SAMPLES

4.1 Introduction

One of the areas of considerable research activity in sample surveys during the last two decades has been to explore methods of computing sample estimates of the variances of non-linear statistics computed from complex surveys. The most familiar example is the ratio estimator but initiative has also been taken for statistics like partial and multiple regression and correlation coefficients which are often used in the analysis of survey data. Correlation analysis not only has an important bearing on sociological studies based on sample surveys but also forms the basis of many multivariate techniques.

Statistical inference, whether it pertains to confidence limits or tests of significance, is based on the assumptions that the sample estimate of the population parameter is approximately unbiased, an approximately unbiased estimate of variance of this estimate is computable from the sample and that the distribution of the ratio of the difference of the sample estimate and its expected value to its estimated standard error could reasonably be approximated by a standard normal distribution within symmetric limits. These assumptions are fairly true if the sample design is a simple random selection of elements, the estimate is a simple statistic like mean or total and the sample size is large. However, departures from simple random selection and descriptive parameter estimates may render the above assumptions dubious.

Many sample surveys are characterised by complex clustered and stratified samples and often depart drastically from a simple random model, see, for instance, Kish (1965, Sec. 14.2). The important question, therefore, arises about how far standard statistical procedures based implicitly on simple random sampling are appropriate when the population has a complex structure. The pioneer work for estimating analytical statistics such as correlation and regression coefficients from complex surveys is due to Frankel (1971) and Kish and Frankel (1974).

Kish and Frankel (1974) studied descriptive P-inference. Besides, Kish and Frankel, the descriptive P-inference has also been studied by Jönrup and Rennermalm (1976) and Shah et al., (1977). The finite population parameter of interest is defined as

$$\underline{B} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y} \ , \tag{4.1}$$

where

$$\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} , \ \underline{B} = \begin{pmatrix} B_1 \\ \vdots \\ B_P \end{pmatrix} \text{ and } \underline{X} = \begin{pmatrix} X_{11} & \cdots & X_{1P} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{NP} \end{pmatrix} , \ \underline{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix} .$$

The error vector $\underline{e}$ has the property that $\sum_{i=1}^{N} e_i = 0$.

Recall that there are two main approaches to making a descriptive inference for samples drawn from the finite populations, the randomisation based approach and a model based approach. The model based approach considers the finite population of N units as a random sample from an infinite population. The population data can be characterised by the regression equation,

$$\underline{Y} = \underline{X}\beta + \underline{\varepsilon} \quad , \tag{4.2}$$

where, $\underline{Y}$ is an N x 1 vector, $\underline{X}$ is an N x P matrix, β is a P x 1 vector of parameters and $\underline{\varepsilon}$ has a probability distribution ξ with $E_\xi(\underline{\varepsilon}) = 0$ and $Var_\xi(\underline{\varepsilon}) = \sigma^2\underline{V}$. A sample of n units is drawn by using the sampling scheme P_s. This sample is assumed to be independent of both Y and X and is ignored for making inference from the model (4.2). A descriptive parameter of interest in this case is the generalised least square solution of (4.2), i.e.

$$\tilde{\underline{B}} = (\underline{X}'\underline{V}^{-1}\underline{X})\underline{X}'\underline{V}^{-1}\underline{Y} \quad . \tag{4.3}$$

In case the population consists of clusters, the covariance matrix will include terms that represent the intra-cluster correlation coefficients. When $\underline{V} = \underline{I}$, the ordinary least squares solution of (4.2) gives the finite population parameter $\underline{B}$. If $\underline{y}_s$ and $\underline{x}_s$ represent the sample values of $\underline{Y}$ and $\underline{X}$, then the ordinary least squares estimator of $\underline{B}$ or $\tilde{\underline{B}}$ is

$$\tilde{\underline{B}} = (\underline{X}_s'\underline{X}_s)^{-1}\underline{X}_s'\underline{Y}_s \, . \tag{4.4}$$

which is ξ-unbiased for $\underline{B}$ or $\tilde{\underline{B}}$ or β.

Smith (1981a) gives an illuminating discussion on both types of inference. He argues that "the interpretation that the survey design does not matter in the model based approach is not correct because design always affects the efficiency of an inference in terms of criteria such as mean squares error." Rubin (1976) describes the situations in which the design can be ignored. Further details regarding these issues could be found in Fuller (1973, 1975), Hartley and Sielken (1975) and Holt, Smith and Winter (1980).

Hartley and Sielken (1975) classify regression models according to these two criteria which is reproduced here as Table 4.1.

Table 4.1 Sampling Theories Classified by Sampling Procedure and Target Parameters.

Target Parameter	Sampling Procedure	
	Repeated sampling from a fixed finite population	Repeated two step sampling from an infinite population
Parameters of finite population B or $\tilde{B}$	Classical finite population sampling theory. P-inference	Superpopulation theory for finite population sampling. ξ-inference.
Parameters of infinite superpopulation β	Infeasible	Inference on infinite population parameters from two step sampling procedure Analytic ξ-inference.

The purpose of this chapter is to consider the problems encountered with the estimation of correlation from sample surveys with complex designs. Frankel (1971) studied the behaviour of simple, partial and multiple correlation and regression from clustered stratified samples. In Section 4.2, we give a brief review of his study for the estimation of correlation and regression for finite populations. We discuss some alternative estimators for the estimation of correlation coefficient from stratified populations in Section 4.3. Non-simple random sampling leads to the biased estimation of the population correlation coefficient. Section 4.4 deals with the consequences of the sampling method on the estimation of correlation and regression as elucidated by Warren (1971). In Section 4.5, we investigate the effects of stratification on the behaviour of simple, partial and multiple correlation through a simulation study. Finally, in Section 4.6, we give a summary of the main conclusions.

4.2 Frankel's Study

Using data from the March 1967 Current Population Survey, Frankel (1971) investigated the behaviour of complex estimators for three clustered stratified sample designs of 6, 12 and 30 strata with two primary sampling units being drawn from each stratum. The population of his study consisted of 3240 primary sampling units each of 14.1 households on the average. The study included 45,737 households in all. This study involves a clustered stratified sampling design with proportional stratification and equal probability of selection which implies a self-weighting sample. Simple random sampling within each stratum is assumed. The strata were formed on the basis of geographical regions and population density (rural - urban). A summary of the sample design appears as Table 4.2.

Table 4.2 Summary of Sample Designs Used by Frankel (1971).

Design	Primary units per Strata	Number of Strata	Number of Primary Units Selected	Number of Samples Selected	Average Sample Size	Coefficient of Variation of Sample Size
I	540	6	12	300	170.3	0.190
II	270	12	24	300	339.5	0.130
III	104	30	60	200	846.5	0.074

The study comprised eight variables for each household which were divided into two groups to form two multiple regression equations, i.e.,

$$Y_i = B_o + B_1X_{1i} + B_2X_{2i} + B_3X_{3i} + \varepsilon_i \qquad (4.5)$$

$$Y'_i = B'_o + B_4X_{4i} + B_5X_{5i} + B_6X_{6i} + \varepsilon'_i \qquad (4.6)$$

The details of the variables and their identification numbers are as under.

Identification Number	Variable
1	Number of persons in household under 18.
2	Number of persons in household.
3	Sex of household head.
Y	Total income of household head.
Y'	Total income of household.
4	Age of household head.
5	Number of persons in household in labour force.
6	Years of school completed by household head.

The estimates studied were based on the following two mutually exclusive groups of variables.

Group	Predictand	Predictors
I	Y	1, 2, 3.
II	Y'	4, 5, 6.

No estimates were formed from variable pairs not within the same group. Thus correlation was computed between variables 1 and 2 but not between 1 and 4.

Frankel considers a finite population of N elements where the values $Y_i, X_{1i}, X_{2i}, X_{3i}, \ldots, X_{ki}$ are associated with the ith element of the population. For measuring the relationship between $\underline{Y}$ and $\underline{X}$ the population parameters B_j are defined such that $\sum_{i=1}^{N} (Y_i - \sum_{j=0}^{K} B_j X_{ji})^2$ is minimum, where $X_{oi} = 1$, that is B_j are the ordinary least squares regression coefficients. The estimates $\hat{B}_j$ are obtained by

treating the sample of n vectors from the population of N vectors as if it is a simple random sample. Thus minimising $\sum_{i=1}^{n} (Y_i - \sum_{j=0}^{K} \hat{B}_j X_{ji})^2$ with respect to $\hat{B}_j$, we get

$$\sum_{j=0}^{K} U_{mj} \hat{B}_j = V_m \ , \ m = 0,1,2,\ldots,K \qquad (4.7)$$

where,

$$U_{mj} = \sum_{i=1}^{n} X_{mi} X_{ji} \quad \text{and} \quad V_m = \sum_{i=1}^{n} Y_i X_{mi} \ .$$

Without assuming that the population under study conforms exactly or more closely to the model being used, the correlation coefficient population values are interpreted as the proportion of variability in Y explained by the particular model under consideration, see, for instance, Frankel (1971, p.20).

The ordinary least squares regression theory, on the other hand, assumes the model

$$Y_i = \underline{\beta}^T X_i + \varepsilon_i \ . \qquad (4.8)$$

where $\underline{X}_i = (X_{1i}, X_{2i}, \ldots, X_{ki})^T$, and $\underline{\beta} = (\beta_1, \ldots, \beta_k)^T$. It further assumes that

(a) $E(\varepsilon_i / X_i) = 0$, for all i

(b) $Var(\varepsilon_i / X_i) = \sigma^2$ for all i

(c) $Cov(\varepsilon_i, \varepsilon_j / X_i, X_j) = 0$ for all $i \neq j$

(d) $\varepsilon_i \sim N(0, \sigma^2)$.

Kish and Frankel argue that this and other well specified models give desirable results but the problem is how to reconcile this model with the real population being studied. For example, clustered

selection introduces positive correlation between the errors of the model and the assumption (c) fails. Besides, these correlations have serious effects on the statistics based on complex samples and also pose formidable theoretical difficulties. For example, as a broader class of analytical statistics, we may think of the coefficient of multiple linear regression. The derivation of their distribution implies independence of observations and not of the variables whereas clustering destroys the independence of observations.

4.2.1 Estimation of Complex Statistics

Kish and Frankel (1974) investigated the properties of the first order and second order estimates for three clustered stratified designs. The first order estimates in their study comprised 8 ratio means, 12 difference of ratio means, 12 simple correlations, 6 partial correlation coefficients, 2 multiple correlation coefficients and 8 partial regression coefficients. For these estimates, they compared the relative biases (ratio of bias to parameter). The formulation is described as under:-

Let X_{phai} denote the value of the Pth variable for the ith element in the a^{th} primary unit of the hth stratum for the population where $p = 1, 2,\ldots,P$, $i = 1, 2,\ldots,I$, $a = 1, 2,\ldots,A$ and $h = 1,2,\ldots,H$. The corresponding values for the sample are denoted by x_{phai} where all subscripts have the same range except $a = 1, 2$. We have

(i) Means:

$$\bar{X}_p = \sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} X_{phai} / HAI_{ah} \quad \text{(parameter)}$$

$$\bar{x}_p = \sum_{h=1}^{L} \sum_{a=1}^{2} \sum_{i=1}^{I_{ah}} x_{phai} / 2HI_{ah} \quad \text{(sample estimate)} \quad (4.9)$$

(ii) Correlation Coefficient:

$$\rho_{pp'} = \frac{\text{Covariance } PP'}{(\text{Var}P\text{Var}P')^{\frac{1}{2}}} = \frac{C_{pp'}}{(S_p^2 S_{p'}^2)^{\frac{1}{2}}} \quad \text{, (parameter)} \qquad (4.10)$$

where

$$C_{pp'} = \left[\sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} X_{phai} X_{p'hai} - \frac{\sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} X_{phai} \sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} X_{p'hai}}{N} \right] .$$

$$S_p^2 = \left[\sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} X_{phai}^2 - \frac{\left(\sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} X_{phai} \right)^2}{N} \right] . \qquad (4.11)$$

The sample estimates are:

$$\hat{\rho}_{pp'} = \frac{C_{pp'}}{(s_p^2 s_{p'}^2)^{\frac{1}{2}}} \qquad (4.12)$$

where

$$C_{pp'} = \left[\sum_{h=1}^{H} \sum_{a=1}^{2} \sum_{i=1}^{I_{ah}} x_{phai} x_{p'hai} - \frac{\sum_{h=1}^{H} \sum_{a=1}^{2} \sum_{i=1}^{I_{ah}} x_{phai} \sum_{h=1}^{H} \sum_{a=1}^{2} \sum_{i=1}^{I_{ah}} x_{p'hai}}{n} \right] ,$$

and

$$s_p^2 = s_{p'}^2 = \left[\sum_{h=1}^{H} \sum_{a=1}^{2} \sum_{i=1}^{I_{ah}} x_{phai}^2 - \left(\sum_{h=1}^{H} \sum_{a=1}^{2} \sum_{i=1}^{I_{ah}} x_{phai} \right)^2 / n \right] \qquad (4.13)$$

(iii) Regression Coefficients

Let the variables be recorded so that $X_1, X_2, \ldots, X_{k-1}$ are k-1 predictors and X_k is the predictand, then

$$\underset{(k\times 1)}{\underline{\beta}} = (\beta_0, \beta_1, \ldots, \beta_{k-1})^T . \qquad (4.14)$$

where $\underset{k\times k}{\underline{\beta}} = \underset{k\times k}{\underline{A}^{*-1}} \underset{k\times k}{\underline{V}}$ and $\underline{A}^*$ is the $k \times k$ matrix with ℓqth element

$$\{A^*_{\ell q}\} = \sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} X_{\ell hai} X_{qhai} , \qquad \begin{matrix} \ell = 1,2,\ldots,k-1 \\ q = 1,2,\ldots,k-1 \end{matrix}$$

and $\underline{V}$ is $k \times 1$ vector with ℓth element $\{V_\ell\} = \sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} X_{khai} X_{\ell hai}$

$\ell = 0, 1, \ldots, k-1$.

The corresponding sample estimates are given by

$$\underset{K\times 1}{\hat{\underline{\beta}}} = (\hat{\beta}_o, \hat{\beta}_1, \ldots, \hat{\beta}_{k-1})^T . \qquad (4.16)$$

where

$\underset{k\times 1}{\hat{\underline{\beta}}} = \underset{k\times k}{\overset{*-1}{a}} \underset{k\times 1}{\underline{v}}$ and $\overset{*}{a}$ is a $k \times k$ matrix with ℓqth element

$$\{a^*_{\ell q}\} = \sum_{h=1}^{H} \sum_{a=1}^{A} \sum_{i=1}^{I_{ah}} x_{\ell hai} x_{qhai} , \quad \ell = 0, \ldots, k-1, \quad q = 0, \ldots, k-1$$

and $\underline{v}$ is a k x 1 vector with ℓth element $\{v_\ell\} = \sum_{h=1}^{H} \sum_{a=1}^{2} \sum_{i=1}^{I_{ah}} x_{khai} x_{\ell hai}$

(iv) Partial Correlation Coefficient

The partial correlation coefficient is defined as the correlation between two variables when all others are fixed. For the population values the partial correlation between X_k and X_ℓ is defined as:

$$R_{k\ell.1,\ldots,k-1,\ k+1\ldots\ldots\ell} = \frac{\{C_{k\ell}\}}{(\{C_{kk}\}\{C_{\ell\ell}\})^{\frac{1}{2}}} \quad \text{(Parameter)} \qquad (4.17)$$

where $C_{\ell q}$ is the ℓqth element of a $k \times k$ matrix $\underline{C} = \underline{\bar{R}}^1$ and $\underline{R}$ is a $k \times k$ population correlation matrix with ℓqth element

$$\{R_{\ell q}\} = \frac{S_{\ell q}}{(S_\ell^2 S_q^2)^{\frac{1}{2}}} , \quad \ell = 1,2,\ldots,k, \quad q = 1,2,\ldots,k$$

The sample estimates are:

$$r_{k/\ell} = \frac{\{c_{k\ell}\}}{(\{c_{kk}\}\{c_{\ell\ell}\})^{\frac{1}{2}}} . \qquad (4.18)$$

where $\{c_{\ell q}\}$ is the ℓqth element of a $k \times k$ matrix $\underline{c} = \underline{r}^{-1}$ and $\underline{r}$ is a $k \times k$ sample correlation matrix with ℓqth element

$$\{r_{\ell q}\} = s_{\ell q}/(s_{\ell}^2 s_q^2)^{\frac{1}{2}} \quad , \quad \begin{matrix} \ell = 1,2,\ldots,k \\ q = 1,2,\ldots,k \end{matrix}$$

(v) Multiple correlation:

Multiple correlation coefficient may be defined as the simple correlation between Y and its linear regression $B_1X_1 + B_2X_2$ $B_1X_1 + B_2X_2,\ldots,+B_kX_k$ on X_1, X_2, X_3 ,...,X_k . Thus the population multiple correlation coefficient is given by

$$R_{k(\ell)} = [1 - \frac{1}{C_{kk}}]^{\frac{1}{2}} \quad \text{(Parameter)} \tag{4.19}$$

Let

$$\underset{k\times k}{\underline{R}} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ \ell \\ \vdots \\ k \end{array} \begin{pmatrix} 1 & 2 & 3 & \ldots. q & \ldots. k \\ 1 & \rho_{12} & \rho_{13} \ldots. & \rho_{1q} \ldots. & \rho_{1k} \\ \rho_{21} & \rho_{22} & \rho_{23} \ldots. & \rho_{2q} \ldots. & \rho_{2k} \\ \cdot & \cdot & \cdot \ldots. \cdot \ldots. & \cdot & \\ \cdot & \cdot & \cdot \ldots. \cdot \ldots. & \cdot & \\ \rho_{\ell 1} & \rho_{\ell 2} & \rho_{\ell 3} & \rho_{\ell q} & \rho_{\ell k} \\ \cdot & \cdot & \cdot \ldots. \cdot \ldots. & \cdot & \\ \cdot & \cdot & \cdot \ldots. \cdot \ldots. & \cdot & \\ \rho_{k1} & \rho_{k2} & \rho_{k3} & \rho_{kq} & \rho_{kk} \end{pmatrix}$$

and $\underset{k\times k}{\underline{C}} = \underset{k\times k}{\underline{R}}^{-1} =$

$$\begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ \ell \\ \vdots \\ k \end{array} \begin{pmatrix} 1 & 2 & \ldots. & q & \ldots. & k \\ \frac{z_{11}}{|z|} & \frac{z_{12}}{|z|} & \ldots. & \frac{z_{1q}}{|z|} & \ldots. & \frac{z_{1k}}{|z|} \\ \frac{z_{21}}{|z|} & \frac{z_{22}}{|z|} & \ldots. & \frac{z_{2q}}{|z|} & \ldots. & \frac{z_{2k}}{|z|} \\ \cdot & \cdot & \ldots. & \cdot & \ldots. & \cdot \\ \cdot & \cdot & \ldots. & \cdot & \ldots. & \cdot \\ \frac{z_{\ell 1}}{|z|} & \frac{z_{\ell 2}}{|z|} & & \frac{z_{\ell q}}{|z|} & & \frac{z_{\ell k}}{|z|} \\ \cdot & \cdot & \ldots. & \cdot & \ldots. & \cdot \\ \cdot & \cdot & \ldots. & \cdot & \ldots. & \cdot \\ \cdot & \cdot & \ldots. & \cdot & \ldots. & \cdot \\ \frac{z_{k1}}{|z|} & \frac{z_{k2}}{|z|} & \ldots. & \frac{z_{kq}}{|z|} & \ldots. & \frac{z_{kk}}{|z|} \end{pmatrix}$$

$$\underset{k\times k}{C} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ \ell \\ \vdots \\ k \end{array} \begin{pmatrix} 1 & 2 & \cdots & 3 & \cdots & q & \cdots & k \\ C_{11} & C_{12} & & C_{13} & & C_{1q} & & C_{1k} \\ C_{21} & C_{22} & \cdots & C_{23} & \cdots & C_{2q} & \cdots & C_{2k} \\ C_{31} & C_{32} & \cdots & C_{33} & \cdots & C_{3q} & \cdots & C_{3k} \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdots & \cdot \\ C_{\ell 1} & C_{\ell 2} & & C_{\ell 3} & & C_{\ell q} & & C_{\ell k} \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdots & \cdot \\ C_{k1} & C_{k2} & & C_{k3} & & C_{kq} & & C_{kk} \end{pmatrix}$$

$$1 - R^2_{1(2,\ldots,k-1)} = \frac{|z|}{z_{11}} = (1-\rho^2_{12})(1-\rho^2_{13.2})\ldots(1-\rho^2_{1k.2}) \ .$$

$$R^2_{k(1,\ldots.k-1)} = \left(1- \frac{|z|}{z_{11}}\right) = \left(1- \frac{1}{C_{kk}}\right) \ ,$$

$$R_{k(1,2,\ldots.k-1)} = \left(1- \frac{|z|}{z_{11}}\right)^{\frac{1}{2}} = \left(1- \frac{1}{C_{kk}}\right)^{\frac{1}{2}}$$

The sample estimates are:

$$r_{k(\ell)} = \left(1 - \frac{1}{c_{kk}}\right)^{\frac{1}{2}} \ . \qquad (4.20)$$

4.2.2 Results:

In Frankel's study the relative biases for 5 in 12 simple correlations for 6 strata design and 3 in 12 simple correlations for 12 strata design are greater than 10 per cent. These relative biases, however, decrease as we proceed to a 30 strata design for which the relative biases for the individual simple correlations are less than 5 per cent. For the partial and multiple correlations the relative biases are much higher as compared to simple correlations. For example, the relative biases for the partial correlations $r_{y1.23}$ for the 6 strata design and $r_{y2.13}$ for 12 strata design are greater than 25 percent and 15 percent respectively. Similarly the relative biases for the multiple correlations $r_{y(123)}$ is greater than 20 percent and that for $r_{y'(456)}$ is greater than 12 percent for 6 strata designs (see Appendix I). The average absolute relative biases for the simple, partial and multiple correlations decrease with the increasing sample

size and are comparatively higher for the multiple correlations except for 30 strata design for which these are higher for the partial correlations than those for the multiple correlations. Kish and Frankel (1974) conjectured that the high relative biases for the multiple correlations were due to the basic defects of the estimator rather than the complexity of the selection design (Table 4.3).

Table 4.3 Average Relative Biases in Complex Samples for Simple, Partial, Multiple Correlations and Regression Coefficients

Estimator	No in Average	Relative Biases		
		6 Strata design 300 samples	12 strata design 300 samples	30 strata design 200 samples
Simple Correlation	12	0.06972	0.05399	0.01748
Partial Correlation	6	0.12333	0.08365	0.05863
Multiple Correlation	2	0.16002	0.11115	0.04670
Regression Coefficients	8	0.04978	0.03320	0.02776

Source: Kish and Frankel (1974) Table 2.

Kish and Frankel advocating the estimation of B (Finite population regression coefficient) argue that some researchers could benefit by knowing the amount of the relationship between Y and X that could be explained or accounted for by a linear model like $y_j = \alpha + \beta x_j + \varepsilon_j$, $i=1, 2,\ldots,N$ with ε_j assumed i.i.d random variables (with zero mean and variance σ^2) or some other model. They stress that it is not necessary to assume models as described above because the purpose is only to know how much of the variability in Y can be explained by the particular model we try. However, a different point of view is expressed by others(Konijn, H.S., and Smith, T.M.F. in the discussion of Kish and Frankel, 1974) who argue that the researcher's primary objective is to discover some (at least approximate) relationship between Y and X and the amount of

variation that can not be explained by such a relationship would as a natural consequence form the basis for the evaluation of the results obtained. This implies that the primary interest of the most of the users is in the estimation of the parameters of an appropriate model rather than the estimation of the descriptive measures like B ,

$$B = \frac{S_{XY}}{S_X^2} = \frac{\sum_{i=1}^{N} (X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{N} (X_i-\bar{X})^2} \quad . \tag{4.21}$$

Frankel employed estimates of (4.21) which are valid for simple random sampling. Since the sampling designs used in his study are self-weighting, these estimates are consistent for large samples. With unequal probabilities π_i , S_{XY} and S_X^2 are estimated term by term using the method of substitution, and the estimates of population expression B are given (Brewer and Mellor, 1973) as under:

$$\hat{B} = \frac{\sum_{i=1}^{n} \pi_i^{-1} (\sum_{i=1}^{n} \pi_i^{-1} y_i x_i) - (\sum_{i=1}^{n} \pi_i^{-1} y_i)(\sum_{i=1}^{n} \pi_i^{-1} x_i)}{(\sum_{i=1}^{n} \pi_i^{-1})(\sum_{i=1}^{n} \pi_i^{-1} x_i^2) - (\sum_{i=1}^{n} \pi_i^{-1} x_i)^2} \quad . \tag{4.22}$$

where n is the number of sample units and π_i is the total probability of inclusion of the ith unit in the sample.

4.3 Some Alternative Estimators:

For a stratified sampling design where n_h elements are drawn from each stratum without replacement and with equal probabilities at each draw, the parent population correlation coefficient ρ is given by

$$\rho_{(X,Y)} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{N_h} (x_{hi}-\bar{X})(y_{hi}-\bar{Y})}{\left[\sum_{h=1}^{L} \sum_{i=1}^{N_h} (x_{hi}-\bar{x})^2 \sum_{h=1}^{L} \sum_{i=1}^{N_h} (y_{hi}-\bar{y})^2\right]^{\frac{1}{2}}} = \frac{T_{XY}}{(T_{XX}.T_{YY})^{\frac{1}{2}}} \quad , \tag{4.23}$$

where

$$\bar{X} = \sum_{h=1}^{L} \sum_{i=1}^{N_h} x_{hi}/N \quad \text{and} \quad \bar{Y} = \sum_{h=1}^{L} \sum_{i=1}^{N_h} y_{hi}/N \ .$$

In this situation, Frankel's estimates for the population expressions T_{XY}, T_{XX} and T_{YY} can be written as

$$\hat{T}_{XY}(K.F) = \sum_{h=1}^{L} \sum_{i=1}^{n_h} (x_{hi}-\bar{x})(y_{hi}-\bar{y}) \ ,$$

where

$$\bar{x} = \sum_{h=1}^{L} \sum_{i=1}^{n_h} x_{hi}/n \quad \text{and} \quad \bar{y} = \sum_{h=1}^{L} \sum_{i=1}^{n_h} y_{hi}/n$$

Thus, we have

$$\hat{T}_{XY}(K.F) = \sum_{h=1}^{L} \sum_{i=1}^{n_h} (x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h) + \sum_{h=1}^{L} n_h(\bar{x}_h-\bar{x})(\bar{y}_h-\bar{y}) \ . \qquad (4.24)$$

$$\hat{T}_{XX}(K.F) = \sum_{h=1}^{L} \sum_{i=1}^{n_h} (x_{hi}-\bar{x}_h)^2 + \sum_{h=1}^{L} n_h(\bar{x}_h-\bar{x})^2 \ . \qquad (4.25)$$

and

$$\hat{T}_{YY}(K.F) = \sum_{h=1}^{L} \sum_{i=1}^{n_h} (y_{hi}-\bar{y}_h)^2 + \sum_{h=1}^{L} n_h(\bar{y}_h-\bar{y})^2 \ . \qquad (4.26)$$

The estimator for ρ is given by

$$\hat{\rho} \quad \frac{\hat{T}_{XY(K.F)}}{[\hat{T}_{XX}(K.F).\hat{T}_{YY}(K.F)]^{\frac{1}{2}}} \ . \qquad (4.27)$$

4.3.1 Koop's Estimator

Consider a finite population divided into L strata, each stratum having N_h identifiable elements (h = 1,2,...,L) with pairs of

measurable values (x_{hi}, y_{hi}), $(h=1,2,\ldots,L;\ i=1,2,\ldots,N_h)$ so that $\sum_{h=1}^{L} N_h = N$. Let n_h elements be drawn from each stratum without replacement and with equal probabilities at each draw. The total sample size is $\sum_{h=1}^{L} n_h = n$. In this case in order to estimate the population correlation ρ given by the expression (4.23), Koop derived unbiased estimates for the population values T_{XY}, T_{XX} and T_{YY} which can be written as

$$\hat{T}_{XY(K)} = \sum_{h=1}^{L}\left[\frac{N_h-1}{n_h-1} - \left(1-\frac{N_h}{N}\right)\frac{N_h-n_h}{n_h(n_h-1)}\right] S_h(x,y) + \sum_{h=1}^{L} N_h(\bar{x}_h-\bar{x}_{st})(\bar{y}_h-\bar{y}_{st}) \tag{4.28}$$

where $S_h(x,y) = \sum_{i=1}^{n_h}(x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h)$, $\bar{x}_{st} = \sum_{h=1}^{L}\left(\frac{N_h}{N}\right)\bar{x}_h$, $\bar{y}_{st} = \sum_{h=1}^{L}\left(\frac{N_h}{N}\right)\bar{y}_h$, $\bar{x}_h$ and $\bar{y}_h$ are the unbiased estimates of $\bar{X}_h$ and $\bar{Y}_h$ respectively.

Similarly we have

$$\hat{T}_{XX}(K) = \sum_{h=1}^{L}\left[\frac{N_h-1}{n_h-1} - \left(1-\frac{N_h}{N}\right)\frac{N_h-n_h}{n_h(n_h-1)}\right] S_h(x,x) + \sum_{h=1}^{L} N_h(\bar{x}_h-\bar{x}_{st})^2 \tag{4.29}$$

$$\hat{T}_{YY}(K) = \sum_{h=1}^{L}\left[\frac{N_h-1}{n_h-1} - \left(1-\frac{N_h}{N}\right)\frac{N_h-n_h}{n_h(n_h-1)}\right] S_h(y,y) + \sum_{h=1}^{L} N_h(\bar{y}_h-\bar{y}_{st})^2 \tag{4.30}$$

For a self weighting design we have, $\frac{n_h}{n} = \frac{N_h}{N}$ and substituting in (4.28), (4.29) and (4.30), we get

$$\hat{T}_{XY}(K) = \frac{N}{n}\sum_{h=1}^{L} S_h(x,y) + \sum_{h=1}^{L}\frac{N_h(N-n)}{n(N_h n-N)} S_h(x,y) + \frac{N}{n}\sum_{h=1}^{L} n_h(\bar{x}_h-\bar{x})(\bar{y}_h-\bar{y})$$

$$= \frac{N}{n}\left[\sum_{h=1}^{L}\sum_{i=1}^{n_h}(x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h) + \sum_{h=1}^{L} n_h(\bar{x}_h-\bar{x})(\bar{y}_h-\bar{y})\right]$$

$$+ \sum_{h=1}^{L}\frac{N_h(N-n)}{n(N_h\cdot n-N)} S_h(x,y),$$

$$= \frac{N}{n}\hat{T}_{XY}(K.F) + \sum_{h=1}^{L}\frac{n_h}{n^2}\frac{N-n}{(n_h-1)}S_h(x,y),$$

$$= \frac{N}{n}\hat{T}_{XY}(K.F) + \sum_{h=1}^{L}\frac{N_h}{n}(1-\frac{n}{N})\widehat{Cov}_h(x,y),$$

(4.31)

where $\widehat{Cov}_h(x,y) = \sum_{i=1}^{n_h}(x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h)/(n_h-1)$. In a similar manner, we obtain

$$\hat{T}_{XX}(K) = \frac{N}{n}\hat{T}_{XX}(K.F) + \sum_{h=1}^{L}\frac{N_h}{n}(1-\frac{n}{N})\hat{V}_h(x,x) \qquad (4.32)$$

where $\hat{V}_h(x,x) = \sum_{i=1}^{n_h}(x_{hi}-\bar{x}_h)^2/(n_h-1)$

$$\hat{T}_{YY}(K) = \frac{N}{n}\hat{T}_{YY}(K.F) + \sum_{h=1}^{L}\frac{N_h}{n}(1-\frac{n}{N})\hat{V}_h(y,y) \qquad (4.33)$$

where $\hat{V}_h(y,y) = \sum_{i=1}^{n_h}(y_{hi}-\bar{y})^2/(n_h-1)$

From expressions (4.31), (4.32) and (4.33) we note that the bias in Kish and Frankel's estimates of covariances and variances is $\sum_{h=1}^{L}\frac{N_h}{n}(1-\frac{n}{N})\hat{Cov}_h(x,y)$, $\sum_{h=1}^{L}\frac{N_h}{n}(1-\frac{n}{N})\hat{V}_h(x,x)$, $\sum_{h=1}^{L}\frac{N_h}{n}(1-\frac{n}{N})\hat{V}_h(y,y)$ respectively. The estimator of the correlation coefficient ρ in this case is therefore given by

$$\hat{\rho} = \frac{\hat{T}_{XY}(K.F) + \sum_{h=1}^{L}\frac{n_h}{n}(1-\frac{n}{N})\hat{Cov}_h(x,y)}{[(\hat{T}_{XX}(K.F) + \sum_{h=1}^{L}\frac{n_h}{n}(1-\frac{n}{N})\hat{V}_h(x,x)(\hat{T}_{YY}(K.F) + \sum_{h=1}^{L}\frac{n_h}{n}(1-\frac{n}{N})\hat{V}_h(y,y))]^{\frac{1}{2}}} \qquad (4.34)$$

4.3.2 Wakimoto's Estimator

In order to estimate the population correlation coefficient Wakimoto (1971) also derived unbiased estimates of the population expressions σ_{xx}, σ_{yy}, σ_{xy} , where

$$\sigma_{xx} = \sum_{h=1}^{L} W_h \sigma_{hxx} + \sum_{h=1}^{L} W_h (\mu_{hx}-\mu_x)^2 \quad .$$

$$\sigma_{yy} = \sum_{h=1}^{L} W_h \sigma_{hyy} + \sum_{h=1}^{L} W_h (\mu_{hy}-\mu_y)^2 \quad .$$

$$\sigma_{xy} = \sum_{h=1}^{L} W_h \sigma_{hxy} + \sum_{h=1}^{L} W_h (\mu_{hx}-\mu_x)(\mu_{hy}-\mu_y) \quad .$$

and

$$\mu_x = \sum_{h=1}^{L} W_h \mu_{hx} \; ; \; \mu_y = \sum_{h=1}^{L} W_h \mu_{hy} \quad .$$

The unbiased estimators of these population variances are given by

$$\hat{S}_{xy}(W) = \sum_{h=1}^{L} \frac{W_h}{n_h(n_h-1)} \sum_{i<j}^{n_h} (x_{hi}-x_{hj})(y_{hi}-y_{hj}) + \sum_{h=1}^{L} W_h (\bar{x}_h-\bar{x})(\bar{y}_h-\bar{y})$$

$$- \sum_{h=1}^{L} \frac{W_h(1-W_h)}{n_h^2(n_h-1)} \sum_{i<j} (x_{hi}-x_{hj})(y_{hi}-y_{hj}) \quad ,$$

$$= \sum_{h=1}^{L} \left[\frac{W_h}{n_h(n_h-1)} - \frac{W_h(1-W_h)}{n_h^2(n_h-1)} \right] \sum_{i<j} (x_{hi}-x_{hj})(y_{hi}-y_{hj})$$

$$+ \sum_{h=1}^{L} W_h (\bar{x}_h-\bar{x})(\bar{y}_h-\bar{y}) \quad ,$$

$$= \sum_{h=1}^{L} \left[\frac{n_h}{n(n_h-1)} - \frac{n-n_h}{n^2(n_h-1)} \right] \sum_{i=1}^{n_h} (x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h)$$

$$+ \frac{1}{n} \sum_{h=1}^{L} n_h (\bar{x}_h-\bar{x})(\bar{y}_h-\bar{y}) \quad . \qquad (4.35)$$

Similarly, we have

$$\hat{S}_{XX}(W) = \sum_{h=1}^{L} \left(\frac{n_h}{n(n_h-1)} - \frac{n-n_h}{n^2(n_h-1)} \right) \sum_{i=1}^{n_h} (x_{hi}-\bar{x}_h)^2 + \frac{1}{n} \sum_{h=1}^{L} n_h(\bar{x}_h-\bar{x})^2 , \quad (4.36)$$

$$\hat{S}_{YY}(W) = \sum_{h=1}^{L} \left(\frac{n_h}{n(n_h-1)} - \frac{n-n_h}{n^2(n_h-1)} \right) \sum_{i=1}^{n_h} (y_{hi}-\bar{y}_h)^2 + \frac{1}{n} \sum_{h=1}^{L} n_h(\bar{y}_h-\bar{y})^2 . \quad (4.37)$$

For a selfweighting design, the above expressions can be written as

$$\hat{S}_{XY}(W) = \sum_{h=1}^{L} \left(\frac{1}{n} + \frac{N_h}{n(N_h n-N)} \right) \sum_{i=1}^{n_h} (x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h) + \frac{1}{n} \sum_{h=1}^{L} n_h(\bar{x}_h-\bar{x})(\bar{y}_h-\bar{y}) ,$$

$$= \sum_{h=1}^{L} \frac{1}{n} S_h(x,y) + \frac{1}{n} \sum_{h=1}^{L} n_h(\bar{x}_h-\bar{x})(\bar{y}_h-\bar{y}) + \sum_{h=1}^{L} \frac{N_h}{n} \frac{1}{(N_h n-N)} S_h(x,y) ,$$

$$= \frac{1}{n} \hat{T}_{XY}(K.F) + \sum_{h=1}^{L} \frac{n_h}{n} \hat{Cov}h(x,y) . \quad (4.38)$$

$$\hat{S}_{XX}(W) = \frac{1}{n} \hat{T}_{XX}(K.F) + \sum_{h=1}^{L} \frac{n_h}{n} \hat{V}_h(x,x) . \quad (4.39)$$

$$\hat{S}_{YY}(W) = \frac{1}{n} \hat{T}_{YY}(K.F) + \sum_{h=1}^{L} \frac{n_h}{n} \hat{V}_h(y,y) . \quad (4.40)$$

The estimator of the correlation coefficient ρ is now given by

$$\hat{\rho} = \frac{\hat{T}_{XY}(K.F) + \sum_{h=1}^{L} n_h \widehat{Cov}_h(x,y)}{\left[\left(\hat{T}_{XX}(K.F) + \sum_{h=1}^{L} n_h \hat{V}_h(x,x)\right)\left(\hat{T}_{YY}(K.F) + \sum_{h=1}^{L} n_h \hat{V}_h(y,y)\right)\right]^{\frac{1}{2}}} \qquad (4.41)$$

4.3.3 Relationship Between Koop and Wakimoto's Estimates

From expression (4.31), we have

$$\hat{T}_{XY}(K) = \sum_{h=1}^{L} N\left(\frac{1}{n} + \frac{N_h}{n}\frac{(1-n/N)}{(N_h n - N)}\right) S_h(x,y) + \frac{1}{n}\sum_{h=1}^{L} n_h(\bar{x}_h - \bar{x})(\bar{y}_h - \bar{y}) \qquad (4.42)$$

If the finite population correction factor is ignored, then we get

$$\hat{T}_{XY}(K) = N[\hat{S}_{XY}(W)] \qquad (4.43)$$

$$\hat{T}_{XX}(K) = N[\hat{S}_{XX}(W)] \qquad (4.44)$$

$$\hat{T}_{YY}(K) = N[\hat{S}_{YY}(W)] \qquad (4.45)$$

4.3.4 Discussion of Results

We note that for a self-weighting design, Koop and Wakimoto's results for the estimation of covariances and variances reduce to those of Kish and Frankel estimates plus bias. This shows that stratification causes bias in the estimation of correlation and that the bias may be of substantial magnitude.

Kish and Frankel have mainly concentrated on the effects of clustering on the variances and have ignored the effects of stratification on the bias. Thus they have been lead to the misinterpretation of results. Also

from expressions (4.43), (4.44) and (4.45) we note that if the finite population correction is ignored then, for a self-weighting design, Koop's estimates reduce to those of Wakimoto multiplied by N. From these theoretical results we conclude that the high relative biases in Kish and Frankel's estimates of correlation may be due to the effect of stratification.

4.4 *Effect of Sampling Method on the Estimation of Correlation and Regression*

Warren (1971) draws a distinction between correlation and regression. Correlation studies the extent of relationship between two variables whereas regression describes the form of relationship between the specified values of one variable (independent) and the means of all corresponding values of the second variable (dependent). In regression, we will seldom be dealing with strictly random samples whereas a random sample is necessary for the conventional theory of the correlation coefficient to be applicable. In regression the general practice is to get a wide range of independent variables so as to have a more precise estimate of the regression coefficient or to point out the existence of the curvilinear relationship. Thus the sample correlation coefficient computed from regression data will usually not be a valid estimate of the population correlation coefficient. Therefore, designing studies both for correlation and regression simultaneously presents a contradiction because if the correlation coefficient is a valid estimate of the population value then it implies that the estimate of regression equation is inefficient. Similarly, if the estimate of regression is the best possible, then the estimate of correlation is likely to be invalid. It means that a good regression design causes bias in the estimates of correlations and a good correlation design

(simple random sampling) gives an inefficient estimate of regression.

In Frankel's study we observe that the mean absolute relative biases for the multiple regression coefficients are 5, 3 and 3 per cent respectively for 6, 12 and 30 strata designs whereas for partial correlation coefficients these relative biases are 12, 8 and 6 per cent and for the multiple correlations 16, 11 and 5 per cent respectively for 6, 12 and 30 strata designs. This is in agreement with Warren's argument that the estimates of correlations obtained from restricted sampling will over estimate the population value substantially.

The survey design in Frankel's study used geographical region and the population density (urban-rural) as the design variable (Z) for stratification. In practice we might expect that regional or area type (urban-rural) differences do exist which would affect the relationship between the variables. For example, sex of the household head may have a strong relationship with the density of population due to social reasons. Warren designs on x variable rather than on the design variable **Z** . We discuss briefly salient features of Warren's results.

4.4.1 Warren's Analysis

Let (X, Y) be a random sample having a bivariate normal distribution with parameters μ_1, μ_2, σ_1^2, σ_2^2 and ρ , where $\rho = \frac{\mu_{12}}{\sigma_1\sigma_2}$ is the correlation coefficient and μ_{12} is the covariance. The frequency function with zero means $(\mu_1 = 0, \mu_2 = 0)$ can be written as

$$dF(x,y) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x}{\sigma_1}\right)^2 - 2\rho\left(\frac{xy}{\sigma_1\sigma_2}\right) + \left(\frac{y}{\sigma_2}\right)^2\right\}\right] dx\,dy \qquad (4.46)$$

Consider a sample of size N, $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$. The correlation coefficient can be written as

$$r = \frac{\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^{N} (X_i - \bar{X})^2 \sum_{i=1}^{N} (Y_i - \bar{Y})^2\right]^{\frac{1}{2}}} \tag{4.47}$$

where

$$\bar{X} = \sum_{i=1}^{N} X_i \quad \text{and} \quad \bar{Y} = \sum_{i=1}^{N} Y_i \ .$$

Warren (1971) derived expressions for the expectation of the square of the correlation coefficient r^2 and precision of the estimated regression under simple random sampling and selective sampling. He studied the impact of these sampling methods on the bias of correlation and the precision of regression. His main thesis is that non-simple random sampling can cause bias in the estimation of correlation and this bias may be of substantial magnitude. We give a brief summary of his results as under.

4.4.1.1 Simple Random Sampling

(a) Expectation of r^2

Under simple random sampling, for a sample size N, the expected value of the multiple correlation coefficient ρ^2 is given by

$$E(r^2) = 1 - \frac{N-2}{N-1}(1-\rho^2) F\left\{1, 1, \frac{1}{2}(N+1), \rho^2\right\} \ . \tag{4.48}$$

where $F(a,b,c;x)$ is the hypergeometric function (Kendall and Stuart, 1970, Vol.2, p.353). $E(r^2)$ may be evaluated as follows

$$F\left\{1, 1, \frac{1}{2}(N+1), \rho^2\right\} = \frac{\Gamma\{\frac{1}{2}(N+1)\}}{\Gamma\{\frac{1}{2}(N-1)\}} \int_0^1 \frac{(1-t)^{\frac{1}{2}(N-3)}}{(1-\rho^2 t)}\, dt,$$

and hence

$$E(r^2) = 1 - \frac{1}{2}(N-2)(1-\rho^2)\int_0^1 \frac{(1-t)^{\frac{1}{2}(N-3)}}{(1-\rho^2 t)}\,dt \ . \qquad (4.49)$$

(b) Precision of the Estimated Regression:

The estimator of the slope and intercept of the regression are

$$\hat{\beta} = \frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2} \quad \text{and} \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \ .$$

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\Sigma(X_i-\bar{X})^2} \ , \ \mathrm{Var}(\hat{\alpha}) = \frac{\Sigma X_i^2}{N\Sigma(X_i-\bar{X})^2}\sigma^2 \ , \ \mathrm{Cov}(\hat{\alpha},\hat{\beta}) = -\frac{\Sigma X_i}{N\Sigma(X_i-\bar{X})^2}\sigma^2 \ . \qquad (4.50)$$

If $\hat{Y}_0 = \hat{\alpha} + \hat{\beta}\,x_0$ is the estimated value of the dependent variable at $x = x_0$, then

$$\mathrm{Var}(\hat{Y}_0/X) = \left[\frac{1}{N} + \frac{(X_0-\bar{X})^2}{\Sigma(X_i-\bar{X})^2}\right]\sigma^2 \qquad (4.51)$$

Now, using the fact that $\mathrm{Var}(Y) = E[V(Y/X)] + V[E(Y/X)]$, we have

$$E\left(\mathrm{Var}(\hat{Y}_0/X)\right) = \left[N^{-1} + \frac{X_0^2+\sigma_1^2/N}{(N-3)\sigma_1^2}\right]\sigma^2 \qquad (4.52)$$

Now, letting $Z_0\sigma_1 = X_0$, we can write (4.52) as

$$E\left(\mathrm{Var}(\hat{Y}_0/X)\right) = \left[N^{-1} + \frac{Z_0^2+N^{-1}}{(N-3)}\right]\sigma^2 \ .$$

Thus

$$\mathrm{Var}(\hat{Y}) = \left[N^{-1} + \frac{Z^2+N^{-1}}{N-3}\right]\sigma^2 \ . \qquad (4.53)$$

1.4.1.2 Selective Sampling

(c) Expectation of r^2

Let the sample consist of n observations made at g equally spaced values of X so that $N = ng$ and

$$X = -(g-1)a, -(g-3)a, \ldots, (g-3)a, (g-1)a \qquad (4.54)$$

Under this situation we have

$$\sum_{i=1}^{n} (X_i-\bar{X})^2 = \frac{na^2(g-1)g(g+1)}{3} = na^2\phi(g) \tag{4.55}$$

where

$$\phi(g) = \frac{(g-1)g(g+1)}{3}$$

$$\sum_{i=1}^{n} (X_i-\bar{X})(Y_i-\bar{Y}) = na\ \Sigma C_j\bar{Y}_j \ , \tag{4.56}$$

$$\sum_{i=1}^{n} (Y_i-\bar{Y})^2 = \sum_{j=1}^{g}\sum (Y_j-\bar{Y}_j)^2 + n \sum_{j=1}^{g} (\bar{Y}_j-\bar{Y})^2$$

$$= [W_1+nW_2+n(\Sigma C_j\bar{Y}_j)^2/\Sigma C_j^2] \ , \tag{4.57}$$

where, $W_1 \sim \sigma^2\chi^2_{g(n-1),0}$ $\quad W_2 \sim \frac{\sigma^2}{n}\chi^2_{g-2,0}$ and $\Sigma C_j\bar{Y}_j \sim N\left(\rho\frac{\sigma_2}{\sigma_1}\phi(g), \frac{\sigma^2}{n}\phi(g)\right)$

Now

$$r = \frac{na\Sigma C_j\bar{Y}_j}{[(na^2\phi(g))(W_1+nW_2+n(\Sigma C_j\bar{Y}_j)^2/(\Sigma C_j^2))]^{\frac{1}{2}}} \tag{4.58}$$

and,

$$r^2 = \frac{n(\Sigma C_j\bar{Y}_j)^2}{\phi(g)\left[W+\frac{n}{\phi(g)}(\Sigma C_j\bar{Y}_j)^2\right]} = \frac{Z}{W+Z} \tag{4.59}$$

where

$$W_1+nW_2 = \sigma^2\chi^2_{g(n-1),0} + n\frac{\sigma^2}{n}\chi^2_{g(n-1),0} = \sigma^2\chi^2_{f,0}$$

and

$$f = g(n-1)+(g-2)=N-2, \quad Z = \frac{n}{\phi(g)}(\Sigma C_j\bar{Y}_j)^2 \sim \sigma^2\chi^2_{1,\delta}$$

and

$$\delta = n\phi(g)\left(\frac{a}{\sigma_1}\right)^2\frac{\rho^2}{1-\rho^2} \ .$$

$$E(r^2) = 1-e^{-\delta/2}\left\{\frac{f}{f+1}\right\}F\left\{\frac{1}{2}(f+1), \frac{1}{2}(f+3), \frac{1}{2}\delta\right\} \ . \tag{4.60}$$

where $F(a,b;x)$ denotes the confluent hypergeometric function. $E(r^2)$ can be evaluated using

$$F\left\{\frac{1}{2}(f+1),\ \frac{1}{2}(f+3),\ \frac{1}{2}\delta\right\} = \frac{\Gamma\left[\frac{1}{2}(f+3)\right]}{\Gamma\left[\frac{1}{2}(f+1)\right]} \int_0^1 t^{\frac{1}{2}(f-1)} \exp\left(\frac{1}{2}\delta t\right) dt;$$

so that

$$E(r^2) = 1 - \frac{1}{2}(f+1) \int_0^1 t^{\frac{1}{2}(f-1)} \exp\left\{-(1-t)\ \frac{1}{2}\delta\right\} dt \ . \qquad (4.61)$$

(d) Precision of the Estimated Regression:-

From (4.51), we have:

$$\mathrm{Var}(\hat{Y}_o/X) = \left[N^{-1} + \frac{(X_o - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]\sigma^2 \ .$$

Under selective sampling $\bar{X} = 0$ and $\Sigma(X_i - \bar{X})^2 = na^2\phi(g)$, therefore

$$\mathrm{Var}(\hat{Y}_0/X) = \left[N^{-1} + \frac{X_o^2}{na^2\phi(g)}\right]\sigma^2 \qquad (4.62)$$

Supposing that the range of interest extends from $x = -\lambda\sigma_1$ to $x = \lambda\sigma_1$, i.e. letting $\lambda\sigma_1 = (g-1)a$ and putting $x_0 = z_0\sigma_1$, in (4.62), we get

$$\mathrm{Var}(\hat{Y}_0) = N^{-1}\left[1 + \frac{3(g-1)}{(g+1)}\left(\frac{z_0}{\lambda}\right)^2\right]\sigma^2 = N^{-1}\left[1 + \frac{3(g-1)}{g+1}\left(\frac{z}{\lambda}\right)^2\right]\sigma^2 \ . \qquad (4.63)$$

From these results we note that,

(i) Under selective sampling, $E(r^2)$ is a function of N, g, λ and ρ^2 , whereas under random sampling it is a function of N and ρ^2 only.

(ii) Similarly, for a given σ^2 , under selective sampling $\mathrm{Var}(\hat{Y})$ is a function of N, g, Z and λ , while under simple random sampling it is a function of N and Z only.

(iii) The efficiency of selective sampling relative to simple random sampling, for a given Z , is defined as the ratio:

$$R.Ef = \frac{\text{Variance Simple Random Sample}}{\text{Variance Selective Sample}}$$

This relative efficiency is a function of Z and of N, g and λ.

Warren conducted an exhaustive numerical study of the behaviour of $E(r^2)$ with varying N, g, λ and ρ^2 and also of the relative efficiency with varying N, g, λ and Z. We present a few graphs from his study in order to illustrate the behaviour of the bias in case of correlation and the relative efficiency in case of regression. The details of these graphs are given as under:

Fig. 4.1 $E(r^2)$vs.n , selected ρ^2, $\lambda = 2$ } $g = 2$

Fig. 4.2 $E(r^2)$vs.n , selected λ , $\rho^2 = 0.5$ } $g = 2$

Fig. 4.3 $E(r^2)$vs.g , selected ρ^2 , $\lambda = 2$ $N = 30$

Fig. 4.4 R.E vs. Z , selected λ , $n = 10$ } $g = 2$

Fig. 4.5 R.E vs. g , selected n , $\lambda = 2$ } $g = 2$

Fig. 4.6 R.E. vs.g , selected Z , $\lambda = 2$ $N = 30$

From these graphs we note that selective sampling gives biased estimates of ρ^2 and this bias can be large. Moreover, bias is an increasing function of the range of the independent variable but a decreasing function of the number of points within the range and also of the sample size but does not approach the true population value in any case. In case of regression, the graphs indicate that the relative efficiency increases as the range of independent variables increases but decreases as the number of points within the range increase. It is usually greater than unity except when the range is small. The relative efficiency also decreases with increasing sample size. Thus, these results show that the bias for correlation and the precision for regression are contradictory. One should be cautious in determining which of the correlations or regressions is appropriate to the problem and select the sampling method accordingly. For example, for $N = 30$, $g = 10$ and $\lambda = 2$,

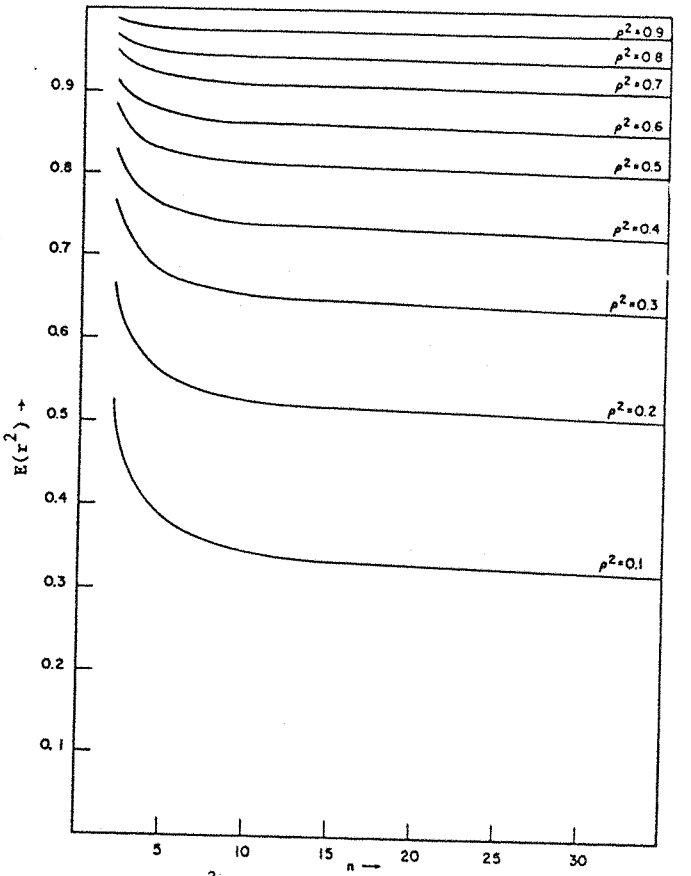

Fig. 4.1 $E(r^2)$ vs. n, $\lambda = 2, g = 2$

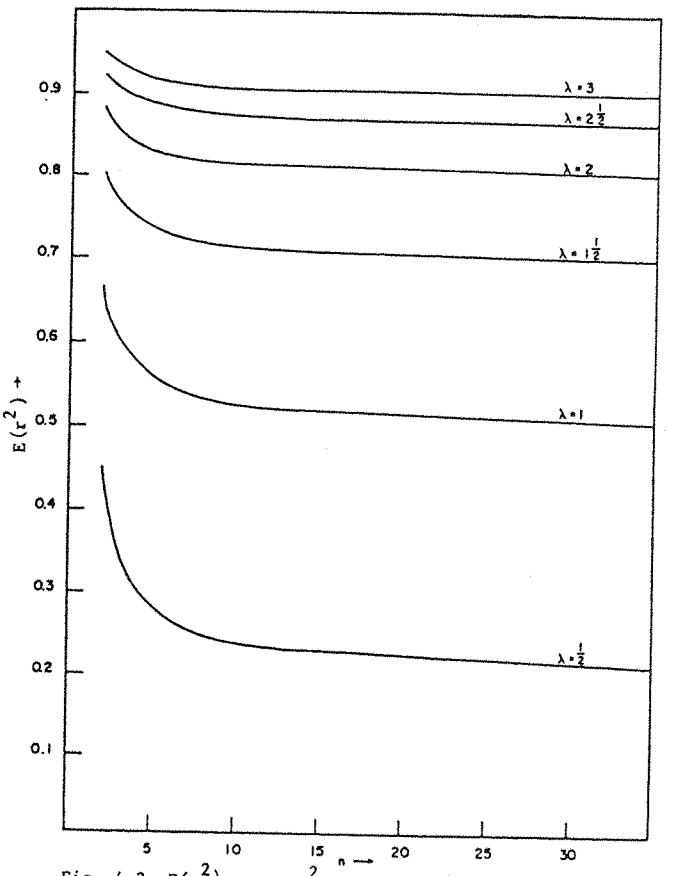

Fig. 4.2 $E(r^2)$ vs. n, $\rho^2 = 0.5$, $g = 2$

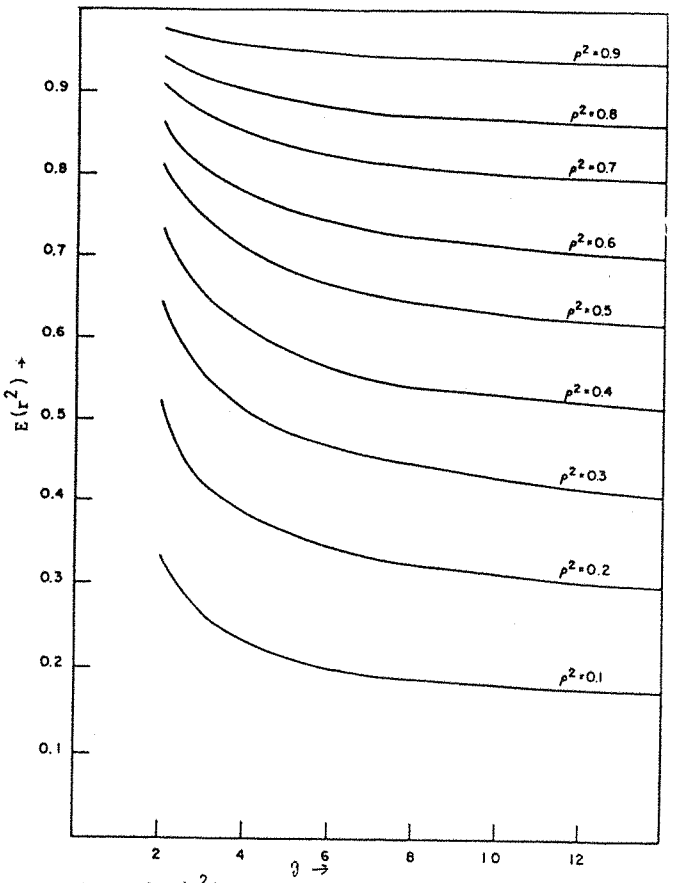

Fig. 4.3 $E(r^2)$ vs. $g, \lambda = 2, N = 30$.

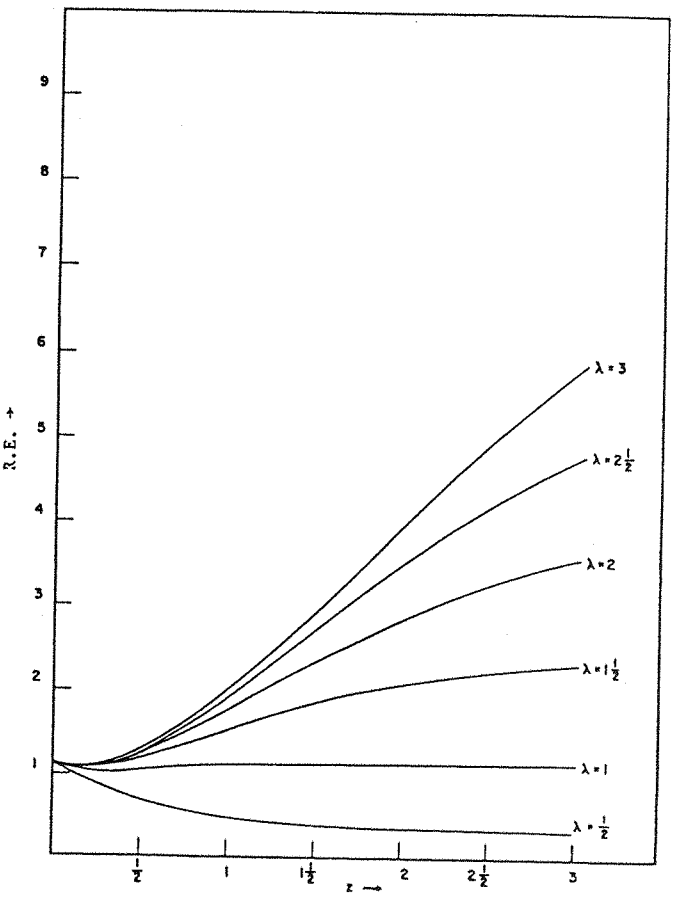

Fig. 4.4 R.E. vs. $z, n = 10, g = 2$

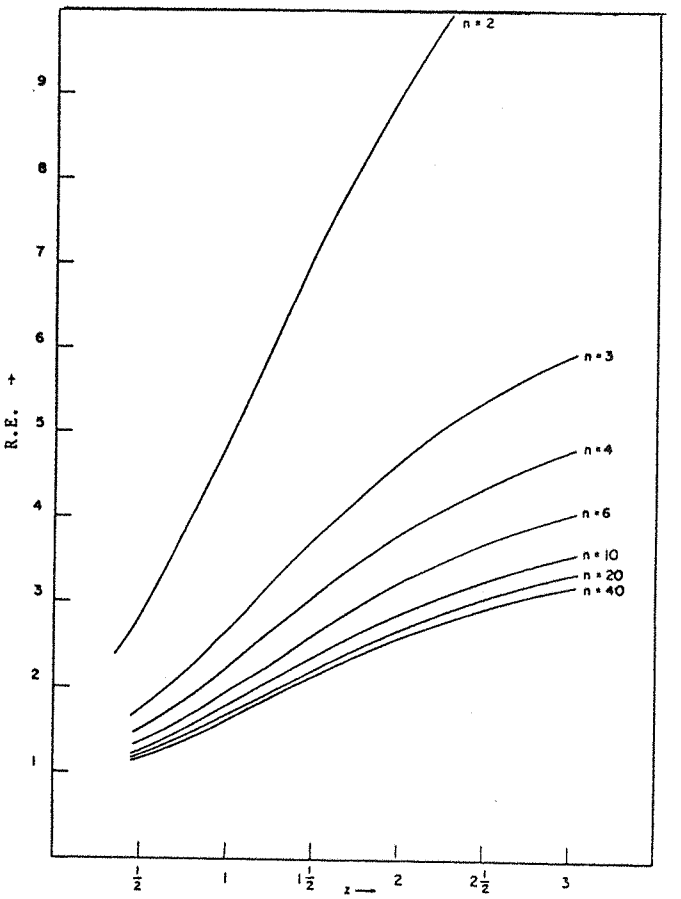

Fig. 4.5 R.E. vs. z, λ = 2, g = 2

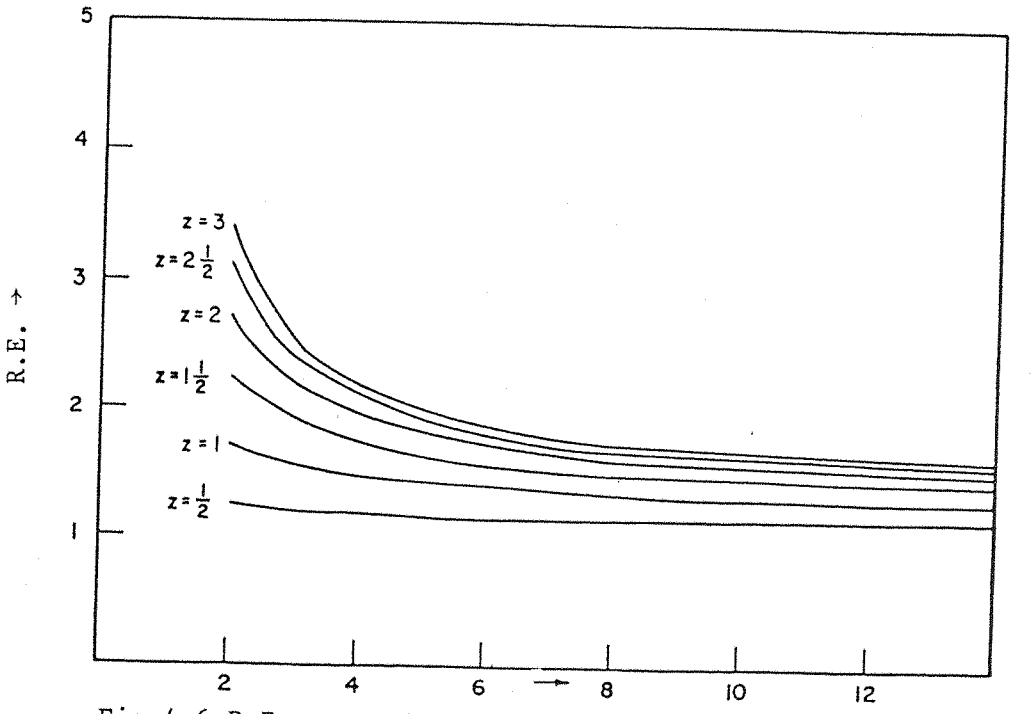

Fig.4.6 R.E. vs. g, λ = 2, N = 30

if $\rho^2 = 0.50$, then $E(r^2) = 0.635$, thus giving an expected overestimate. On the other hand, the relative efficiency for estimating the regression, except close to the mean, is of the order of 150 per cent.

4.5 Effect of Stratifying Variable on the Behaviour of Correlations

Let θ be the population parameter of interest and $\hat{\theta}$ be the estimate of θ computed from the sth sample $(s = 1,2,\ldots,S)$.

Then $\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$.

$$\text{Relative Bias} = \frac{E(\hat{\theta}) - \theta}{\theta} \qquad (4.64)$$

In Frankel's study, the relative biases for the multiple correlation were the highest as compared to the biases for the simple and partial correlations for 6 and 12 strata designs (Table 4.3). However, these biases for the multiple correlation decreased with the increasing sample size at a good rate, i.e. 16, 11 and 5 percent for 6, 12 and 30 strata designs. This suggests that the relative bias is a function of sample size and for large sample sizes the bias may decrease considerably.

The important point to consider is whether the poor performance of the multiple correlation is caused mainly by the correlation between elements due to the complexity of the design or due to some other reason. Kalton (discussion of Kish and Frankel, 1974) expresses the opinion that the multiple correlations in Frankel's study have been derived only from two regression equations and there may be something atypical in these cases. Sampford (discussion of Kish and Frankel, 1974) says that the poor performance of the multiple correlation may be due to the effect of clustering rather than the nature of the estimate. However, Smith, T.M.F. (discussion of Kish and Frankel) points out that if the objective of the study is regression analysis, then the sample multiple correlation coefficient is being used in the second order sense of measuring the

goodness of fit of the model to the data rather than in the first order sense of estimating the population multiple correlation. However, if used to estimate the population correlation coefficient for a sample design involving stratification, then Warren's results show that the sample multiple correlation coefficient may be badly biased.

Kish and Frankel's study is based on a stratified clustered sampling design. The estimates of correlation in their study are computed from two regression equations. It is possible that the high relative biases for the multiple correlations in their study may be due to the strong relationship of the stratification variable (geographical region and population density) with one or more of the variables under study. Kish and Frankel present seemingly anomalous results about partial and multiple correlations. In fact, Kish and Frankel's analysis using deffs confounds clustering and stratification. These authors have mainly concentrated on the effects of clustering and have thus ignored the stratification. Stratification can cause bias and earlier in Section 4.3, we have seen that Koop and Wakimoto's results show that stratification has an effect on variances and covariances. Also from Warren's theoretical results and graphs (sub-sections 4.4.1.1 and 4.4.1.2) we have observed that stratification can cause bias in the estimation of correlations and that the bias can be of substantial magnitude. It, therefore, seems worthwhile to examine the effects of stratification on the behaviour of simple, partial and multiple correlations. We attempt to reproduce Kish and Frankel's results employing correlations modified by taking into consideration stratification effects.

4.5.1 Simulation Study

Consider the regression of Y on $\underline{X}$. Let $X_1, X_2, \ldots, X_K$ $(i = 1,2,\ldots,K)$ be the independent variables and Z the stratification

(design) variable. Suppose that the variable X_i is highly dependent on Z. Let r_{yi} be the sample correlation between the dependent variable Y and the independent variables X_i, r_{yj} the sample correlation between Y and X_j $(j = 1,2,\ldots,k-1)$ and r_{ij} the sample correlation between X_i and X_j. We denote the corresponding population correlation by ρ_{yi}, ρ_{yj} and ρ_{ij} respectively. Further, we suppose that Δ_i is the bias in r_{yi} caused by stratification. These relationships can be expressed in general form as:

$$E(r_{yi}) = \rho_{yi} + \Delta_i \; ; \; E(r_{yj}) = \rho_{yj} \quad \text{for} \quad j \neq i = 1,2,\ldots,K$$

and

$$E(r_{ij}) = \rho_{ij} \quad \text{for} \quad i \neq j = 1,2,\ldots,K \; . \tag{4.65}$$

Using data from Frankel's table 5.1 (see Appendix I) we conducted an exhaustive simulation study and computed fairly extensive tables of relative biases under the above assumptions for different values of Δ's and different combinations of independent and dependent variables. A sample of two cases from these different simulation results is presented as Tables 4.4 and 4.5. It is hoped that these will be sufficient to illustrate the salient points of the effects of bias caused by the compexity of the design on the behaviour of the simple, partial and multiple correlations.

Frankel considered 2 different three variable regression equations. First, the regression of Y (total income of the household head) on X_1 (number of persons in household under 18), X_2 (number of persons in household) and X_3 (sex of household head). The second was the regression of Y' (total income of household) on X_4 (age of household head), X_5 (number of persons in household in labour force) and X_6 (years of school completed by the household head). A study of these variables indicates that possibly X_3 (sex of household head) for

regression equation 1 and X_6 (years of school completed by the household head) for regression equation 2 may be highly dependent on the design variable Z (geographical region and the population density). We have examined this and other possibilities in detail and report here only two cases which may explain the behaviour of the correlation coefficients.

4.5.1.1 Special Case I

We assume that X_3 (sex of household head) for regression equation 1 and X_6 (years of school completed by the household head) for regression equation 2 are highly correlated with the stratification (design) variable Z (geographical regions and population density) such that:

$$E(r_{y3}) = \rho_{y3} + \Delta_3 \ , \ E(r_{y1}) = \rho_{y1}, \ E(r_{y2}) = \rho_{y2} \ , \ E(r_{12}) = \rho_{12} \ ,$$
$$E(r_{13}) = \rho_{13}, \ E(r_{23}) = \rho_{23} \ .$$

$$E(r_{y6}') = \rho_{y6}' + \Delta_6 \ , \ E(r_{y4}) = \rho_{y4} \ , \ E(r_{y5}) = \rho_{y5} \ , \ E(r_{45}) = \rho_{45} \ ,$$
$$E(r_{46}) = \rho_{46} \ , \ E(r_{56}) = \rho_{56} \ . \qquad (4.66)$$

where r_{y3} and r_{y6}' are the estimated correlations between y and X_3 , and y' and X_6 respectively. Δ_3 and Δ_6 are the biases in r_{y3} and r_{y6}' respectively.

From Table 4.4, we observe that for $\Delta = 0.01$ the average absolute relative biases for the multiple correlations are higher as compared to the average absolute relative biases for the partial and simple correlations except for 30 strata design where the average biases for the partial correlations are higher than those of the multiple correlations. Moreover, these average absolute relative biases decrease with the increasing sample size. Thus, we note that for $\Delta = 0.01$ these average absolute relative biases are quite consistent with and very close to Kish and Frankel (1974) results (Table 4.3).

Table 4.4 Average Absolute Relative Biases Under Special Case I for Selected Values of

	$\Delta = 0.01$			$\Delta = 0.21$			$\Delta = 0.50$		
Correlations	6 Strata	12 Strata	30 Strata	6 Strata	12 Strata	30 Strata	6 Strata	12 Strata	30 Strata
Simple	0.0762	0.0603	0.0236	0.5786	0.5502	0.4962	0.5007	0.4819	0.4436
Partial	- 0.1600	0.0963	0.0803	4.9016	4.7147	4.8617	1.2262	1.2008	1.1964
Multiple	0.1854	0.1335	0.0676	0.4947	0.4286	0.3450	0.1744	0.1781	0.1810

Table 4.5 Average Absolute Relative Biases Under Special Case II for Selected Values of

	$\Delta = 0.02$			$\Delta = 0.20$			$\Delta = 0.50$		
Correlations	6 Strata	12 Strata	30 Strata	6 Strata	12 Strata	30 Strata	6 Strata	12 Strata	30 Strata
Simple	0.1046	0.0997	0.0637	1.4999	1.5133	1.4569	1.1459	1.1421	1.1062
Partial	0.1642	0.1022	0.0927	3.5489	3.4093	3.2589	1.1593	1.1339	1.1306
Multiple	0.1983	0.1456	0.0789	0.4543	0.3892	0.3070	0.0230	0.0328	0.1020

We also note that as Δ varies the average absolute relative biases also vary and become larger for simple and partial correlations as compared to the multiple correlations and are no longer consistent with Kish and Frankel's results. Thus, at $\Delta = 0.21$ and $\Delta = 0.50$, we observe that the average absolute relative biases for the simple and partial correlations are much higher as compared to multiple correlations which is contradictory to Kish and Frankel's results.

These results support Warren's argument discussed earlier in this section as pointed out by Smith (1974). Thus our results indicate that the stratification in the design affects the behaviour of the partial and multiple correlations. This also contradicts Kish and Frankel's assumption that the design affects only the variances.

4.5.1.2 Special Case II

We assume that X_3 (sex of the household head) for regression equation 1 and X_6 (years of school completed by the household head) for regression equation 2 are related to the design variable Z (geographical region and population density) such that:

$E(r_{y3}) = \rho_{y3} + \Delta_3$, $E(r_{13}) = \rho_{13} + \Delta_{13}$, $E(r_{23}) = \rho_{23} + \Delta_{23}$, $E(r_{12}) = \rho_{12}$, $E(r_{y1}) = \rho y_1$, $E(r_{y2}) = \rho_{y2}$.

$E(r_{y'6}) = \rho_{y'6} + \Delta_6$, $E(r_{46}) = \rho_{46} + \Delta_{46}$, $E(r_{56}) = \rho_{56} + \Delta_{56}$, $E(r_{45}) = \rho_{45}$

and

$$E(r'_{y4}) = \rho'_{y4} \quad , \quad E(r'_{y5}) = \rho'_{y5} \tag{4.66}$$

where $r_{y.}$, $r'_{y.}$ and $r..$ are the correlation estimates for $\rho_{y.}$, $\rho_{y.}$ and $\rho'_{y.}$. The Δ's are the biases in the estimated correlations.

We note (Table 4.5) that the average absolute relative biases for $\Delta = 0.02$ for the simple, partial and multiple correaltions are quite

consistent with and very close to the average absolute relative biases reported by Kish and Frankel (1974). This indicates that Warren's model reproduces Kish and Frankel's results. For moderate values of bias ($\Delta = 0.20$) the average absolute relative biases for the partial correlations are higher as compared to simple correlations which in turn are higher than those for multiple correlations. For $\Delta = 0.50$, we observe that the average absolute relative biases for the simple correlations are much higher than those for the partial and multiple correlations.

These results indicate that Kish and Frankel's results can hold under specific situations but can be varied by varying Δ's and the assumptions. Thus by suitable choices, which fit the stratification almost exactly, we conclude that the high relative biases in Kish and Frankel are due to the effect of stratification rather than the basic defect in the estimator which contradicts Kish and Frankel conjecture stated earlier in Section 4.2. Thus there is no anomaly.

4.6 Summary and Main Conclusions

Problems concerning the estimation of correlation from complex surveys for finite populations have been discussed. Kish and Frankel argue that the complexity of the design affects only the variances and that the high relative biases for the multiple correlations are due to the basic defects of the estimator rather than the complexity of the design. By concentrating mainly on the intra-cluster correlation, these authors ignored the effects of stratification and have been led into a misinterpretation of results. Koop and Wakimoto's results indicate that stratification does effect variances and covariances. Warren's analysis

indicates that stratification can cause bias in the estimation of the correlation and this bias may be of substantial magnitude. Our analysis of the effects of stratification from a simulation study on the behaviour of simple, partial and multiple correlations indicate that:

(i) The complexity of the design does affect the behaviour of the simple, partial and multiple correlation.

(ii) The relative biases for the multiple correlations, in general, are much smaller than those for the simple and partial correlations, when the design variable is strongly related to one of the X variables under study.

(iii) Different stratifications can produce different results.

(iv) Kish and Frankel's results can hold under specific situations and for specific populations but are not true in general. They can be varied under varying conditions.

We conclude that the high relative biases for the multiple correlations in Kish and Frankel's study are due to the effect of stratification rather than in defects of the estimator. Thus, we disapprove of Kish and Frankel's conjecture regarding high relative biases in multiple correlations. We also remark that different surveys with different variables and different stratifications may produce quite different results, i.e., the relative biases may be higher for simple correlations or partial correlations than for multiple correlations.

CHAPTER 5

THE JACK-KNIFE METHOD: A REVIEW

5.1 Introduction

Over the years the jack-knife method, introduced briefly in Chapter 3 (Section 3.2.2.9) has found a wide applicability in a variety of fields. It is the subject of this chapter to discuss the ideas which motivated the universal application of the method. Our intention is not to trace the history of the method but to highlight the developments of significant relevance on the subject for a specialist user of the technique. It seems to us that a review of main developments concerning the jack-knife method will contribute significantly to the understanding of the technique as a whole and its various applications in survey sampling.

Quenouille (1949) introduced the jack-knife technique for reducing the bias of a serial correlation estimator by dividing the total sample into two half samples. He noted that the bias was inversely proportional to the sample size and the bias to order $1/n$ could be removed by this method. Later in 1956, he generalised this idea by splitting the sample into g groups of size m each and examined its general properties. Since then the general problem of bias removal has been the subject of extensive study. This aspect of the jack-knife method is surveyed in Section 5.2.

One of the earliest applications of the jack-knife method, both as a means of bias reduction and variance estimation, is in the case of

ratio estimators which are frequently employed to improve the efficiency of ratio estimation in sample surveys. These estimators are known to be biased and this can be of importance if a fairly small sample is only available or if the sample is heavily stratified. Perhaps the most important is the ability of the jack-knife to produce an estimate of variance where alternative techniques appear to be cumbersome and time consuming. This important feature of the jack-knife for complex surveys have been explored among others by Brillinger (1966), McCarthy (1966) and Kish and Frankel (1974) for estimating the variances of complex statistics from complex samples. The jack-knife technique has also proved to be a useful device for variance components estimation for complex sample statistics. The issues regarding the performance of the jack-knife method in ratio estimation, comparison of variances and applications to complex survey situations are discussed in Section 5.3.

The jack-knife method enjoys the position of being a widely used technique in aiding the researcher to obtain approximate test or confidence intervals for parameters of interest. There are, however, counter examples as well, see, for example, Section 3.2.2.9. The motivation for using the jack-knife for interval estimation is to overcome the problems of non-normality and hence provide robust inference procedures. Developments along this line are summarised in Section 5.4.

Finally, in Section 5.5 we give a summary and conclusions.

5.2 Bias Reduction Properties

Suppose that we have a random sample $Y_1, Y_2, \ldots, Y_n$ and we want to estimate the value of an unknown parameter θ. Let $\hat{\theta}(Y_1, Y_2, \ldots, Y_n)$

be an estimator of θ such that $E(\hat{\theta}) \neq \theta$ and $\hat{\theta}_{-i}(i=1,2,\ldots,n)$ be the estimate obtained by deleting the ith observation. Then

$$\hat{\theta}_{-i} = \hat{\theta}(Y_1, Y_2, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n) \tag{5.1}$$

The ith pseudovalue is given by

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i} \tag{5.2}$$

The jack-knife estimate of θ is the average of these pseudovalues and is given by

$$\tilde{\theta} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\theta}_i \quad (i=1,2,\ldots,n) \tag{5.3}$$

This estimate will generally have smaller bias than the original estimate and sometimes smaller variance as well. If n is large, the jack-knife can be modified by splitting the sample into g groups of size m each so that $n = mg$ and deleting one of the groups successively, the jack-knife estimate of θ is given by

$$\tilde{\theta} = \frac{1}{g} \sum_{i=1}^{n} \tilde{\theta}_i \quad \text{where} \quad \tilde{\theta}_i = g\hat{\theta} - (g-1)\hat{\theta}_{-i} \tag{5.4}$$

The jack-knife method was introduced by Quenouille (1949, 1956) as a means of eliminating the bias of order n^{-1} from an estimator. Suppose that the expected value of the estimator $\hat{\theta}$ based on a sample of size n is represented as a series expansion such that

$$E(\hat{\theta}) = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \ldots = \theta + O(n^{-1}) \tag{5.5}$$

Then the jack-knifed estimator $\tilde{\theta}$ possesses the interesting property that it eliminates the leading bias term, i.e.,

$$E(\tilde{\theta}) = \theta + \frac{a_2}{n^2} + \ldots = \theta + O(n^{-2}) \tag{5.6}$$

A re-application of the method could remove bias of order n^{-2} which remains after the first application. Mantel (1967) also considered the bias reduction problem and gave an alternative formula for the re-application which eliminated the n^{-1} and n^{-2} terms. The drawback with the estimators given by Quenouille and Mantel is that if the entering estimator may be written exactly as

$$E(\hat{\theta}-\theta) = \frac{a_1}{n} + \frac{a_2}{n^2} ,$$

then their estimators are not free of bias, that is they contain an n^{-3} term.

A generalisation of the jack-knife procedure for the elimination of bias was suggested by Schucany, Gray and Owen (1971). Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of θ with expectations of the form

$$\begin{aligned} E(\hat{\theta}_1) &= \theta + f_1(n)B(\theta) \\ E(\hat{\theta}_2) &= \theta + f_2(n)B(\theta) \end{aligned} \qquad (5.7)$$

Then, the estimator

$$\tilde{\theta}^* = \det \begin{pmatrix} \hat{\theta}_1 & \hat{\theta}_2 \\ f_1(n) & f_2(n) \end{pmatrix} \Big/ \det \begin{pmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{pmatrix} \qquad (5.8)$$

is exactly unbiased estimator for θ . The estimator $\tilde{\theta}^*$ is called the generalised jack-knife. The standard jack-knife is a special case of (5.8) with $\hat{\theta}_1 = \theta$, $\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{-i}$, $f_1(n) = \frac{1}{n}$ and $f_2(n) = \frac{1}{(n-1)}$. The definition (5.8) can be extended to include three or more estimators and the second or higher order bias terms in expansion (5.5) can be exactly eliminated by the generalised jack-knife. Further details on these generalisations appear in Schucany et al., (1971), Adams et al.,

(1971) and the monograph by Gray and Schucany (1972). Properties of the generalised jack-knife have also been discussed by Sharot (1976 a,b).

For the estimation problem where the quantity to be estimated is $f(\theta)$, f is a known function and θ being the unknown parameter, Miller (1964) established some results for $\theta = \mu$ and $\hat{\theta} = \bar{x}$. Arvesen (1969) extended these results to U statistic and Gray et al., (1972) established the validity of these results for the generalised jack-knife.

The jack-knife method could also be improved if the observations are grouped so as to follow some special pattern or structure in the data for calculating the jack-knife estimator. Such a grouping will eliminate the first order bias correction term, whereas in the absence of such grouping only reduction of these terms could be done, see for instance, Farewell (1978). Krewski(1978b), discusses that the jack-knife is also a useful method for bias reduction for small samples.

5.3 Application of Jack-knife in Sample Surveys

5.3.1 Ratio Estimation (SRS)

In sample surveys we frequently come across the situations where the use of auxiliary information may improve the precision of the estimates. Consider a population of N units. Let x and y be two variables of interest and X_i and Y_i denote the values of these variables for the ith unit in the population. Suppose a simple random sample of n units is drawn from this population and let x_i, y_i denote the values of the variables for the ith unit in the sample.

We know that the sample mean $\bar{y} = \sum_{i=1}^{n} y_i/n$, provides an unbiased estimate of the population mean $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$. Now if the variable x is positively correlated with the variable y and the population mean $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$ of the variable x is known, then the estimator $\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}} \bar{X}$ may also be used as an estimator of $\bar{Y}$ due to the reason that the ratio of the variable y to the variable x may have smaller variance than the variable y alone. Sometimes, it is desired to estimate the population ratio rather than the mean and for this purpose we use the sample ratio $\hat{R} = \frac{\bar{y}}{\bar{x}}$; where $\bar{y}$ and $\bar{x}$ are the sample means of the characteristic y and x respectively and $\bar{x} \neq 0$ for every n.

These ratio estimators are consistent but biased for many of the selection procedures commonly used in samples surveys. This bias is, however, negligible for large samples. For moderate size samples, the distribution of the ratio estimate is positively skewed for most types of the populations for which the ratio estimate is generally used. For example, the bias may be considerable in surveys with many strata or small or moderate samples within strata if separate ratio estimates are used for each stratum. The bias, however, may be negligible if a combined ratio estimate is used. Moreover, if it is required that proper confidence statements be made, it is often necessary that the bias of an estimator may be negligibly small. The other problem with these estimators is that the variance estimators $\hat{V}(\hat{\bar{Y}}_R)$ and $\hat{V}(\hat{R})$ are only large sample approximations and their derivations depend upon the validity of the curtailed Taylor Series expansion.

To overcome these difficulties, considerable importance has been given to investigate the properties of these estimators during the last three decades. Consequently, several modifications both under the

design based and the model based approaches have been suggested, see for example, Quenouille (1956), Robson (1957), Goodman and Hartley (1958), Hartley and Ross (1954), Mickey (1959), Beale (1962), Murthy (1962), Tin (1965), Royall (1970), Royall and Eberhardt (1975) and Royall and Cumberland (1978b),etc.

One of the earliest applications of the jack-knife method was in the case of ratio estimators which are frequently employed to improve the efficiency of estimation in survey samples. We have discussed the bias and variance estimation properties of the ratio estimator under the model based approach in Chapter 3. In this section we discuss the bias and the variance estimation for the jack-knifed ratio estimator under the design based approach. We will also refer to models but only as a method for evaluation. Before discussing the jack-knifed ratio estimator we would very briefly give the bias and variance of the classical ratio estimator so as to facilitate comparison when the need arises.

5.3.1.1 Approximation to the Bias and the Variance of Ratio Estimate

Consider a simple random sample of n units (without replacement) from a population of N units. For non-negative integers r and s, we have from David and Sukhame (1974),

$$E(\bar{x}-\bar{X})^r(\bar{y}-\bar{Y})^s = \begin{cases} O(n^{-\frac{1}{2}(r+s)}) & \text{if } r+s \text{ is even} \\ O(n^{-\frac{1}{2}(r+s+1)}) & \text{if } r+s \text{ is odd} \end{cases}$$

and

$$(\bar{x}-\bar{X})^r(\bar{y}-\bar{Y})^s = O_p(n^{-\frac{1}{2}(r+s)}) \qquad (5.9)$$

Let $\bar{a} = \frac{\bar{x}-\bar{X}}{\bar{X}}$, $\bar{b} = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$, then bias of $\hat{R} = \frac{\bar{y}}{\bar{x}}$ is given by

$$B(\hat{R}) = E(\hat{R}-R) = R\left[E\left(\frac{\bar{y}}{\bar{x}} - \frac{\bar{Y}}{\bar{X}}\right)\right]$$

$$= R\left[E(\bar{b}-\bar{a})\left(\frac{\bar{x}}{\bar{X}}\right)^{-1}\right] \qquad (5.10)$$

Now consider the approximation to $B(\hat{R})$ given by

$$B_k(\hat{R}) = R[E(\bar{b}-\bar{a})(1-\bar{a}+\bar{a}^2-\bar{a}^3+\ldots.+(\bar{a})^{2k-2}\ldots.(\bar{a})^{2k-1}] \qquad (5.11)$$

where k is some positive integer. From (5.9) we note that $B_k(\hat{R})$ consists of terms in $\bar{a}$ and $\bar{b}$ up to and including $0(n^{-k})$. Now from (5.10) and (5.11) we have

$$B(\hat{R}) - B_k(\hat{R}) = R\left[E(\bar{b}-\bar{a})\left(\frac{\bar{X}}{\bar{x}}\right)\right] - R\left[E(\bar{b}-\bar{a})(1-\bar{a}+\ldots.+(\bar{a})^{2k-2}\ldots.(\bar{a})^{2k-1}\right] \qquad (5.12)$$

Thus, if x is positive for every unit in the population and if x_o is positive lower bound, then

$$|B(\hat{R})-B_k(\hat{R})| \leq \left(\frac{\bar{Y}}{x_o} E\left\{(\bar{a})^{2k}(\bar{b}) - (\bar{a})^{2k+1}\right\}\right) = 0(\bar{n}^{(k+1)}) \qquad (5.13)$$

David and Sukhatme (1974) argue that by a proper choice of n and k , $B(\hat{R}) - B_k(\hat{R})$ can be made very small and hence $B_k(\hat{R})$ provides a useful approximation to $B(\hat{R})$. Now it could be easily seen that

$$B_1(\hat{R}) = R\cdot[E(\bar{b}-\bar{a})(1-\bar{a})] = R\cdot\frac{N-n}{Nn}(C_{20}-C_{11}) = 0(n^{-1}) \qquad (5.14)$$

and

$$B_2(\hat{R}) = R\,[E(\bar{b}-\bar{a})(1-\bar{a}+\bar{a}^2-\bar{a}^3)]$$

$$= R\left[\frac{N-n}{Nn}(C_{20}-C_{11}) + \frac{(N-n)(N-2n)}{(N-1)(N-2)n^2}(C_{21}-C_{30}) + \frac{(N-n)(N^2+N-6Nn+6n^2)}{(N-1)(N-2)(N-3)n^3}\right.$$

$$\left.(C_{40}-C_{31}) + \frac{3N(N-n)(N-n-1)(n-1)}{(N-1)(N-2)(N-3)n^3} C_{20}(C_{20}-C_{11})\right]$$

$$= 0(n^{-1})$$

where

$$C_{rs} = \frac{\sum_{i=1}^{N} (x_i-\bar{X})^r (y_i-\bar{Y})^s}{(\bar{X})^r (\bar{Y})^s} \tag{5.15}$$

The mean square error is

$$MSE(\hat{R}) = E(\hat{R}-R)^2 = R^2\left[E(\bar{b}-\bar{a})^2\left(\frac{\bar{X}}{\bar{x}}\right)^2\right] \tag{5.16}$$

and

$$MSE_k(\hat{R}) = R^2[E(\bar{b}-\bar{a})^2(1-2\bar{a}+3\bar{a}^2)....(-2k-2)(\bar{a})^{2k-3}+(2k-1)(\bar{a})^{2k-2}] \tag{5.17}$$

Arguments as above lead to the conclusion that $MSE_k(\hat{R})$ provides a useful approximation to $MSE(\hat{R})$. Thus, we have

$$MSE_1(\hat{R}) = R^2[E(\bar{b}-\bar{a})^2] = R^2\left[\frac{N-n}{Nn}(C_{02}-2C_{11}+C_{20})\right] \tag{5.18}$$

we note from this expression that $MSE(\hat{R})$ decreases as n increases. The variance could be obtained as

$$V(\hat{R}) = E(\hat{R})^2 - [E(\hat{R})]^2 = R^2\frac{N-n}{N}[C_{02}+C_{20}-2C_{11}] \tag{5.19}$$

A second approximation to $MSE(\hat{R})$ can be obtained by including terms up to $O(n^{-3})$.

$$\begin{aligned} MSE_2(\hat{R}) &= R^2[E(\bar{b}-\bar{a})^2(1-2\bar{a}+3\bar{a}^2)] \\ &= R^2\left[\frac{N-n}{(N-1)n}(C_{20}-2C_{11}+C_{02}) + \frac{2(N-n)(N-2n)}{(N-1)(N-2)n^2}(2C_{21}-C_{12}-C_{30})\right. \\ &\quad + 3\frac{(N-n)(N^2+N-6Nn+6n^2)}{(N-1)(N-2)(N-3)n^3}(C_{40}-2C_{31}+C_{22}) + \frac{3N(N-n)(N-n-1)(n-1)}{(N-1)(N-2)(N-3)n^3} \\ &\quad \left.(3C_{20}^2-6C_{20}C_{11}+C_{20}C_{02}+2C_{11}^2)\right] . \end{aligned} \tag{5.20}$$

For large N , the expression (5.20) reduces to

$$MSE_2(\hat{R}) = R^2\left[\frac{1}{n}(C_{20}-2C_{11}+C_{02}) + \frac{2}{n^2}(2C_{21}-C_{12}-C_{30}) + \frac{3}{n^3}(C_{40}-2C_{31}+C_{22}) + \frac{3(n-1)}{n^3}(3C_{20}^2-6C_{20}C_{11}+C_{20}C_{02}+2C_{11}^2)\right]. \quad (5.21)$$

Thus we note that the accuracy of $Bias(\hat{R})$ and $MSE(\hat{R})$ depends in fact on the sample and the population under study.

Several modifications have been suggested to develop unbiased ratio type estimators. One approach to obtain the unbiased estimators is to modify the sampling procedure in order to make the available estimator unbiased, see for instance, Lahiri (1951), Midzuno (1950) and Raj (1968). The second approach is to modify the usual biased ratio estimator so that an unbiased ratio type estimator is obtained. This approach has lead to estimators like $\hat{R}_{HR}$ (Hartley and Ross, 1954), $\hat{R}_M$ (Mickey 1959), $\hat{R}_B$ (Beale, 1962), $\hat{R}_T$ (Tin, 1965) and $\hat{R}_P$ (Pascual, 1961) as alternatives to the classical ratio estimator $\hat{R}$. These estimators have, however, found little applications in large scale sample surveys due to the complications involved either with the ratio estimators or the estimators of the variances of these estimates. Besides, the conditions under which the unbiased ratio estimators are expected to be more efficient than the usual ratio estimator are sometimes restrictive. An alternative ratio type estimator which removes the bias to $O(n^{-1})$ was proposed by Quenouille (1956). This estimator, known as the jack-knifed ratio estimator, has been employed extensively in a wide variety of situations with a fair degree of success. We discuss the jack-knifed ratio estimator in some detail in the next section.

5.3.1.2 The Jack -knifed Ratio Estimator

Consider a simple random sample of size n. Let (y_i,x_i) be the values observed for the ith selected unit in the sample. Let the n values in the sample be divided at random into g groups, each

of size m, so that $n = mg$. We remove the ith group and use the remaining $(g-1)$ groups to compute the ratio estimate $\hat{R}_{(j)}, j=1,2,\ldots,g$. Let $\bar{y}_{(j)}$ and $\bar{x}_{(j)}$ denote the means obtained after omitting the jth group. Then $\hat{R}_{(j)} = \frac{\bar{y}_{(j)}}{\bar{x}_{(j)}}$ where $\bar{y}_{(j)} = \frac{n\bar{y}-m\bar{y}_j}{n-m}$, $\bar{x}_{(j)} = \frac{n\bar{x}-m\bar{x}_j}{n-m}$; $\bar{y}_j$ and $\bar{x}_j$ are the means for the jth group and $\bar{y}$ and $\bar{x}$ are the sample means. The pseudovalues are defined as

$$\tilde{R}_j = g\hat{R} - (g-1)\hat{R}_{(j)} \tag{5.22}$$

and the jack-knifed ratio estimator $\tilde{R}_J$ is the average of these pseudo-values

$$\tilde{R}_J = \frac{1}{g}\sum_{j=1}^{g}\tilde{R}_j = g\hat{R} - \frac{(g-1)}{g}\sum_{j=1}^{g}\hat{R}_{(j)} \tag{5.23}$$

Proceeding as in Subsection 5.3.1.1 and using (5.9), the expressions for the bias and mean square error for $\tilde{R}_J$ for $g = n$ to $O(n^{-3})$ can be derived. Let

$$\bar{a}_{(j)} = \frac{\bar{x}_{(j)}-\bar{X}}{\bar{X}} \quad \text{and} \quad \bar{b}_{(j)} = \frac{\bar{y}_{(j)}-\bar{Y}}{\bar{Y}}$$

Then

$$\hat{R}_{(j)} = R\left(\frac{1+\bar{b}_{(j)}}{1+\bar{a}_{(j)}}\right) = R[1+\bar{b}_{(j)}-\bar{a}_{(j)}-\bar{a}_{(j)}\bar{b}_{(j)}+\bar{a}^2_{(j)}+\bar{a}^2_{(j)}\bar{b}_{(j)}-\bar{a}^3_{(j)}-\bar{a}^3_{(j)}\bar{b}_{(j)}+\bar{a}^4_{(j)}]$$

$$E(\hat{R}_{(j)}) = R\left[1+\frac{g}{n(g-1)}(C_{20}-C_{11}) + \frac{g^2}{n^2(g-1)^2}(C_{21}-C_{30}) + \frac{3g^2}{n^2(g-1)^2}(C^2_{20}-C_{20}C_{11})\right.$$
$$\left. - \frac{3g^3}{n^3(g-1)^3}(C^2_{20}-C_{20}C_{11}) + \frac{g^3}{(g-1)^3n^3}(C_{40}-C_{31})\right]. \tag{5.24}$$

Now

$$E(\tilde{R}_J) = gE(\hat{R}) - \frac{(g-1)}{g} \sum_{j=1}^{g} E(\hat{R}_{(j)})$$

$$= R\left[1 - \frac{g}{n^2(g-1)}(C_{21}-C_{30}) - \frac{3g}{(g-1)n^2}(C_{20}^2-C_{20}C_{11}) - \frac{g(2g-1)}{n^3(g-1)^2}\right. \quad (5.25)$$

$$\left.(C_{40}-C_{31}) + \frac{g(2g-1)}{(g-1)^2n^3}(C_{20}^2-C_{20}C_{11})\right] . \quad (5.25)$$

Hence

$$B(\tilde{R}_J) = O(n^{-2}) . \quad (5.26)$$

Thus we note from (5.26) that the bias of $\tilde{R}_J$ is at the most of $O(n^{-2})$ whereas from (5.15) we note that $B(\hat{R})$ is of $O(n^{-1})$.

The mean square error for the jack-knifed ratio estimator $\tilde{R}_J$ could be worked in a similar way.

$$MSE(\tilde{R}_J) = E(\tilde{R}_J-R)^2$$

$$= g^2E(\hat{R}-R)^2 - 2g \sum_{j=1}^{g} E(\hat{R}-R)(\hat{R}_{(j)}-R) + \frac{(g-1)^2}{g^2} \sum_{j=1}^{g} E(\hat{R}_{(j)}-R)^2$$

$$+ \frac{(g-1)^2}{g^2} \sum_{i\neq j} \sum E(\hat{R}_{(j)}-R)(\hat{R}_{(k)}-R)$$

$$= R^2\left[\frac{1}{n}(C_{20}+C_{02}-2C_{11}) + \frac{g}{(g-1)}\frac{1}{n^2}\left\{(C_{11}-C_{20})^2 + C_{20}(C_{20}+C_{02}-2C_{11})\right\}\right.$$

$$\left. + \frac{2g}{(g-1)}\frac{1}{n^3}\left\{(C_{20}C_{02}+2C_{11}^2-6C_{11}C_{20}+3C_{20}^2) - (C_{22}-2C_{31}+C_{40})\right\}\right]. \quad (5.27)$$

If only terms of $O(n^{-2})$ are retained then (5.27) reduces to

$$MSE(\tilde{R}_J) = R^2\left[\frac{1}{n}(C_{20}+C_{02}-2C_{11}) + \frac{g}{g-1}\frac{1}{n^2}\left\{(C_{11}-C_{20})^2 + C_{20}(C_{20}+C_{02}-2C_{11})\right\}\right] \quad (5.28)$$

From (5.28) it is apparent that the mean square error for $\tilde{R}_J$ decreases as g increases so that $g = n$ is the optimum choice. Also for large n , the variance and MSE of $\tilde{R}_J$ to order n^{-1} are equal

and given by (5.28). This implies that the variance of $\tilde{R}_J$ like the mean square error is also a decreasing function of g and has a minimum at $g = n$.

5.3.1.3 Model Based Evaluation of the Jack-knifed Ratio Estimator

We now consider the behaviour of the jack-knifed ratio estimator under some specified infinite population models. Durbin (1959) was the first to consider the ratio estimator $\hat{R} = y/x$ and the jack-knifed ratio estimator $\tilde{R}_{J_2} = 2\hat{R} - \frac{1}{2}(\hat{R}_1+\hat{R}_2)$ for $g = 2$, where $\hat{R}_1 = y_1/x_1$, $\hat{R}_2 = y_2/x_2$, $y = (y_1+y_2)/2$ and $x = \frac{1}{2}(x_1+x_2)$, under two infinite populations models. The models considered were,

(1) the regression of y on x is linear and x is normally distributed.

(2) the regression of y on x is linear and x has a gamma distribution.

These models form the basis of most of the work done to investigate the bias and approximation to mean square error of the various ratio-type estimators during the last two decades.

Durbin derived approximations to the moments of $O(n^{-3})$ and showed that the jack-knife, applied with two groups, not only reduced the bias of the estimator but also decreased variance under the model with x normally distributed. In the case of the model with gamma distribution the exact results indicate that with coefficient of variation less than 0.25, the jack-knifed ratio estimator reduces the bias, although the variance is increased slightly, but the mean square error is reduced. Clearly the reduction of bias and variability is a strong recommendation for the technique. Tin (1965) also examined this and other related problems for the jack-knifed ratio estimator with $g = 2$. Rao (1965) showed that $g = n$ is the optimum choice for the normal auxiliary distribution whereas Rao and Webster (1966) through a combination of

theoretical and numerical study demonstrated that this also holds true for the gamma distribution.

In this subsection we briefly outline the derivations for the bias and variance of the jack-knifed ratio estimator under Durbin's models for g = n and also under a more general model discussed by Rao and Kuzik (1974). We will also refer to other competing estimators when the need arises but will not go into the derivational details for these estimators. Following Rao (1965) and Rao and Webster (1966) we derive the expressions for the bias and variance of the jack-knifed ratio estimator.

5.3.1.3a Bias and Variance of the Jack-knifed Ratio Estimator (g = n) Under Model I

Let $(y_i,x_i), i = 1,2,\ldots,n$ be a random sample of n observations on the characters y and x. Let x_i be normally distributed and we choose the units of measurements such that $E(x_i) = 1$. Let $x_i = 1-a_i$ and $Var(\bar{x}) = h$ of $O(n^{-1})$ where $\bar{x}$ is the sample mean of x_i's. Then for sufficiently large n,

$$E(\bar{x})^{-1} = E(1+\bar{a}+\bar{a}^2+\ldots) = 1+h+3h^2+15h^3+0(n^{-4}), \tag{5.29}$$

and

$$E(\bar{x})^{-2} = E(1+2\bar{a}+3\bar{a}^2+\ldots) = 1+3h+15h^2+105h^3+0(n^{-4}) . \tag{5.30}$$

Consider $\hat{R} = \frac{\bar{y}}{\bar{x}}$ as an estimator of $\rho = \frac{E(y_i)}{E(x_i)}$ based on n observations in the sample. Suppose that the regression of y on x is linear of the form

$$y_i = \alpha+\beta x_i + u_i . \tag{5.31}$$

where $E(u_i/x) = 0$, $Var(u_i/x) = n\delta$ and δ is a constant of $O(n^{-1})$. Thus $Var(\bar{u}/x) = \delta$, where $\bar{u}$ is the sample mean of u_i's, and $\rho = \alpha+\beta$.

Suppose that the sample is split at random into g groups of size m each so that $n = mg$. Recall that the grouped jack-knife can be written as

$$\tilde{R}_J = g\hat{R} - \frac{g-1}{g}\sum_{j=1}^{g}\hat{R}_{(j)}$$

$$= \beta + \alpha\left(\frac{g}{\bar{x}} - \frac{g-1}{g}\sum_{j=1}^{g}\frac{1}{\bar{x}'_j}\right) + g\,\frac{\bar{u}}{\bar{x}} - \frac{g-1}{g}\sum_{j=1}^{g}\frac{\bar{u}'_j}{\bar{x}'_j} \qquad (5.32)$$

where $\hat{R}_{(j)} = \frac{\bar{y}'_j}{\bar{x}'_j}$, $\bar{y}'_j = \frac{n\bar{y}-m\bar{y}_j}{n-m}$, $\bar{x}'_j = \frac{n\bar{x}-m\bar{x}_j}{n-m}$, $\bar{u}'_j = \frac{n\bar{u}-m\bar{u}_j}{n-m}$ and $\bar{y}_j$, $\bar{x}_j$ and $\bar{u}_j$ are the sample means for the jth group. Now, under the model (5.31) we have

$$E(\tilde{R}_J-\beta) = \alpha E\left[\frac{g}{\bar{x}} - \frac{g-1}{g}\sum_{j=1}^{g}\frac{1}{\bar{x}'_j}\right] + E\left[g\,\frac{\bar{u}}{\bar{x}} - \frac{g-1}{g}\sum_{j=1}^{g}\frac{\bar{u}'_j}{\bar{x}'_j}\right]$$

After simplifications, we get

$$E(\tilde{R}_J-\beta) = \alpha\left[1 - \frac{3g}{g-1}h^2 - 15g\,\frac{(2g-1)}{(g-1)^2}h^3\right], \qquad (5.33)$$

and

$$\text{Bias}\,(\tilde{R}_J) = \frac{3gh^2}{(g-1)} + 15\,\frac{g(2g-1)}{(g-1)^2}h^3 . \qquad (5.34)$$

It is apparent from (5.34) that h^2 and h^3 are decreasing functions of $g(\geq 2)$ which implies that the bias will be minimum for $g = n$.

The variance of $\tilde{R}_J$ is given by

$$\text{Var}(\tilde{R}_J) = E(\tilde{R}_J-\beta)^2 - [E(\tilde{R}_J-\beta)]^2$$

$$= \alpha^2\left[h + \frac{2g}{g-1}h^2 - 6\,\frac{g(4g^2-14g+11)}{(g-1)^3}\right] + \delta\left[1 + \frac{g}{g-1}h - \frac{2g(2g^2-9g+8)}{(g-1)^3}\right.$$

$$\left. - \frac{3g(28g^4-149g^3+260g^2-173g+32)}{(g-1)^5}h^3\right] \qquad (5.35)$$

For $g = 2$, the expression (5.35) reduces to $V(\tilde{R}_{J2})$ given by Durbin (1959) i.e.,

$$\text{Var}(\tilde{R}_{J2}) = \alpha^2(h+4h^2+12h^3)+\delta(1+2h+8h^2+108h^3) \qquad (5.36)$$

It may be seen from (5.35) that the coefficients of h, h^2 and h^3 are decreasing functions of $g(\geq 2)$ and hence the smallest value of $\text{Var}(\tilde{R}_J)$ will occur at $g = n$. Also we have seen that the bias is minimum at $g = n$. This suggests that $g = n$ is the optimum choice. Thus the estimator $\tilde{R}_{J_n} = n\hat{R} - \frac{n-1}{n}\sum_{j=1}^{n}\hat{R}_{(j)}$, where $\hat{R}_{(1)}, \hat{R}_{(2)},\ldots,\hat{R}_{(n)}$ denote the estimates obtained by deleting the first, second and the nth observation respectively, may be preferable.

5.3.1.3b Bias and Variance of the Jack-knifed Ratio Estimate (for $g = n$) Under Model II

Let us now consider the case where the regression of y on x has the same form as has been stipulated in Subsection 5.3.1.3a but where x has now gamma distribution. This model can be described as

$$y_i = \alpha + \beta x_i + u_i \quad (i = 1,2,\ldots,n) \qquad (5.37)$$

Where $E(u_i/x_i) = 0$, $E(u_i^2/x_i) = n\delta$, a constant of order n^{-1} and the variate x_i/n has a gamma distribution with parameter $nh = m$. Under this model, we have

$$E\left(\frac{1}{\bar{x}}\right) = \frac{1}{m-1} \quad \text{and} \quad E\left(\frac{1}{\bar{x}^2}\right) = \frac{1}{(m-1)(m-2)} \qquad (5.38)$$

Suppose that the sample of n observations is now split into g groups, each of size q so that $n = qg$. Following from subsection 5.3.1.3a, we have $\hat{R}_{(j)} = \frac{\bar{y}'_j}{\bar{x}'_j}$ where in this case $\bar{y}'_j = \frac{n\bar{y}-q\bar{y}_j}{n-q}$, $\bar{x}'_j = \frac{n\bar{x}-q\bar{x}_j}{n-q}$ and $\bar{y}_j$, $\bar{x}_j$ are defined similarly as in (5.32).

Now proceeding in the same way as in subsection 5.3.1.3a, the bias and variance of $\tilde{R}_J$ defined in (5.32) could be obtained. Since $\frac{(g-1)}{g}\bar{x}'_j$ has the gamme distribution with parameter $(n-q)h = \frac{m(g-1)}{g}$, therefore in this case we have

$$E(\tilde{R}_J-\beta) = \alpha E\left[\left\{\frac{g}{\bar{x}} - \frac{g-1}{g}\sum_{j=1}^{g}\frac{1}{\bar{x}'_j}\right\} + \left\{g\,\frac{\bar{u}}{\bar{x}}\,\frac{g-1}{g}\sum_{j=1}^{g}\frac{\bar{u}'_j}{\bar{x}'_j}\right\}\right]$$

$$= \frac{\alpha}{m-1}\left\{\frac{(g-1)(m-1)-g}{(g-1)m-g}\right\} \quad , \tag{5.39}$$

and the bias of $\tilde{R}_J$ is

$$\text{Bias}(\tilde{R}_J) = E(\tilde{R}_J) - \beta - \frac{\alpha}{m} = -\frac{\alpha}{m(m-1)} \cdot \frac{g}{\{m(g-1)-g\}} \tag{5.40}$$

This suggests that the smallest value of the absolute bias will occur at $g = n$, since $|\text{Bias}(\tilde{R}_J)|$ is a monotonically decreasing function of g.

The variance of $\tilde{R}_J$ is now given by

$$V(\tilde{R}_J) = E(\tilde{R}_J-\beta)^2 - [E(\tilde{R}_J-\beta)]^2$$

$$= \alpha^2\left[\frac{g^2}{(m-1)(m-2)} - \frac{2g(g-1)^2}{(m-2)[m(g-1)-g]} + \frac{(g-1)^4}{[m(g-1)-g][m(g-1)-2g]} + \frac{(g-1)^3}{g}\right.$$

$$E\left(\frac{1}{\bar{x}'_i\bar{x}'_j}\right)\Bigg] + \delta\left[\frac{g^2}{(m-1)(m-2)} - \frac{2g(g-1)^2}{(m-2)[m(g-1)-g]} + \frac{(g-1)^3}{[m(g-1)-g][m(g-1)-2g]}\right.$$

$$\left. + (g-1)(g-2)E\left(\frac{1}{\bar{x}'_i\bar{x}'_j}\right)\right] \tag{5.41}$$

where

$$E\left(\frac{1}{\bar{x}_i \bar{x}_j}\right) = \frac{\Gamma(2a+b-2)}{\Gamma^2(a)\Gamma(b)} \sum_{k=0}^{a-2} (-1)^k \frac{\Gamma(a-k-1)\Gamma(b+k)}{\Gamma(2a+b-k-2)} + (-1)^{a-1}\left\{2 \sum_{k=1}^{a+b-2} (-1)^{k+1} \frac{1}{(a+b-k-1)^2}\right.$$

$$\left. + (-1)^{a+b} \frac{\pi^2}{6}\right\}, \quad \text{if} \quad a \geq 2, \tag{5.41a}$$

$$= 2 \sum_{k=1}^{b-1} (-1)^{k+1} \frac{1}{(b-k)^2} + (-1)^{b+1} \frac{\pi^2}{6}, \quad \text{if} \quad a = 1, \tag{5.41b}$$

$$= \frac{\pi^2}{6}, \quad \text{if} \quad a = 1, \; b = 1, \tag{5.41c}$$

and $a = \frac{m}{g}$, $b = \frac{m(g-1)}{g}$.

It can be seen that the variance given by (5.41) is a function of m and g only. It is difficult to examine the behaviour of $V(\tilde{R}_J)$ as a function of g for fixed m analytically since the expression (5.41) would not be in a closed form after the substitution of (5.41a), (5.41b) and (5.41c). However, numerical results by these authors, which compare $\tilde{R}_J$, $\hat{R}_T$ and $\hat{R}$, indicate that the variance of $\tilde{R}_J$ is a decreasing function of g and the smallest value of $V(\tilde{R}_J)$ occurs at $g = n$. Thus, both the bias and variance of $\tilde{R}_J$ are decreasing functions of g (for integer m) and hence the optimum choice would be $g = n$. The discussion of numerical results is deferred until subsection 5.3.1.4. Rao and Webster also derived expressions for the exact bias and variance of Tin's estimator $\hat{R}_T$ under the model (5.37).

Rao (1967) investigated the precision of Mickey's estimator $\hat{R}_M$ under the above models and compared it with $\hat{R}_{HR}$, $\hat{R}_P$, $\tilde{R}_J$ and $\hat{R}_{Tg}$ (Tin, 1965). The derivation details are similar to those discussed earlier in this section and hence omitted. Roa's results indicate that for Model I, the asymptotic variance of $\hat{R}_M$ is less than the asymptotic variance of $\hat{R}$ but slightly greater than $V(\tilde{R}_J)$ with $g = n$. Theoretical comparisons under Model II are not possible, however, the Monte Carlo results support the jack-knifed ratio estimator $\tilde{R}_J$ and

the modified ratio restimator $\hat{R}_{Tg}$. Since the main reason for using the unbiased and approximately unbiased ratio estimators as compared to the classical ratio estimator is to reduce or eliminate bias where the elimination of bias is important. Therefore, the jack-knife estimator $\tilde{R}_J$ with $g = n$ and the estimator $\hat{R}_{Tg}$ may be preferable. The jack-knife estimator has the additional advantage of having the simple variance estimator.

5.3.1.3c Bias and Variance of the Jack-knifed Ratio Estimate Under Model III

Consider a bivariate sample (x_i, y_i), $i = 1,2,\ldots,n$, from a population of size N . Suppose that the regression of y on x is linear of the form

$$y_i = \alpha + \beta x_i + u_i \quad (i = 1,2,\ldots,n) \tag{5.42}$$

with $E(u_i/x_i) = 0$, $E(u_i^2/x_i) = \delta x_i^{\ell}$ and $E(u_i u_j/x_i x_j) = 0$, u is independent of x and x has a gamma distribution with parameter h . Generally ℓ is found to be between 0 and 2 . Under this model, Rao P.S.R.S. (1969) compares four ratio estimators $\hat{\bar{Y}}_R$, $\hat{\bar{Y}}_{R_{J2}}$, $\hat{\bar{Y}}_{RP}$ and $\hat{\bar{Y}}_{RGH}$ (Goodman and Hartley , 1958). Rao obtains exact expressions for the mean square errors of these estimators following Rao and Webster (1966). The derivation details are therefore omitted and the main results for the jack-knifed ratio estimate are summarised below:

$$\text{Bias}(\hat{\bar{Y}}_{R_{J2}}) = \frac{-2\alpha}{(nh-1)(nh-2)} \tag{5.43}$$

$$\text{MSE}(\hat{\bar{Y}}_{R_{J2}}) = \alpha^2 \frac{n^3h^3-5n^2h^2+12nh+16}{(nh-1)(nh-2)^2(nh-4)} + \delta\left[\frac{nh^2(n^2h^2+6nh\ell-7nh+9\ell^2-27\ell+18}{(nh+\ell-1)(nh+\ell-2)(nh+2\ell-4)} \cdot \frac{(h+\ell-1)!}{(h-1)!}\right] \tag{5.44}$$

This study reveals that for $\ell = 0$, the mean square errors of these estimators decrease with increasing h. For fixed n and h the mean square errors increase as ℓ increases. It is also observed that the MSE's for $\hat{\bar{Y}}_R$, $\hat{\bar{Y}}_{R_{J2}}$ and $\hat{\bar{Y}}_{RP}$ are equal to $0(n^{-1})$ and are given by

$$MSE(\hat{\bar{Y}}_R) = MSE(\hat{\bar{Y}}_{R_{J2}}) = MSE(\hat{\bar{Y}}_{RP}) = \frac{\alpha^2}{nh} + \frac{\delta}{n}\frac{(h+\ell-1)!}{(h-1)!} \tag{5.45}$$

The other point emerging from this study is that if $\alpha = 0$ the MSE's of the four estimators are equal and if $\alpha \neq 0$, the $V(\hat{\bar{Y}}_{RGH})$ is slightly larger than the MSE's of the other three estimators for large n. A comparison of the MSE $(\hat{\bar{Y}}_R)$ and $MSE(\hat{\bar{Y}}_{R_{J2}})$ indicates that the former is less than the later provided $nh > 16$ and $\ell \leq 1/2$.

In case of finite populations, Rao and Rao (1971a) report some theoretical results on the expected mean square errors of $\hat{R}$, $\tilde{R}_{J2}$ and $\hat{R}_{JJ}$ (Jones, 1962 version of the jack-knife). Under the model $y_i = \alpha + \beta x_i + e_i$ $(i = 1,2,\ldots,N)$ with $E(e_i/x_i) = 0$, $E(e_i e_j/x_i,x_j)$ for $i \neq j$, $E(e_i^2/x_i) = \delta x_i^{\ell}$ and with the further assumption that x has a gamma distribution with parameter h. These authors show that

$$E[MSE(\hat{R}) - MSE(\tilde{R}_{J2})] = \alpha^2\left[\frac{n(nh-16)}{h(nh-1)(nh-2)^2(nh-4)} - \frac{4n}{h(nh-1)(nh-2)(Nh-2)}\right] +$$
$$\delta m\left[\frac{nh(-2\ell+1)-5(\ell-1)(\ell-2)}{2(mh+\ell-1)(mh+\ell-2)(nh+\ell-1)(nh+\ell-2)}\right.$$
$$\left. + \frac{2(\ell-1)}{(mh+\ell-1)(nh+\ell-1)(Nh+\ell-2)}\right] \times \frac{\Gamma(\ell+h)}{\Gamma(h)} . \tag{5.46}$$

This expression implies that the expected mean square error of $\tilde{R}_{J2}$ is smaller than that of $\hat{R}$ provided $nh > 16$ and $\ell = 0$. Earlier Durbin (1959) showed this result for an infinite population model.

However, Durbin's result is not valid for a finite N. This study concludes that, for finite populations, Jones (1963) version of the jack-knife ratio estimator may be preferable. However, if the population size is infinite, the mean square error of $\hat{R}$ will be smaller than the mean square error of $\hat{R}_{JJ}$ provided that $nh < 16$ and $\frac{1}{2} < \ell < 2$. Similar results were obtained earlier by Rao (1969*a*) from an empirical study comparing the ratio and regression estimators for 16 natural populations with N ranging from 7 to 270, the coefficient of variation of x from 0.14 to 1.19 and ℓ from 0.535 to 0.995. The limitation of an empirical study is that the results are strictly applicable only to the particular populations studied. However, the utility of such studies can not be denied since they provide useful guidelines on the performance of various methods of estimation.

In a semi-empirical study Rao and Kuzik (1974) compared the average mean square errors of seven ratio estimators of the population mean $\bar{Y}$ and the average biases and mean square errors of two classical variance estimators and the jack-knife variance estimator under the model (5.42). Rao (1979) may also be mentioned in this context; for he derived expressions for the bias and MSE of the jack-knifed ratio estimator $\tilde{R}_J$ for $g = n$ and compared it with the biases and the MSE's of the classical ratio estimator $\hat{\bar{Y}}_R$ and a regression type estimator $\hat{\bar{Y}}_{RJg} = \bar{y} + \tilde{R}_J(\bar{x}-\bar{X})$. The efficiencies of these estimators have been compared through a small simulation study. We will address ourselves to these results in the next section.

5.3.1.3d Stability of the Jack-knifed Variance Estimator

The estimate of variance associated with an estimator is often used in drawing statistical inferences. It is, therefore, desirable that a variance estimator should be as stable as possible. Kokan (1963)

investigated the large sample stabilities of the conventional variance estimator $\hat{V}_{1C}$ and an unbiased variance estimator

$$\hat{V}(\bar{y}/\bar{X}) = \frac{s_y^2}{(n\bar{x})^2} \tag{5.47}$$

where $\bar{y}/\bar{X}$ is the ratio estimator of R not using the sample x - information. Through a Monte Carlo study, Rao and Beegle (1966) investigated the small sample properties of $\hat{V}_{1C}$, $\hat{V}(\bar{y}/\bar{X})$ and $\hat{V}_J(\tilde{R}_J)$ where

$$\hat{V}_J(\tilde{R}_J) = \frac{1}{g(g-1)} \sum_{j=1}^{g} (\tilde{R}_j - \tilde{R}_J)^2 \ . \tag{5.48}$$

The discussion of these numerical results is deferred until subsection 5.3.1.4. In this subsection, we give results on the stabilities of $\hat{V}(\bar{y}/\bar{X})$, $\hat{V}_{1C}$ and $\hat{V}_J(\tilde{R}_J)$ following Chakrabarty and Rao (1967). The method of derivation of these results is the same as given by Rao and Webster (1966) and hence we confine ourselves to main results so as to facilitate comparison.

Under the model (5.37) the variance of $\hat{R}$ can be given as

$$\begin{aligned} V(\hat{R}) &= E(\hat{R}-\beta)^2 - [E(\hat{R}-\beta)]^2 \\ &= \frac{\alpha^2}{(m-1)^2(m-2)} + \frac{\delta}{(m-1)(m-2)} \end{aligned} \tag{5.49}$$

(i) Bias and Variance of $\hat{V}_{1C}$ (Cochran, 1977) as an Estimator of $V(\hat{R})$

$$\begin{aligned} \text{Bias}(\hat{V}_{1C}) &= E(V(\hat{R})) - V(\hat{R}) \\ &= -\frac{(5m^2-5m+2)\alpha^2}{m^2(m^2-1)(m-1)(m-2)} - \frac{2\delta(m^2+2m-2)}{m^2(m^2-1)(m-2)} \\ &= A_1\alpha^2 + A_2\delta \end{aligned} \tag{5.50}$$

The variance of $\hat{V}_{1C}$ under the model (5.37) is

$$V(\hat{V}_{1C}) = E[\hat{V}_{1C}]^2 - [E(\hat{V}_{1C})]^2$$

$$= \frac{\delta^2}{m^4}\left[3\eta + \frac{(n+1)(n+3)}{(n-1)(m+1)} - \frac{(m+2)^2}{(m+1)^2}\right] + \frac{2\alpha^2\delta}{m^4}\left[3\eta + \frac{2m-n+3}{(n-1)(m+1)^2}\right]$$

$$+ \frac{\alpha^4}{m^4}\left[\eta - \frac{1}{(m+1)^2}\right] . \qquad (5.51)$$

where

$$\eta = \frac{(n+1)(m+6)-12}{(n-1)(m+3)(m+2)(m+1)} .$$

$$MSE(\hat{V}_{1C}) = Var[\hat{V}_{1C}] + [Bias\ \hat{V}_{1C}]^2$$

$$= \frac{\delta^2}{m^4}\left[3\eta + \frac{(n+1)(n+3)}{(n-1)(m+1)} - \frac{(m+2)^2}{(m+1)^2} + \frac{4(m^2+2m-2)}{(m^2-1)^2(m-2)^2}\right]$$

$$+ \frac{2\alpha^2\delta}{m^4}\left[3\eta + \frac{2m-n+3}{(n-1)(m+1)^2} + \frac{2(5m^2-5m+2)(m^2+2m-2)}{(m^2-1)^2(m-1)(m-2)^2}\right]$$

$$+ \frac{\alpha^4}{m^4}\left[\eta - \frac{1}{(m+1)^2} + \frac{(5m^2-5m+2)}{(m^2-1)^2(m-1)^2(m-2)^2}\right]$$

$$= A_3\delta^2 + A_4\alpha^4 + A_5\alpha^2\delta . \qquad (5.52)$$

where the coefficients A_3, A_4 and A_5 are functions of m and n .

(ii) *Variance of* $\hat{V}(\bar{y}/\bar{X})$

The variance of $(\bar{y}/\bar{X})$ under the model (5.37) is

$$V(\bar{y}/\bar{X}) = \frac{\delta}{m^2} + \frac{\beta^2}{m} \qquad (5.53)$$

Now the variance of $\hat{V}(\bar{y}/\bar{X})$ as an unbiased estimator of $V(\bar{y}/\bar{X})$ can be given as

$$V[\hat{V}(\bar{y}/\bar{X})] = E[\hat{V}(\bar{y}/\bar{X})]^2 - [E(\hat{V}(\bar{y}/\bar{X}))]^2$$

$$= \frac{2\delta^2}{m^4(n-1)} + \frac{4\beta^2\delta}{m^3(n-1)} + \frac{\beta^4}{m^3}[\eta(m+1)(m+2)(m+3)-m] \qquad (5.54)$$

(iii) Bias and Variance of $\hat{V}_J(\tilde{R}_J)$

The jack-knife variance estimator can be written as

$$\hat{V}_J(\tilde{R}_J) = \frac{(g-1)^2}{g(g-1)}\left[\alpha^2\left\{\frac{1}{\bar{x}'_j} - \frac{1}{g}\sum_{j=1}^{g}\frac{1}{\bar{x}'_j}\right\}^2 + \left\{\frac{\bar{u}'_j}{\bar{x}'_j} - \frac{1}{g}\sum_{j=1}^{g}\frac{\bar{u}'_j}{\bar{x}'_j}\right\}^2 + \right.$$

$$\left. + 2\alpha\sum_{j=1}^{g}\left(\frac{1}{\bar{x}'_j} - \frac{1}{g}\sum_{j=1}^{g}\frac{1}{\bar{x}'_j}\right)\left(\frac{\bar{u}'_j}{\bar{x}'_j} - \frac{1}{g}\sum_{j=1}^{g}\frac{\bar{u}'_j}{\bar{x}'_j}\right)\right].$$

Following the same arguments as in subsection 5.3.1.3b, we get

$$E[\hat{V}_J(\tilde{R}_J)] = \alpha^2\left[\frac{(g-1)^4}{g[m(g-1)-g][m(g-1)-2g]} - \frac{(g-1)^2}{g}E\left(\frac{1}{\bar{x}'_j\bar{x}'_k}\right)\right]$$

$$+ \delta\left[\frac{(g-1)^3}{[m(g-1)-g][m(g-1)-2g]} - (g-2)\,E\left(\frac{1}{\bar{x}'_j\bar{x}'_k}\right)\right]$$

$$\text{Bias}[\hat{V}_J(\tilde{R}_J)] = E[\hat{V}_J(\tilde{R}_J)] - V(\hat{R})$$

$$= \alpha^2\left[\frac{(g-1)^4}{g[m(g-1)-g][m(g-1)-2g]} - \frac{1}{(m-1)^2(m-2)} - \frac{(g-1)^2}{g}\right.$$

$$\left.E\left(\frac{1}{\bar{x}'_j\bar{x}'_k}\right)\right] + \delta\left[\frac{(g-1)^3}{[m(g-1)-g][m(g-1)-2g]} - \frac{1}{(m-1)(m-2)} - (g-2)\right.$$

$$\left.E\left(\frac{1}{\bar{x}'_j\bar{x}'_k}\right)\right] \qquad (5.55)$$

This expression is not in a closed form and hence it is difficult to examine its behaviour analytically. However, the numerical results indicate that the absolute bias of $\hat{V}_J(\tilde{R}_J)$ with $g = n$ is less than that of $\hat{V}_{1C}$ for $n \geq 6$ and $m > 8$.

Now we examine the stability of the jack-knife variance estimator $\hat{V}_J(\tilde{R}_J)$ under the model (5.37) for $g = 2$.

$$\mathrm{Var}[\hat{V}_J(\tilde{R}_{J2})] = E[\hat{V}_J(\tilde{R}_{J2})]^2 - [E(\hat{V}_J(\tilde{R}_{J2}))]^2 .$$

Under the model,

$$\hat{V}_J(\tilde{R}_{J2}) = \frac{1}{4}\left[\alpha^2\left(\frac{1}{\bar{x}_1'} - \frac{1}{\bar{x}_2'}\right)^2 + \left(\frac{\bar{u}_1'}{\bar{x}_1'} - \frac{\bar{u}_2'}{\bar{x}_2'}\right)^2 + 2\alpha\left(\frac{1}{\bar{x}_1'} - \frac{1}{\bar{x}_2'}\right)\left(\frac{\bar{u}_1'}{\bar{x}_1'} - \frac{\bar{u}_2'}{\bar{x}_2'}\right)\right] \qquad (5.56)$$

where $\bar{x}_1'$ and $\bar{x}_2'$ are the means of the first and second half samples respectively and they are independent gamma variates with parameters $m/2$. From Chakrabary and Rao (1967), we have

$$E\left(\frac{1}{\bar{x}_1^{t}}\right) = E\left(\frac{1}{\bar{x}_2^{t}}\right) = \frac{1}{\prod_{i=1}^{t}(m-2i)} \quad , \quad t \geq 1$$

Hence,

$$E(\hat{V}_J(\tilde{R}_{J2})) = \frac{\alpha^2}{(m-2)^2(m-4)} + \frac{\delta}{(m-2)(m-4)} .$$

Bias $\hat{V}_J(\tilde{R}_{J2})$ as an estimator of $V(\hat{R})$ is

$$\mathrm{Bias}(\hat{V}_J(\tilde{R}_{J2}) = E[\hat{V}_J(\tilde{R}_{J2})] - V(\hat{R})$$

$$= \frac{4m-7}{(m-1)^2(m-2)^2(m-4)} + \frac{3\delta}{(m-1)(m-2)(m-4)} \qquad (5.57)$$

This expression shows that the bias of $\hat{V}_J(\tilde{R}_{J2})$ decreases as m increases.

Now

$$E(\hat{V}_J(\tilde{R}_{J2}))^2 = \frac{1}{16} E\left[\alpha^4\left(\frac{1}{\bar{x}_1'} - \frac{1}{\bar{x}_2'}\right)^4 + \left(\frac{\bar{u}_1'}{\bar{x}_1'} - \frac{\bar{u}_2}{\bar{x}_2'}\right)^4 + 6\,\alpha^2\left(\frac{1}{\bar{x}_1'} - \frac{1}{\bar{x}_2'}\right)^2 \left(\frac{\bar{u}_1'}{\bar{x}_1'} - \frac{\bar{u}_2'}{\bar{x}_2'}\right)^2\right].$$

The variance is therefore given by

$$V[\hat{V}_J(\tilde{R}_{J2})] = E[\hat{V}_J(\tilde{R}_{J2})]^2 - [E(\hat{V}_J(\tilde{R}_{J2})]^2$$

$$= \frac{2\alpha^4(m^2+m-18)}{(m-1)^2(m-4)^2(m-6)(m-8)} + \frac{2\delta^2(m^2-8m+19)}{(m-2)^2(m-4)^2(m-6)(m-8)}$$

$$+ \frac{4\alpha^2\delta(m^2+m-18)}{(m-2)(m-4)^2(m-6)(m-8)} \qquad (5.58)$$

Using (5.57) and (5.58), we can obtain $MSE(\hat{V}_J(\tilde{R}_{J2}))$ which is

$$MSE[\hat{V}_J(\tilde{R}_{J2})] = \alpha^4\left[\frac{2(m^2+m-18)(m-1)^4+(4m-7)^2(m-6)(m-8)}{(m-1)^4(m-2)^4(m-4)^2(m-6)(m-8)}\right] +$$

$$\frac{2(m^2-8m+19)(m-1)^2+9(m-6)(m-8)}{(m-1)^2(m-2)^2(m-4)^2(m-6)(m-8)}\,\delta^2 +$$

$$\alpha^2\delta\,\frac{4(m^2+m-18)(m-1)^3+6(4m-7)(m-6)(m-8)}{(m-1)^3(m-2)^3(m-4)^2(m-6)(m-8)}$$

$$= A_6\delta^2 + A_7\alpha^4 + A_8\alpha^2\delta \quad . \qquad (5.59)$$

where the coefficients A_6, A_7 and A_8 are functions of $m(>8)$ only. The results of numerical investigation (Chakrabarty and Rao, 1967) for selected values of m and n indicate that the jack-knife variance estimator $\hat{V}_J(\tilde{R}_{J2})$ with $g = 2$ is not very stable. The bias of $\hat{V}_J(\tilde{R}_J)$ with $g = n$ is minimum and $MSE[\hat{V}_J(\tilde{R}_{J2})]$ with $g = 2$ is considerably larger than that of $\hat{V}_{1C}$.

Through an empirical study, Rao (1969a) compares the stabilities of these variance estimates from 16 natural populations. Assuming that the errors u_i are independently and normally distributed and using the measure $\xi E[\hat{V}_i - \varepsilon MSE(\hat{\bar{Y}}_R)]^2$ instead of $\xi E[\hat{V}_i - MSE(\hat{\bar{Y}}_R)]^2$, where ξ is the expectation operator, Rao and Kuzik (1974) also compare the stabilities of these variance estimators under Model III.

5.3.1.4 An Evaluation of Empirical Results for Jack-knifed Ratio Estimator

In the previous section we have reviewed the theoretical work done in connection with the jack-knifed ratio estimator and other competing estimators. Along with these theoretical developments considerable attention has also been given to empirical work in order to compare the efficiencies of these estimators and the corresponding variance estimators in small and moderate size samples from natural and artificial populations. In this section we evaluate the empirical work on ratio estimation. Our discussion is based on two main approaches, as outlined earlier in Chapters 2 and 3, the randomisation based approach and the model based approach. Firstly we address ourselves to the design oriented results and thereafter to prediction approach results.

5.3.1.4a Randomisation Based Results

The randomisation based approach, as pointed out earlier in Chapter 2, provides a measure of error based on the probability distribution induced by the randomisation scheme employed in selecting the sample. Under this approach we discuss empirical results in two categories, that is, variances of the estimators and the estimators of the variances of these estimates.

(1) Efficiencies of the Estimates

Assuming a linear regression of y on x where x is normally distributed, Rao and Bugle (1966) investigated the small sample efficiencies of eight ratio estimators $\hat{R}$, $\hat{R}_{HR}$, $\tilde{R}_{J2}$, $\hat{R}_B$, $\hat{R}_T$, $\hat{R}_M$ and $\hat{R}_p$ through a Monte Carlo study based on 1000 samples. These authors used Lauh and William (1963) and Tin (1965) models for their study. Lauh and William's model is ideal for the ratio estimator since it assumes that the regression is through the origin and the coefficient of variation of x is small. In Tin's model, the regression is not through the origin and the coefficient of variation of x is **not small**.

This study reveals that under Lauh and William model all the ratio estimators are unbiased and there is very little difference in the variances of these estimators even for small n. This study, therefore, suggestes that so far as this model is concerned it does not matter which of these eight ratio estimators is used. The optimum number of groups for the jack-knifed ratio estimator $\tilde{R}_J$ is n. The estimator $\hat{R}_{HR}$ is less efficient than the rest of the estimators.

Under Tin's model the differences in the efficiencies of these ratio estimators are quite promising. The estimator $\hat{R}_{Tg}$ is most efficient at $g = 2$ which may be due to the reason that $\hat{R}_{Tg}$ is biased under this model and the bias increases as g increases. For $n = 4$ and 6 , the efficiency of $\hat{R}_B$ is greater than the rest of the estimators whereas for $n>10$ efficiency of $\tilde{R}_J(g = n)$, $\hat{R}_M$, $\hat{R}_B$ and $\hat{R}_T$ are approximately the same.

It could be concluded from this study that in case of approximately unbiased estimators, $\tilde{R}_J$ (with $g = n$) and $\hat{R}_{Tg}$ appear to be more promising. Earlier, Lauh and William (1963) examined the behaviour of $\tilde{R}_J$ and $\hat{R}$ under non-normal populations and concluded that $\tilde{R}_J$ is

definitely superior to $\hat{R}$. These results also find support from another study conducted by Rao (1967). The important point conveyed by Rao's study is that the jack-knifed ratio estimator $\tilde{R}_J$ (with $g = n$) performs considerably better than the competing estimators $\hat{R}_{HR}$, $\hat{R}_M$, $\hat{R}_{Tg}$, $\hat{R}_p$ and $\hat{R}$ for $n > 2$. With $g = 2$, the estimator $\tilde{R}_{J2}$ is as efficient as $\hat{R}_{Tg}$ which is in agreement with the results reported earlier by Rao and Webster (1966) regarding the efficiencies of these two estimators.

In short, based on the empirical studies by Rao (1969b), Rao and Kuzik (1974), Rao (1979) and others,under the more general model (5.42), the following conclusions may be drawn.

(i) When $\alpha = 0$ and $\ell = 0$, the jack-knifed ratio estimator $\tilde{R}_J$ is more efficient than the classical ratio estimator $\hat{R}$ and the regression type estimator $\hat{\bar{Y}}_{R_{Jg}}$. Among other estimators, for $n = 2$ and the coefficient of variation of x less than 32 per cent, $MSE(\hat{\bar{Y}}_{R_{J2}}) < MSE(\hat{\bar{Y}}_{RP})$. For $n > 2$ and the coefficient of variation of x less than 32 per cent, $MSE(\hat{\bar{Y}}_{R_{J2}}) < MSE(\hat{\bar{Y}}_R)$.

(ii) When $\alpha \neq 0$ and $\ell = 0$, the jack-knifed ratio estimator $\hat{\bar{Y}}_{\tilde{R}_J}$ is more efficient than $\hat{\bar{Y}}_R$ and the regression type estimator $\hat{\bar{Y}}_{R_{Jg}}$. If the coefficient of variation is less than 25 per cent then $MSE(\hat{\bar{Y}}_{R_{J2}}) < MSE(\hat{\bar{Y}}_R)$, see for instance, Durbin (1959) and Rao (1969b).

(iii) When $\alpha \neq 0$ and $\ell \geq \frac{1}{2}$, then $MSE(\hat{\bar{Y}}_R) < MSE(\hat{\bar{Y}}_{R_{J2}})$ provided the coefficient of variation of x is less than 25 per cent.

(iv) When $\alpha \neq 0$ and $\ell = 1$, the jack-knife ratio estimator $\hat{\bar{Y}}_{\tilde{R}_J}$ is more efficient than $\hat{\bar{Y}}_R$.

(v) When $\alpha = 0$ and $\ell = 1$ or $\ell = 2$, the classical ratio estimator $\hat{\bar{Y}}_R$ is more efficient than $\hat{\bar{Y}}_{\tilde{R}_J}$, $\hat{\bar{Y}}_{RP}$ and $\hat{\bar{Y}}_{RGH}$. In case $n \geq 4$ and $\ell = 2$, $\hat{\bar{Y}}_{RT}$ has substantial gain in efficiency over $\hat{\bar{Y}}_{\tilde{R}_J}$ provided the coefficient of variation of x is greater than 0.75, see for instance, Rao and Kuzik (1974).

It could be seen from these empirical results that it is difficult to draw any definite general conclusions about the superiority of any of these estimators over the others. However, the jack-knifed ratio estimator $\tilde{R}_J$ (with $g = n$) and the modified ratio estimator $\hat{R}_{Tg}$ are slightly preferable. The jack-knifed estimator has the additional advantage of having a simple variance estimator.

(2) Variance Estimators

The properties of the jack-knife and other competing variance estimators, for small and moderate size samples, under models I, II and III have been studied among others by Tin (1965), Rao and Beegle (1966), Chakrabarty and Rao (1967), Rao and Rao (1971b), Rao and Kuzik (1974), Krewski and Chakrabarty (1981), Rao (1981) and Wu (1982). We have explored these studies in a greater detail. The details appear to be messy and hence we summarise the main conclusions below.

(i) When $\alpha = 0$, (i.e. the regression line is through the origin) and the coefficient of variation of x is small, $g = n$ is the optimal choice.

(ii) The variance estimators $\hat{V}_J(\tilde{R}_J)$ and $\hat{V}_{1C}$ are quite stable for any n as compared to $\hat{V}(\bar{y}/\bar{X})$, provided the regression is approximately through the origin and the coefficient of variation of x is small. The stabilities of $\hat{V}_J(\tilde{R}_J)$ and $\hat{V}_{1C}$ are approximately the same.

However, in case of exponential distribution the jack-knife variance estimator $\hat{V}_J(\tilde{R}_J)$ is much better than $\hat{V}_{1C}$.

(iii) Under Tin's model ($\alpha = 0$ and the coefficient of variation of x is not small), the coefficients of variation of $\hat{V}_J(\tilde{R}_J)$ and $\hat{V}_{1C}$ are essentially equal as under Lauh and William's model in (ii) above. These coefficients of variation are, however, greater than that for $\hat{V}(\bar{y}/\bar{X})$.

(iv) Under model (5.42), the variance estimators $\hat{V}_{1C}$ and $\hat{V}_{2C}$ underestimate $MSE(\hat{\bar{Y}}_R)$ whereas $\hat{V}_J$ over estimates $MSE(\hat{\bar{Y}}_R)$ for all values of ℓ. However, when $\alpha = 0$ and $\ell = 0$ as well, $\hat{V}_{2C}$ overestimates $MSE(\hat{\bar{Y}}_R)$. Rao and Kuzik (1974) through a semi-empirical study investigated the performance of $\hat{V}_{1C}$, $\hat{V}_{2C}$ and $\hat{V}_J$ under model (5.42) and reported that $\hat{V}_{1C}$ and $\hat{V}_{2C}$ are more stable than $\hat{V}_J$. Earlier, Rao and Rao(1971b) showed that $\hat{V}_{2C}$ has a smaller absolute bias as compared to $\hat{V}_{1C}$. Their results indicate that $\hat{V}_{1C}$ is more stable than $\hat{V}_{2C}$ for $\ell = 0$ and 1 whereas for $\ell = 2$, the variance estimator $\hat{V}_{2C}$ is more stable than $\hat{V}_{1C}$. This conclusion is also supported by Krewski and Chakrabarty (1981) who investigated the properties of $\hat{V}_{1C}$, $\hat{V}_{2C}$ and a slightly modified version of the jack-knifed variance estimator $\hat{V}_{JKC} = \frac{1}{g(g-1)} \sum_{j=1}^{g} (\tilde{R}_j - \hat{R})^2$ under the model (5.42) when u_i's are normally distributed. Considering model (5.42)and different distributions for x, Wu (1982) also supports above conclusions.

(v) Rao (1981) compared the biases and the mean square errors of $\hat{V}_{1C}$, $\hat{V}_L$, $\hat{V}_J$ and a modified jack-knifed variance estimator $\hat{V}_{JR} = \frac{(1-f)}{g} \frac{\bar{X}^2}{\bar{x}} \sum_{j=1}^{g} (y_i - \hat{R}x_i)^2/(g-1) = \frac{\bar{X}^2}{\bar{x}} \hat{V}_{1C}$ for $g = n$. He considered model (5.42) and showed that $\hat{V}_{1C}$ underestimates $V(\hat{\bar{Y}}_R)$ when $\ell = 0$, 1 or 2. This supports the results earlier given by Rao and Kuzik (1974).

$\hat{V}_L$ is unbiased for $\ell = 1$ but overestimates $V(\hat{\bar{Y}}_R)$ for $\ell = 0$ and underestimates for $\ell = 2$. The variance estimator $\hat{V}_{JR}$ overestimates $V(\bar{Y}_R)$ when $\ell = 0$ and underestimates when $\ell = 1$ and 2. The important features of this study is that $\hat{V}_{JR}$ is less stable than $\hat{V}_{1C}$ and $\hat{V}_J$.

(vi) When $\alpha = 0$, that is the regression line is through the origin $\hat{V}_{JKC}$ and $\hat{V}_{2C}$ are somewhat equally stable for $\ell = 0$ provided n is moderately large or C_x is relatively small. However, for $\ell = 2$, the results concerning the mean square error of $\hat{V}_{1C}$, $\hat{V}_{2C}$ and $\hat{V}_{JKC}$ are not clear.

It could be observed from the above results that no clear picture emerges about the relative merits of the ratio estimates and the estimators of the variances of these estimates under the randomisations approach. The reason is that these authors have followed the rules of the conventional theory in studying the average performance of these estimators and by averaging over all possible samples they have been led to ambiguous results.

5.3.1.4b Model Based Results

The model based measure of uncertainty, as discussed earlier in Chapter 3, is a property of the units in the sample actually observed and does not depend on the manner in which the units are selected. In this section we give some empirical results regarding the properties of the jack-knifed variance estimator $\hat{V}_J$, and other competing estimators, $\hat{V}_{1C}$, $\hat{V}_L$, $\hat{V}_H$ and $\hat{V}_D$ under the model ξ. Royall and Cumberland, through a series of investigations (1978, 1981a, 1982, 1983) have compared the properties of these variance estimators using simple random samples, restricted random samples and best fit samples from six real populations under model (3.91). These studies show that $\bar{x}_s$ and $\sum_s x_i^2/n$ are the key sample characteristics. The ratio estimator

for the population total $\hat{T}_R$ can be badly biased in samples where the averages $\bar{x}_s$ and $\sum_s x_i^2/n$ are not close to the population averages $\bar{x}$ and $\sum_{i=1}^{N} x_i^2/N$ respectively. Moreover, the behaviour of the variances and variance estimators is strongly dependent on $\bar{x}_s$.

The results for simple random samples and error curves of Royall and Cumberland (1978a, 1981a) and Cumberland and Royall (1983) indicate that the variance estimator $\hat{V}_{1C}$ tends to be too small when $\bar{x}_s$ is smaller than $\bar{x}$ and too large when $\bar{x}_s$ is larger. The performance of $\hat{V}_J$, $\hat{V}_D$ and $\hat{V}_H$ in tracking the actual mean square error, as $\bar{x}_s$ varies, is much superior to $\hat{V}_{1C}$ and $\hat{V}_L$. The performance of $\hat{V}_L$, obtained directley from the prediction approach, is very poor as compared to $\hat{V}_J$, $\hat{V}_D$ and $\hat{V}_H$ when the model (3.91) does not hold. This implies that the variance estimators $\hat{V}_H$, $\hat{V}_D$ and $\hat{V}_J$ are more robust to model deviations in the variance function as compared to $\hat{V}_{1C}$ and $\hat{V}_L$. We also note from these studies that the performance of the bias robust variance estimator $\hat{V}_H$ is quite similar to that of $\hat{V}_D$. The important feature of these studies is that they illustrate that even in well balanced samples the least squares variance estimator $\hat{V}_L$ is biased if $\mathrm{Var}(Y_i)$ is not proportional to x_i. Their error curves also lead to an important conclusion that on the whole the best samples for these statistics are those that are well balanced on x.

In case of restricted random sampling, the variance estimates $\hat{V}_{1C}$, $\hat{V}_D$ and $\hat{V}_H$ give similar results. The error curves for the jack-knife variance estimate $\hat{V}_J$ are very close to $\hat{V}_{1C}$, $\hat{V}_D$ and $\hat{V}_H$ whereas those for $\hat{V}_L$ are considerably below $\hat{V}_{1C}$. The general pattern of these curves is $\hat{V}_J > \hat{V}_D > \hat{V}_H > \hat{V}_{1C}$.

The results for the best samples give no general pattern for the superiority of any of the variance estimators $\hat{V}_J$, $\hat{V}_{1C}$, $\hat{V}_H$ and $\hat{V}_D$. However these results indicate that in spite of the absence of randomisation, the best fit samples provide estimates of precision of the variance estimates $\hat{V}_{1C}$, $\hat{V}_D$, $\hat{V}_H$ and $\hat{V}_J$ which appear to be reasonable.

Royall and Cumberland (1982), and Cumberland and Royall (1983) through an empirical study investigated the coverage properties of confidence intervals for the ratio estimator $\hat{T}_R$, using the standard normal approximation with variance estimates $\hat{V}_{1C}$, $\hat{V}_L$, $\hat{V}_D$ and $\hat{V}_J$, for the samples drawn at random from the same six populations used in their 1981 study. In essence, these investigations reveal that the standard regression theory variance estimator $\hat{V}_L$ produced excess of standardised errors (SEZ's) in every group and for all populations and hence this estimator is useless in practice. The variance estimators $\hat{V}_J$ and $\hat{V}_D$ showed comparatively more stable and consistent performance over all groups for some populations. However, these two bias robust variance estimators also performed badly at the extremes in some populations (Hospitals and Counties) and gave an excess of large negative SZE's in every group, even those where $\bar{x}_s$ was near to $\bar{x}$, in certain populations (Counties 70). The conventional variance estimator $\hat{V}_{1C}$ performed better than $\hat{V}_L$ in almost all the populations but was in no case better than $\hat{V}_J$ and $\hat{V}_D$.

These empirical results indicate that although the performance of the robust variance estimators $\hat{V}_J$ and $\hat{V}_D$ is better than $\hat{V}_{1C}$ in almost all the populations yet it is not quite satisfactory because both these estimators produced a large excess of extreme SZE's in samples showing the worst imbalance (e.g. Hospitals and Counties 70) as well as in samples that are well balanced (e.g. Counties 70). The

problem, however, diminishes to some extent when the sample size increases but persists.

5.3.1.5 Application to Complex Surveys

The jack-knife method, as discussed earlier in Chapter 3, derives the estimates of variance from the variability among a number of replication estimates of θ computed from overlapping subsamples of the total sample. Miller (1974a) reviews the properties of the jack-knife method for simple random sampling with replacement. Let $\hat{\theta}_{-i}$ denote the estimator of θ computed from the sample after omitting the ith observation, then jack-knife variance estimator for simple random sampling with replacement is given by

$$\hat{V}_J(\tilde{\theta}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\tilde{\theta}_i - \tilde{\theta})^2 , \qquad (5.60)$$

where $\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i}$ are the pseudo values (see Section 3.2.2.9).

Jones (1962, 63) suggested a modification of the jack-knife method for use in sampling without replacement from a finite population. Jones suggested that for $m = 1$, the pseudovalues could be defined as

$$\tilde{\theta}^*_j = w\hat{\theta} - (w-1)\hat{\theta}_{-j} ,$$

where $w = \left(\frac{n}{N-n}\right) \Big/ \left(\frac{n}{N-n} - \frac{n-1}{N-n+1}\right)$.

The jack-knife estimator is

$$\tilde{\theta}^* = \frac{1}{n} \sum_{j=1}^{n} \tilde{\theta}^*_j = w\hat{\theta} - \frac{w-1}{n} \sum_{j=1}^{n} \hat{\theta}_{-j} \qquad (5.62)$$

where $(w-1) = (n-1)(N-n)/N$.

The estimator of the variance based on pseudovalues $\tilde{\theta}^*_j$ is given by

$$\hat{V}^*_J = \frac{(N-n)(n-1)}{Nn} \sum_{j=1}^{n} (\tilde{\theta}^*_j - \tilde{\theta}^*)^2/(w-1) \ .$$

$$= \frac{N-n}{N} \frac{1}{n(n-1)} \sum_{j=1}^{n} (\tilde{\theta}_j - \tilde{\theta})^2$$

$$= \frac{N-n}{N} \hat{V}_J(\tilde{\theta}) \ . \qquad (5.63)$$

Some large sample results for simple random sampling without replacement, using U statistic have been derived by Krewski (1978a) and Majumdar and Sen (1978).

The extension of the jack-knife to complex surveys is not immediate and no general definition is available for jack-knifing in a stratified multistage sample. However, several different versions of the jack-knife procedures have been suggested for complex surveys. In this section, we briefly discuss the applicability of the jack-knife for estimation of variance from complex surveys.

5.3.1.5a Stratified Sampling

McCarthy (1966) extended the application of the jack-knife method to stratified sampling without replacement. However, most of the developments for the application of the jack-knife assume that the PSU's are selected with replacement. In an attempt to extend the application of the jack-knife to complex surveys different definitions of pseudovalues have emerged which have resulted in different jack-knife estimates. We attempt to summarise various approaches for obtaining jack-knifed variance estimates from stratified samples.

(i) $n_h = 2$

Consider a stratified random sample based on L strata and sample observations $(y_{11}, y_{12}), \ldots, (y_{L1}, y_{L2})$ per stratum. An unbiased estimate of the population mean $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$$

where $\bar{y}_h = \sum_{j=1}^{n_h} y_{hj}/n_h$ denotes the sample mean in the hth stratum and W_h are the relative stratum sizes with $\sum_{h=1}^{L} W_h = 1$.

The usual sample estimate of the $V(\bar{y}_{st})$ is

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{s_h^2}{n_h} = \frac{1}{4} \sum_{h=1}^{L} W_h^2 d_h^2 \qquad (5.64)$$

where $d_h = y_{h1} - y_{h2}$.

Case I Now, if we delete one observation per stratum and denote by $\hat{\theta}_{-hj}$ the estimator based on the entire sample but with the jth observation in the hth stratum deleted and $\hat{\theta}$ as the estimator based on the entire sample then the pseudovalues can be defined as

$$\tilde{\theta}_{hj} = 2\hat{\theta} - \hat{\theta}_{-hj} = \hat{\theta} \mp \frac{1}{2} W_h d_h \qquad (5.65)$$

The jack-knifed estimator is

$$\tilde{\theta}_1 = \frac{1}{2L} \sum_{h=1}^{L} \sum_{j=1}^{2} \tilde{\theta}_{hj} \qquad (5.66)$$

The jack-knifed variance estimator in this case is

$$\hat{V}_{J_1}(\tilde{\theta}_1) = \sum_{h=1}^{L} \frac{1}{2} \sum_{j=1}^{2} (\tilde{\theta}_{hj} - \tilde{\theta}_1)^2$$

$$= \frac{1}{4} \sum_{h=1}^{L} W_h^2 d_h^2 = \hat{V}(\bar{y}_{st}) \quad . \tag{5.67}$$

which is the ordinary estimate of the variance of $\bar{y}_{st}$.

Case II Kish and Frankel (1970) proposed a variant of the jack-knife method known as the jack-knife repeated replications technique (JRR) where each replication gives a measure of the variance contributed by a single stratum. This technique measures the stratum variance contribution for variance estimates by leaving out replicates. The method consists in forming a set of replicates by deleting one primary selection from each stratum and adding twice the stratum's other primary selection. This method yields four variance estimators. Let θ_i^* denote the jack-knifed replication and θ_i^{**} be its compliment, then variance estimators based on the half-sample, complementary sample and those based on the difference and the sum of the half-sample and complementary sample are:

$$\hat{V}_{JRR\text{-}HS}(\tilde{\theta}) = (1-f) \sum_{i=1}^{L} (\theta_i^* - \tilde{\theta})^2 \tag{5.68}$$

$$\hat{V}_{JRR\text{-}C}(\tilde{\theta}) = (1-f) \sum_{i=1}^{L} (\theta_i^{**} - \tilde{\theta})^2 \tag{5.69}$$

$$\hat{V}_{JRR\text{-}D}(\tilde{\theta}) = (1-f) \sum_{i=1}^{L} (\theta_i^* - \theta_i^{**})^2 \tag{5.70}$$

$$\hat{V}_{JRR\text{-}S}(\tilde{\theta}) = (1-f) \sum_{i=1}^{L} [(\theta_i^* - \tilde{\theta})^2 + (\theta_i^{**} - \tilde{\theta})^2] \tag{5.71}$$

where $f = \frac{2}{N_h}$

The variance estimates (5.68) and (5.69) are cheaper to compute but less precise than (5.71). Lee (1973) also proposed a variance estimator which is equivalent to $\hat{V}_{JRR\text{-}S}(\tilde{\theta})$. Lee's variance estimator is

$$\hat{V}_{JL}(\tilde{\theta}) = \sum_{h=1}^{L} \frac{1}{2} \sum_{j=1}^{2} (\tilde{\theta}_{hj} - \tilde{\theta})^2 = \frac{1}{4} \sum_{h=1}^{L} W_h^2 d_h^2 \qquad (5.72)$$

(ii) $n_h > 2$

Case I: Suppose that n_h units are selected from the N_h units in the stratum $h(h = 1,2,\ldots,L)$ by simple random sampling without replacement in each stratum. Then Jones (1974) derived jack-knife estimators by employing Taylor Series expansion. Let $\hat{\theta}_{-hj}$ denote the estimator of θ obtained by omitting the jth observation from the hth stratum and let $\tilde{\theta}_{(h)}$ denote the mean of n_h observations, then Jones estimators are:

$$\tilde{\theta}_J = \left[1 + \sum_{h=1}^{L} (n_h - 1)\left(1 - \frac{n_h}{N}\right)\hat{\theta} - \sum_{h=1}^{L} (n_h - 1)\left(1 - \frac{n_h}{N}\right)\hat{\theta}_{(h)}\right] \qquad (5.73)$$

where $\hat{\theta}_{(h)} = \sum_{j=1}^{n_h} \hat{\theta}_{-hj}/n_h$.

The variance estimator is given by

$$\hat{V}_{J_2}(\tilde{\theta}_J) = \sum_{h=1}^{L} \left(1 - \frac{n_h}{N_h}\right) \frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} (\hat{\theta}_{-hj} - \hat{\theta}_{(h)})^2 \qquad (5.74)$$

The estimators (5.73) and (5.74) have also been suggested by Brillinger (1977). If $n_h = 2$, then $\left(1 - \frac{n_h}{N_h}\right) = 1$ and (5.73) reduces to McCarthy's estimator (5.65) and the variance estimator (5.74) reduces to $\hat{V}_J(\tilde{\theta}) = \sum_{h=1}^{L} \frac{1}{2} \sum_{j=1}^{2} (\hat{\theta}_{-hj} - \tilde{\theta})^2$ which is the same as considered by Lee (1973) and Kish and Frankel (1974). Now if we replace $\hat{\theta}_{(h)}$ in (5.74) by $\sum_h \sum_j \hat{\theta}_{-hj}/n$ and $\frac{1}{L}\sum_h^L \hat{\theta}_{(h)}$ respectively, we get two variants of $\hat{V}_{J_2}(\tilde{\theta}_J)$.

It is interesting to note that if the pseudo values (5.65) are defined as

$$\tilde{\theta}_{hj} = n_h\hat{\theta} - (n_h-1)\hat{\theta}_{-hj} \ , \tag{5.75}$$

then the jack-knife estimator is given by

$$\tilde{\theta}_{J_1} = \frac{1}{n}\sum_{h=1}^{L}\sum_{j=1}^{n_h}\tilde{\theta}_{hj} = \hat{\theta} + \frac{n_h-1}{n}\sum_{h=1}^{L}\sum_{j=1}^{n_h}\hat{\theta}_{-hj} \tag{5.76}$$

The corresponding jack-knifed variance estimator could be obtained as

$$\hat{V}_{J_3}(\tilde{\theta}_{J_1}) = \sum_{h=1}^{L}\sum_{j=1}^{n_h}(\tilde{\theta}_{hj}-\tilde{\theta}_{J_1})^2/n_h(n_h-1) \ . \tag{5.77}$$

An alternative jack-knife estimator may be obtained by averaging over L strata, that is

$$\tilde{\theta}_{J_2} = \frac{1}{L}\sum_{h=1}^{L}\frac{1}{n_h}\sum_{j=1}^{n_h}\tilde{\theta}_{hj} \tag{5.78}$$

The variance estimator in this case is

$$\hat{V}_{J_4}(\tilde{\theta}_{J_2}) = \sum_{h=1}^{L}\frac{1}{n_h}\sum_{j=1}^{n_h}(\tilde{\theta}_{hj}-\tilde{\theta}_{J_2})^2/(n_h-1) \tag{5.79}$$

These estimators are the extensions of the jack-knife estimator proposed by McCarthy for the special case $n_h=2$. Folsom, Bayless and Shah (1971) have also considered the jack-knifed estimator for stratified sampling. Their jack-knife variance estimator can be obtained by replacing $\tilde{\theta}_{J1}$ in (5.77) by $\sum_{j=1}^{n_h}\tilde{\theta}_{hj}/n_h$ and is equivalent to Jones variance estimator (5.74).

We illustrate the above discussion through an example of ratio estimation from Rao (1975).

Example 5.1

Consider a stratified simple random sample design. Let n_h units be selected from the N_h units in stratum $h(h=1,2,\ldots,L)$. The combined ratio estimator of the population ratio R is

$$\hat{R} = \frac{\sum_{h=1}^{L} W_h \bar{y}_h}{\sum_{h=1}^{L} W_h \bar{x}_h} \tag{5.80}$$

where $W_h = N_h/N$ and $\bar{y}_h$ and $\bar{x}_h$ are stratum sample means. Let $\hat{R}_{(hj)}$ denote the estimator of R obtained by deleting the jth unit in the stratum h and $\hat{R}_{(h)} = \sum_{j=1}^{n_h} \hat{R}_{(hj)}/n_h$. Then Jones estimator (5.73) for R is given by

$$\hat{R}_J = \left[1+ \sum_{h=1}^{L} (n_h-1)\left(1-\frac{n_h}{N_h}\right)\right]\hat{R} - \sum_{h=1}^{L} (n_h-1)\left(1- \frac{n_h}{N_h}\right)\hat{R}_{(h)} \ . \tag{5.81}$$

The estimate of the $MSE(\hat{R})$ is

$$\hat{V}_J(\hat{R}) = \sum_{h=1}^{L} \left(1- \frac{n_h}{N_h}\right)\frac{1}{n_h} s^2_{(h)} \ , \tag{5.82}$$

where

$$s^2_{(h)} = \sum_{j=1}^{n_h} (\hat{R}_{(hj)}-\hat{R}_{(h)})^2/(n_h-1) \ . \tag{5.83}$$

For $n_h = 2$, $1 - \frac{n_h}{N_h} \doteq 1$ and (5.81) reduces to $\hat{R}_J = 2\hat{R} - \sum_{h=1}^{L} \hat{R}_{(h)}$ which was proposed by McCarthy (1966). The variance estimator (5.82) reduces to

$$\hat{V}_J(\hat{R}) = \sum_{h=1}^{L} [\tfrac{1}{2} \sum_{j=1}^{2} (\hat{R}_{(hj)}-\hat{R}_{(h)})^2] = \sum_{h=1}^{L} [\tfrac{1}{4} (\hat{R}_{(h1)}-\hat{R}_{(h2)})^2] \tag{5.84}$$

Similarly, by making appropriate substitutions we can obtain other variance estimators.

Case II

Brillinger (1966) proposed n mutually exclusive and exhaustive groups each containing one observation per stratum and computed the jack-knife estimates by excluding an entire group. He considered $n_h = rm_h$ where r, the number of groups remains fixed. Assuming $f_h = \frac{n_h}{N_h}$ does not change too much with r, he suggests that the sample elements in the h^{th} stratum should be randomly split into r groups of size m_h and then randomly join together a group from each stratum to form r global groups. Brillinger also discusses the jack-knifing of two stage sampling and cluster sampling with unequal cluster sizes. He demonstrates how complicated samples should be divided into groups in order to obtain an unbiased estimate of variance. However, the method adopted by Brillinger is like a part sample technique and does not have the characteristics of randomisation or balancing. Moreover, we have seen earlier in subsection 5.3.1.3 that $r = n$ is the optimum choice whereas Brillinger assumes fixed r.

Case III

Recall that McCarthy applied the jack-knife to stratified sampling by leaving one observation per stratum whereas Kish and Frankel's jack-knifing consists in deleting one PSU per stratum and adding twice the stratum's other primary selection. Lemeshow and Epp (1977) discuss the jack-knifing of individuals or groups by collapsing the strata. Suppose we wish to estimate the variance of the population mean from a stratified simple random sample with n_h observations per stratum. Then the estimates of the population mean may be formed by

omitting a single y_{hj} and adding back the average of those observations remaining in the same stratum. In this way we compute $n_h.L$ values,

$$\hat{\theta}_{hj} = \hat{\theta} - W_h \frac{y_{hj}}{n_h} + W_h \frac{\sum_{j' \neq j} y_{hj'}}{n_h(n_{h-1})} \tag{5.85}$$

The pseudovalues are now given by

$$\tilde{\theta}_{hj} = n_h \hat{\theta} - (n_h - 1)\hat{\theta}_{hj} \tag{5.86}$$

The average of these values is the jack-knife estimator,

$$\tilde{\theta}_L = \frac{1}{L} \sum_{h=1}^{L} \frac{1}{n_h} \sum_{j=1}^{n_h} \tilde{\theta}_{hj} \tag{5.87}$$

and the variance estimate is

$$\hat{V}_J(\tilde{\theta}_L) = \sum_{h=1}^{L} \sum_{j=1}^{n_h} (\tilde{\theta}_{hj} - \tilde{\theta}_L)^2 / n_h(n_h - 1) \tag{5.88}$$

These authors discuss 4 variant cases. Suppose we have L strata, then the four cases considered are:

Case 1

$n_h = km_h$ observations in each stratum which are collapsed into k equal sized groups of $m_h = n_h/k$. Instead of omitting a single observation, an entire group is omitted to obtain the jack-knife variance estimate.

Let

$$y^*_{h1} = \sum_{h=1}^{m_h} y_{hj}, \; y^*_{h2} = \sum_{h=1}^{2m_h} y_{hj}, \ldots, y_{hk} = \sum_{h=(k-1)m_h+1} y_{hj} \tag{5.89}$$

Then

$$\hat{\theta} = \sum_{h=1}^{L} W_h \sum_{j=1}^{k} y^*_{hj}/km_h \ , \tag{5.90}$$

and

$$\hat{\theta}_{hj} = \hat{\theta} - W_h \frac{y^*_h}{km_h} + W_h \frac{\sum_{j' \neq j} y^{*'}_{hj}}{k(k-1)m_h} \tag{5.91}$$

The pseudovalues and the jack-knifed variance estimate are obtained as in (5.86-5.88). Denoting $Z_{hj} = W_h y^*_{hj}/km_h$ in expressions (5.90), the variance estimate is given by

$$\hat{V}_J(\tilde{\theta}_{L1}) = k^2 \sum_{h=1}^{L} \sum_{j=1}^{k} (Z_{hj} - \bar{Z}_h)^2/k(k-1)$$

$$= \frac{1}{k(k-1)} \sum_{h=1}^{L} \left(\frac{W_h}{m_h}\right)^2 \sum_{j=1}^{k} (y^*_{hj} - \bar{y}^*_h)^2 \ , \tag{5.92}$$

where

$$\tilde{\theta}_{L1} = \hat{\theta} + (k-1)Z_{hj} - \sum_{j' \neq j} Z_{hj}' \quad \text{and} \quad \bar{y}^*_h = \frac{1}{k} \sum_{j=1}^{k} y^*_{hj}$$

Case 2

If $k = n_h$ and $m_h = 1$, then from (5.92), we get

$$\hat{V}_J(\tilde{\theta}_{L1}) = \sum_{h=1}^{L} W_h^2 \sum_{j=1}^{n_h} (y^*_{hj} - \bar{y}^*_h)^2/n_h(n_h-1) \tag{5.93}$$

Case 3

If we take $k = 2$, i.e., $n_h = 2m_h$, then from (5.92) we have

$$\hat{V}_J(\tilde{\theta}_{L1}) = \sum_{h=1}^{L} \left(\frac{W_h}{m_h}\right)^2 \sum_{j=1}^{2} (y^*_{hj} - \bar{y}^*_h)^2/2 \tag{5.94}$$

Case 4

When $k = 2$ and $m_h = 1$, then from (5.92)

$$\hat{V}_J(\tilde{\theta}_{L1}) = \sum_{h=1}^{L} W_h^2 \sum_{j=1}^{2} (y_{hj}-\bar{y}_h)^2$$

$$= \frac{1}{4} \sum_{h=1}^{L} W_h^2 (y_{h1}-y_{h2})^2$$

$$= \hat{V}(\bar{y}_{st}) \text{ , given by (5.64)} \qquad (5.95)$$

5.3.1.5b Jack-knifing for Variance Components in Complex Surveys

The jack-knife has also been extended for variance components in complex surveys. The details on this aspect could be seen in Folsom, et al., (1971). These authors have also developed a multistage extension of the jack-knife for non-linear statistics from complex surveys.

Consider a stratified three stage design with PPS selection at the first two stages and equal probability sampling at the last stage. It is assumed that first stage units are sampled with replacement and are subsampled independently each time they are selected. Second and third stage units are selected without replacement.

Let $\hat{\theta}$ be the non-linear statistics, then in the notation of Folsom, et al.,

$$\hat{\theta} = f[y_{+\ldots}(1), \ldots . y_{+\ldots}(g)] \qquad (5.96)$$

where the (plus) in the y's (sample totals) indicates the summation over $h(h = 1,\ldots,H)$ strata. The estimates for θ are formed from pseudo-replicates obtained by successively deleting the data from sampling units at a particular level of the design. These estimates are

$$J\theta_{hijk} = n_h s_{hi} m_{hij} \hat{\theta} - (n_h-1)\hat{\theta}_{-hi} - n_h(s_{hi}-1)\hat{\theta}_{-hij} - n_h s_{hi}(m_{hij}-1)\hat{\theta}_{-hijk} \tag{5.97}$$

When a first stage unit is deleted, the replicate estimator is

$$\hat{\theta}_{-hi} = f\{Y_{-hi}(1),.....Y_{-hi}(y)\} \tag{5.98}$$

with

$$Y_{-hi}(r) = y_{+}...(r) - [y_{hi..}(r) - y_{h...}(r)]/(n_h-1) \tag{5.99}$$

The jack-knife estimate is the average of the pseudo-values and is given by

$$\hat{\theta}_{JK} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{s_{hi}} \sum_{k=1}^{m_{hij}} J\theta_{hijk} / Hn_h s_{hi} m_{hij} \tag{5.100}$$

The variance estimate for the first stage component in (5.100) is

$$J \sum_{p}^{2} = (JMS_p - JMS'_{s/p}) \tag{5.101}$$

where

$$JMS_p = \sum_{h=1}^{H} \sum_{i=1}^{n_h} (J\theta_{hi..} - J\theta_{h...})^2/(n_h-1)$$

$$JMS'_{S/p} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{s_{hi}} (J\theta_{hij.} - J\theta_{hi..})^2/n_h s_{hi}(s_{hi}-1)$$

5.4 Robust Interval Estimation

Tukey (1958) suggested that the pseudo values $\tilde{\theta}_i$ could be treated as n independent, identically distributed random variables with mean θ and hence standard error of $\tilde{\theta}$ can be estimated and could be used to set up approximate confidence intervals. Thus the statistic

$$\frac{\sqrt{n}(\tilde{\theta}-\theta)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\tilde{\theta}_i-\tilde{\theta})^2}} \qquad (5.102)$$

has an approximate t distribution with (n-1) degrees of freedom. The approximate confidence intervals of the form $\tilde{\theta} \pm S.E.(\tilde{\theta})\ t_{n-1}^{\alpha/2}$ could be constructed, where $t_{n-1}^{\alpha/2}$ is 100 $(1-\alpha/2)$ per cent point of Student's t-distribution.

However, with exception of Brillinger (1964), most of the literature on the distribution theory of the jack-knife is concerned with establishing the asymptotic normality. The reason for this is that, in practice, the jack-knife t-intervals are conservative (i.e. wider than necessary for the nominally listed coverage) so it may be better to use the normal critical constant. Miller (1964, 68) has discussed under what circumstances Tukey's assertion regarding the use of the functions of the pseudovalues to estimate the variance and set confidence intervals are valid. Let $\theta = f(\mu)$ where $\mu = E(X_i)$. Then Miller (1964) showed that if $\{X_i\}$ is a sequence of independent and identically distributed random variables with mean $\mu = 0$ and variance $0 < \sigma^2 < \infty$, and if f is a function on the real line which in the neighbourhood of the origin has a bounded second derivative, then as $n \to \infty$, $\sqrt{n}(\tilde{\theta}-\theta)$ is asymptotically normally distributed with mean zero, and variance $\sigma^2(f'(0))^2$ and that $(n-1)^{-1}\sum_{i=1}^{n}(\tilde{\theta}_i-\tilde{\theta})^2 \xrightarrow{P} \sigma^2(f'(0))^2$. In a subsequent paper, Miller (1968) extended his results to potentially more useful case, where $\hat{\theta}_n$ is the sample variance or the log of the sample variance.

The conditions under which the original statistic $\hat{\theta}_n$ and the jack-knifed statistic $\tilde{\theta}_n$ have the same asymptotic properties have been described by Thorburn(1976). In his subsequent paper, Thorburn (1977)

discusses that if the pseudo-values $\hat{\theta}_i$ converge in mean square to a sequence of random variables X_i, as $n \to \infty$, then random variables X_i are independently and identically distributed. He shows that the variance estimate based on the pseudovalues is consistent and that the jack-knife estimator is asymptotically normal.

Miller (1964) also pointed out the situations where the jack-knife procedure fails. His three counter examples are based on the simplest non-linear estimate, that is, the maximum of $X_1,\ldots,X_n$. He demonstrated that for the largest order statistic, the asymptotic behaviour of (5.102) can be non-normal, degenerate and may even drift to infinity.

5.4.1 Functions of Maximum Likelihood Estimate

Brillinger (1964) presents a different limiting approach by holding g fixed and letting $m \to \infty$ for the grouped jack-knife (5.4). Let $y_1,y_2,\ldots,y_n$ be a random sample from a distribution with density function $f(y,\theta)$ then the likelihood function is given by

$$L(y_1,y_2,\ldots,\theta) = \prod_{i=1}^{n} f(y_i;\theta) \tag{5.103}$$

$$Log(y,\theta) = \sum_{i=1}^{n} \log f(y_i;\theta) \tag{5.104}$$

$$\frac{\partial \log L(y,\theta)}{\partial \theta} = \sum_{i=1}^{n} \partial \log f(y_i;\theta)/\partial\theta \tag{5.105}$$

Under the regularity conditions, the maximum likelihood estimate $\hat{\theta}$ of θ based on n observations satisfies the condition,

$$\sum_{i=1}^{n} \frac{\partial \log f(y_i,\theta)}{\partial \theta}\bigg|_{\theta = \hat{\theta}} = 0 \tag{5.106}$$

Now expanding (5.106) in Taylor series and making appeal to strong law of large number and central limit theorem, after some algebraic manipulations, it could be shown that

$$\sqrt{n}\ (\hat{\theta}-\theta) \xrightarrow{d} N(0,1/I)\ , \tag{5.107}$$

where

$$I = -E\left[\frac{\partial^2 \log f(y,\theta)}{\partial\theta^2}\right] \quad \text{is the Fisher's information.}$$

Let $\hat{\theta}_{(j)}$ denote the maximum likelihood estimate derived from $(g-1)m$ observations obtained by deleting the jth group, then the grouped jack-knife can be defined as

$$\tilde{\theta}_{pj} = g\hat{\theta} - (g-1)\hat{\theta}_{(j)}\ , \quad j = 1,2,\ldots,g \tag{5.108}$$

Now, holding g fixed and letting $n \to \infty$, it could be shown by an appeal to central limit theorem that

$$\sqrt{n}(\tilde{\theta}_{pj}-\theta) \xrightarrow{d} N(0,1/I)$$

Also using results from Mann and Wald (1943), after some simplifications it follows that

$$\frac{\sqrt{g}(\tilde{\theta}_{p}-\theta)}{\sqrt{\sum_{j=1}^{g}(\tilde{\theta}_{pj}-\theta)^2/(g-1)}} \overset{d}{\sim} t(g-1) \tag{5.109}$$

This implies that the t-test can be applied to $\tilde{\theta}_{pj}$ as if they were independently distributed.

We note the main difference between Brillinger approach and that of Miller is that the former regards the number of groups to be held finite as $n \to \infty$ and letting the size of group become larger whereas the

latter assumes $g = n$. However, one would prefer to have $g >> m$, so the asymptotics do not establish the t-approximation in many problems. Also it has been shown (Section 5.3) that $g = n$ is optimal. Fryer (1970), in an abstract, reports to have overcome the difficulty that g remains finite as $n \to \infty$. He also claims to have extended Brillinger's results to the multiparameter case but these results have not yet been published. The question thus arises regarding the relative size of g and m , if n is large. However, at present the only solution appears to be that g should be large enough so that adequate degrees of freedom are available for t - statistics.

Reeds (1978) has proved the asymptotic normality for $\tilde{\theta}_g$ when $g \to \infty$, with $m = 1$. He shows that the jack-knife gives the correct asymptotic variance for $\tilde{\theta}_g$ and $\hat{\theta}$ even if the model is incorrect. Fisher's information does not do this because it is computed theoretically on the basis of assumed density $f(y,\theta)$. Brillinger (1977) considers the version of jack-knife, as considered by Miller and Reeds, etc and obtains the results which are very close to Reed's results. To obtain these results, Brillinger uses Chibisov's (1973) expansion in order to prove convergence. The appeal to Chibisov's requires much stronger moment and smoothness conditions than the Cramer's conditions used by Reeds. Reeds uses a reversion of series specially fitted to the jack-knife application instead of a general purpose result like Chibisov's. Thus, on the basis of the above discussion and the earlier results that $g = n$ is the optimum choice (Section 5.3) Reed's results appear to be preferable.

5.4.2 U-Statistic and Functions of U-Statistic

The results of Miller (1964) were extended to U-statistic and functions of several U-statistic by Arvesen, (1969). Any statistic of the form

$$U(X_1,X_2,\ldots,X_n) = \binom{n}{\ell}^{-1} \sum_{C} f^*(X_{i1},X_{12},\ldots,X_{i\ell}) \ , \qquad (5.110)$$

where the kernel function $f^*(x_1,\ldots,x_\ell)$ is symmetric in its ℓ arguments and the summation is over all combinations of ℓ variables $X_{i1},X_{12},\ldots,X_{i\ell}$ out of n variables $X_1,\ldots,X_n$, is termed as U-statistic.

Let the parameter of interest be denoted by $f(\mu)$ and the estimator to be jack-knifed by $\hat{\theta} = f(U)$, where $\mu = E[f^*(X_1,\ldots,X_\ell)]$. Then considering $(X_1,\ldots,X_n)$ as independent and identically distributed random varaibles and $f^*(X_1,\ldots,X_\ell)$ as any real valued symmetric statistic such that $E[f^*(X_1,\ldots,X_\ell)]^2 < \infty$, where ℓ is the smallest number of observations needed to estimate θ, Arvesen shows that the limiting distribution of (5.102) is unit normal as $n \to \infty$, provided f has a bounded second derivative near μ. These results have also been extended to functions of several U-statistics $f^*(U_1,U_2,\ldots,U_r)$, where each U-statistic has U_j different kernel functions $f^*(j)$ for the same set of independent, identically distributed random varialbles $(X_1,\ldots,X_n)$. Arvesen also discusses the case of non-identically distributed random variables and applications to ratio-estimators and product moment correlation.

Considering the X's as independent and identically distributed random variables and expressing the unjack-knifed statistic as a function of several U-statistic, Arvesen and Layard (1975) established the asymptotic normality and the consistency of the jack-knife estimator for

sampling with replacement. In case of sampling without replacement the X's are no longer independent. Utilizing the concept of a sequence of finite populations (Madow, 1948; Hajek, 1960), Nandi and Sen (1963) established the consistency and asymptotic normality of U-statistic in sampling without replacement from finite populations. Recently, Krewski (1978a) shows that many non-linear statistics can be expressed as functions of several U-statistics. He also establishes the asymptotic normality of non-linear statistic and the consistency of the corresponding linearization variance estimator.

5.4.3 Functions of Regression Estimates

We have noted from the previous sections that the applications of the jack-knife technique are mainly confined to handle the balanced situations with independent and identically distributed random variables. Miller (1964, 1968,1974a),Thorburn (1976, 1977) and others have established the conditions under which the jack-knife estimator is asymptotically normally distributed with a consistent estimate of its variance.

Miller (1974a) extended the applicability of the jack-knife technique to the linear model,

$$\underline{Y} = \underline{X}\beta + \underline{e} \tag{5.111}$$

where $\underline{Y} = (Y_1,\ldots,Y_n)$ is a vector of n independent observations $\underline{\beta} = (\beta_1,\beta_2,\ldots\beta_p)$ is a vector of unknown parameters, $\underline{e} = (e_1,e_2,\ldots,e_n)$ is a vector of independent and identically distributed error variables with $E(e_i) = 0$, $Var(e_i) = \sigma^2$, $E(e_i^4) < \infty$. The $\underline{X}$ is a matrix of known values.

The parameter of interest are of the form $\theta = f(\underline{\beta})$, where f is a smooth function of the model parameters. The estimator considered is

$\hat{\underline{\theta}} = f(\hat{\underline{\beta}})$, where $\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$ is the least squares estimator. The jack-knife is applied in the usual way be deleting successively each row of $\underline{X}$ matrix and each column of $\underline{Y}$. Let $\hat{\underline{\beta}}_{-i}$ be the least squres estimator for $\underline{\beta}$ obtained by deleting the ith row of $\underline{X}$ matrix and the ith element of $\underline{Y}$. The pseudovalues are defined as

$$\tilde{\theta}_i = nf(\hat{\underline{\beta}}) - (n-1)f(\hat{\underline{\beta}}_{-i}) \tag{5.112}$$

and the jack-knifed estimator for θ is

$$\tilde{\theta} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i \tag{5.113}$$

Miller (1974a) shows that under the conditions that $\frac{\underline{X}'\underline{X}}{n} \to \underline{\Sigma}$, a positive definite matrix, as $n \to \infty$ the jack-knife estimator of θ is asymptotically normally distributed, that is,

$$\sqrt{n}\,(\tilde{\theta}-\theta) \xrightarrow{d} N(0,\sigma^2 f'(\underline{\beta})^T \underline{\Sigma}^{-1} f'(\underline{\beta}))$$

for any function $f(.)$ with bounded second derivatives in an open sphere about $\underline{\beta}$, where $f'(\underline{\beta}) = \left(\frac{\partial f(\underline{\beta})}{\partial \beta_1}, \ldots, \frac{\partial f(\underline{\beta})}{\partial \beta_p}\right)^T$. The pseudovalues (5.111) are defined symmetrically with respect to the observations $(\underline{x}_i, Y_i)$ but in general linear model situations the observations do not affect the least squares estimates in the same way.

Hinkley (1977b) suggested a weighted jack-knifed estimator $\tilde{\theta}_W$ with the assocaited variance estimator $\tilde{V}_{WJ}$ (see Chapter 3). Under the similar conditions as given by Miller (1974a), the weighted jack-knifed estimator $\tilde{\theta}_W$ of θ is also asymptotically normally distributed, that is, if $\sigma^2 < \infty$ and there is some positive definite matrix $\underline{\Sigma}$ such that $n^{-1}(\underline{X}'\underline{X}) \to \underline{\Sigma}$, as $n \to \infty$, then $\sqrt{n}(\tilde{\theta}_W-\theta)$ converges in distribution to $N(0, \sigma^2 f'(\underline{\beta})^T \underline{\Sigma}^{-1} f'(\underline{\beta}))$. Also if f has a continuous

first derivative in a neighbourhood of $\underline{\beta}$, then $\hat{V}_J$ and $\hat{V}_{WJ}$ converge in probability to $\sigma^2 f'(\underline{\beta})^T \underline{\Sigma}^{-1} f'(\underline{\beta})$ as $n \to \infty$. These results extend the use of jack-knife to the estimation of linear or nonlinear $f(\underline{\beta})$ in unbalanced situations.

5.4.4 Correlation Coefficient

Another area where the normal theory procedure is not robust is the interval estimation for the correlation coefficient. The test that $\rho = 0$ is robust to non-normality but for $\rho \neq 0$, the asymptotic variance of $\hat{\theta} = \tanh^{-1} r = \frac{1}{2} \ln(\frac{1+r}{1-r})$ is not $\frac{1}{n-3}$ unless the underlying distribution is normal. Duncan and Layard (1973) studied jack-knifing $\hat{\theta}$ and found that it works well with most of the distributions considered, namely, normal, lognormal, linear regression uniform, linear regression double exponential, gamma and contaminated normal. The normal theory test showed poor results for certain non-normal distributions whereas the jack-knife method provided better confidence intervals for most of the sampled distributions.

Recent work on improving the jack-knife in connection with the correlation is due to Hinkley (1977a, 1978). Hinkley (1977a) assumes $n = gm$ and on the basis of a simulation study based on 500 replicates for normal, contaminated normal and double exponential, he shows that $T_n = \tanh^{-1} r$ has not only stable variance for elliptically symmetric distributions but, to second order, it is symmetrically distributed. Hinkley concludes that the approximate normal jack-knife confidence limits are often inaccurate even for moderately large n and the student t approximation with $(g-1)$ degrees of freedom is better for large m. The grouping of data usually improves the probability accuracy of confidence limits with only a slight loss of efficiency. However,

random grouping of data is objectionable due to the reason that the resulting inference is not unique. Moreover the pseudovalues are not capable of indicating deviant points.

Hinkley (1978) uses z transformation of the sample correlation estimate r and argues that this transformation is preferable in practice because it is more nearly normal than r and because inadmissible values are possible for the jack-knifed untransformed correlation. His results based on the artificial data show that the jack-knifed estimator is more seriously affected by a deviant point than the unjack-knifed estimator. Similar behaviour of the jack-knife have earlier been observed by Miller (1974b) and Wainer and Thissen (1975). Using well known expansions for differentiable statistical functions (von Mises, 1947), Hinkley suggests robust alternatives based on the analysis of pseudovalues. He concludes that the routine deletion of smallest and largest pseudovalues before averaging produces a very definite improvement over T_n^* in case of correlation, where $T_n^* = \frac{1}{n}\sum_{i=1}^{n} P_{n,i}$, $P_{n,i} = nT_n - (n-1)T_{n,-i}(i=1, 2,\ldots,n)$ and $T_n = t(Y_1,Y_2,\ldots,Y_n)$ is an estimate of the parameter of interest θ. Knott and Frangos (1983) pointed out that the results of Hinkley (1978), on variances associated with jack-knife estimation are incomplete due to the non-inclusion of third order terms in a von Mises expansion. Thus the estimators for the variances in Hinkley (1978) are not adequate for moderate sample sizes and can be improved. These authors give Monte Carlo simulation results in support of their conclusions.

5.5 Concluding Remarks

This brief review has had to be selective, with some issues excluded, but some final points should be made. We have attempted to discuss the important aspects (particularly relevant to survey sampling)

that has been significant in bringing the jack-knife method to its present status. Presumably one could identify two types of inference problems, namely, bias reduction and interval estimation where the jack-knife has proved itself to be a useful device. We observe that the jack-knife method introduced originally to reduce bias in estimation was progressively extended as a means of obtaining robust confidence intervals.

Considering the theoretical properties and numerical studies, discussed in this chapter, we reach the following conclusions.

(1) The jack-knife appears to behave reliably both as a means of bias reduction and of providing a realistic assessment of error in cases where the estimator is a function of all data values concerned. This includes situations involving variances, maximum likelihood or least squares estimator, regression, correlation and ratios. However, the order statstistic does not come under this category.

(2) We also note that the jack-knife method can be extended to more complicated samples, provided adequate care is taken to ensure that a properly representeative method is used for subdividing the sample.

(3) Based on the mean square error considerations in ratio and other problems, the choice $g = n$ appears to be the best.

(4) The theoretical and numerical results on ratio estimation provide no clear picture about the superiority of the jack-knifed ratio estimator and the estimator of its variance as compared to other competing estimators under the randomisation based approach.

(5) The model based approach does point out flaws in the conventional variance estimator $\hat{V}_{1C}$ and recommends that balance is a necessary

condition for robust inference. However, even under this approach, the empirical results of Royall and Cumberland (1982, 1983) regarding the performance of robust variance estimators for restricted random sampling, even in case of balanced samples, for robust interval estimation using standard normal approximation are quite confusing and do not differ from those for $\hat{V}_{1C}$.

Before closing the discussion we would like to make a final point that it has been noted that the standard jack-knife procedure is sensitive to deviant data points (see Chapter 3, Section 3.2.2.11). Thus an outlying pseudovalue indicates that an estimate is being excessively influenced by the corresponding observation (or groups of observations). This suggests that there is need for considering alternative more robust methods, as discussed in subsequent chapters.

CHAPTER 6

THE UNWEIGHTED JACK-KNIFE FOR REGRESSION ESTIMATION

6.1 Introduction

During the last three decades considerable attention has been devoted to studying the properties of the jack-knife method. We find that the technique has since been extended to handle more general forms of applications such as estimation of variances, ratios and correlation, etc. The conditions under which the jack-knife estimator is asymptotically normally distributed together with a consistent estimate of its variance have also been established. As could be seen from Chapter 5, a considerable amount of literature also exists on the performance of the jack-knife in problems of ratio estimation and comparison of variances under the randomisation based and the model based approaches. The jack-knife has also been generalised to tackle more general forms of bias and to estimation in specialised stochastic processes. The bulk of this literature deals with balanced cases which assume equal sample sizes and equal variances. We have reviewed the relevant literature on the jack-knife in Chapter 5.

Linear regression models are of considerable importance in a variety of applications. Miller (1974a) extended the jack-knife technique to the estimation of the parameters associated with the linear regression model $\underline{Y} = \underline{X}\underline{\beta} + \underline{e}$, where $\underline{Y}$ is a vector of n indepedent observations, $\underline{X}$ is a matrix of known values, $\underline{\beta}$ is a vector of unknown parameters and $\underline{e}$ is a vector of independent and identically distributed error variables.

In this chapter, we discuss the application of the unweighted jack-knife procedure to regression estimation. In Section 6.2 we give basic concepts necessary for later developments in this and the subsequent chapters. In multiple regression an important set of quantities is $w_i = \underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i$ which measures the distance of a single design point from the centre of the data. In Section 6.3, we discuss the interpretation of w_i as a distance measure. Section 6.4 is devoted to the properties of the jack-knifed regression estimator $\tilde{\beta}$ under the linear model described above. The important feature of the jack-knife is that it provides a simple variance estimator for the variance of a parameter estimator. In Section 6.5, we examine the properties of the jack-knife variance estimator. Finally, in Section 6.6 we give the main conclusions.

6.2 Preliminary Concepts

In this section we give some basic concepts which will be utilized for derivations in the subsequent developments.

Lemma 6.1

Let $\underline{A}$ be a non-singular $n \times n$ matrix and $\underline{U}$ and $\underline{V}$ be two $n \times 1$ column vectors, then

$$(\underline{A} + \underline{U}\underline{V}^T)^{-1} = \underline{A}^{-1} - \frac{(\underline{A}^{-1}\underline{U})(\underline{V}^T\underline{A}^{-1})}{1+\underline{V}^T\underline{A}^{-1}\underline{U}} \tag{6.1}$$

This is a standard result in matrix algebra, see for example, Rao (1965), and Draper and Smith (1981).

Corollary 6.1

If $\underline{X}$ is a $P \times P$ non-singular matrix, $\underline{Y}$ is a $P \times 1$ vector and C is a scalar, then

$$(\underline{X} + C\underline{Y}\underline{Y}^T)^{-1} = \underline{X}^{-1} - \frac{C\underline{X}^{-1}\underline{Y}\underline{Y}^T\underline{X}^{-1}}{1+C\underline{Y}^T\underline{X}^{-1}\underline{Y}} \tag{6.2}$$

Lemma 6.2 (Miller 1974a)

For fixed $\underline{x}$ the sequence $\underline{x}^T(\underline{X}_n^T\underline{X}_n)^{-1}\underline{x}$ is non-increasing as n increases.

Proof

We can write $\underline{X}_{n+1}^T\underline{X}_{n+1} = \sum_{i=1}^{n} \underline{x}_i^T\underline{x}_i = \underline{X}_n^T\underline{X}_n + \underline{x}_{n+1}^T\underline{x}_{n+1}$. By Lemma 6.1, we have

$$(\underline{X}_{n+1}^T\underline{X}_{n+1})^{-1} = (\underline{X}_n^T\underline{X}_n + \underline{x}_{n+1}^T\underline{x}_{n+1})^{-1}$$

$$= (\underline{X}_n^T\underline{X}_n)^{-1} - \frac{(\underline{X}_n^T\underline{X}_n)^{-1}\underline{x}_{n+1}^T\underline{x}_{n+1}(\underline{X}_n^T\underline{X}_n)^{-1}}{1+\underline{x}_{n+1}(\underline{X}_n^T\underline{X}_n)^{-1}\underline{x}_{n+1}^T}$$

Thus

$$\underline{x}^T(\underline{X}_{n+1}^T\underline{X}_{n+1})^{-1}\underline{x} = \underline{x}^T(\underline{X}_n^T\underline{X}_n)\underline{x} - \frac{(\underline{x}^T(\underline{X}_n^T\underline{X}_n)^{-1}\underline{x}_{n+1}^T)(\underline{x}_{n+1}(\underline{X}_n^T\underline{X}_n)^{-1}\underline{x})}{1+\underline{x}_{n+1}(\underline{X}_n^T\underline{X}_n)^{-1}\underline{x}_{n+1}^T} \tag{6.3}$$

Since the numerator and the denominator of the last term are non-negative and greater than or equal to one, respectively, the conclusion follows.

Lemma 6.3

If $\underline{X}^T\underline{X}/n \to \Sigma$ positive definite, then

$$\max_{1\le i\le n} \underline{x}_i(\underline{X}_n^T\underline{X}_n)^{-1}\underline{x}_i^T \to 0$$

Proof

Consider a sequence ℓ for which $\max \to \varepsilon > 0$

Let $$\underline{x}_{i_\ell}(\underline{X}_\ell^T\underline{X}_\ell)^{-1}\underline{x}_{i_\ell} = \max_{1\le i\le \ell} \underline{x}_i(\underline{X}_\ell^T\underline{X}_\ell)^{-1}\underline{x}_i^T \tag{6.4}$$

Since $\ell(\underline{X}_\ell^T\underline{X}_\ell)^{-1} \to \Sigma^{-1}$, the sequence i_ℓ must tend to infinity.

Now by Lemma 6.2, $\underline{x}_{i_\ell}(\underline{X}_{i_\ell}^T \underline{X}_{i_\ell})\underline{x}_{i_\ell}^T \geq \underline{x}_{i_\ell}^T (\underline{X}_\ell^T \underline{X}_\ell)^{-1} \underline{x}_{i_\ell}^T$. But we have,

$$P = \mathrm{Tr}(\underline{X}_{i_\ell}^T \underline{X}_{i_\ell})^{-1}(\underline{X}_{i_{\ell-1}}^T \underline{X}_{i_{\ell-1}} + \underline{x}_{i_\ell}^T \underline{x}_{i_\ell})$$

$$= \mathrm{Tr}(\underline{X}_{i_\ell}^T \underline{X}_{i_\ell})^{-1}(\underline{X}_{i_{\ell-1}}^T \underline{X}_{i_{\ell-1}}) + \underline{x}_{i_\ell}(\underline{X}_{i_\ell}^T \underline{X}_{i_\ell})^{-1}\underline{x}_{i_\ell}^T \qquad (6.5)$$

Now since, $i_\ell \to \infty$ and $(\underline{X}_n^T \underline{X}_n)/n \to \underline{\Sigma}$, as $n \to \infty$ it implies that

$$\mathrm{Tr}(\underline{X}_{i_\ell}^T \underline{X}_{i_\ell})^{-1}(\underline{X}_{i_{\ell-1}}^T \underline{X}_{i_{\ell-1}}) \to P \qquad (6.6)$$

Thus from (6.5) we get $\underline{x}_{i_\ell}(\underline{X}_{i_\ell}^T \underline{X}_{i_\ell})^{-1}\underline{x}_{i_\ell}^T \to 0$ which is a contradiction. Hence,

$$\max_{1\leq i\leq n} \underline{x}_i (\underline{X}_n^T \underline{X}_n)^{-1} \underline{x}_i^T \to 0 \quad \text{as} \quad \frac{(\underline{X}^T \underline{X}_n)}{n} \to \underline{\Sigma}(\text{p.d.}). \qquad (6.7)$$

Lemma 6.4

Consider the linear regression model,

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{e} \qquad (6.8)$$

where $\underline{Y}^T = (Y_1, Y_2, \ldots, Y_n)$, $\underline{\beta}^T = (\beta_1, \beta_2, \ldots, \beta_p)$, $\underline{e}^T = (e_1, e_2, \ldots, e_n)$ and

$$\underline{X} = \begin{pmatrix} x_{11}, x_{12}, \ldots, x_{1p} \\ x_{21}, x_{22}, \ldots, x_{2p} \\ \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \\ x_{n1}, x_{n2}, \ldots, x_{np} \end{pmatrix} = \begin{pmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_n^T \end{pmatrix} .$$

The error variables e_i, $i = 1,2,\ldots,n$ are assumed to be independently and identically distributed with zero mean and constant variance σ^2. Let $\hat{\underline{\beta}}$ be the least squares estimator of $\underline{\beta}$ with the complete data set and $\hat{\underline{\beta}}_{-i}, i = 1,2,\ldots,n$ be the least squres estimator of $\underline{\beta}$ when

the i^{th} row $(\underline{x}_i^T, Y_i)$ is deleted. Further let $\tilde{\underline{X}}$ be the matrix $\underline{X}$ with the ith row deleted and $\tilde{\underline{Y}}$ represent the vector $\underline{Y}$ with ith row deleted. Then,

$$\hat{\underline{\beta}} - \hat{\underline{\beta}}_{-i} = \frac{(\underline{X}^T\underline{X})\underline{x}_i(Y_i - \underline{x}_i^T\hat{\underline{\beta}})}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i} \tag{6.9}$$

Proof

We can write $\tilde{\underline{X}}^T\tilde{\underline{X}} = (\underline{X}^T\underline{X} - \underline{x}_i\underline{x}_i^T)$ where $\underline{x}_i^T$ is the ith row of $\underline{X}$. By Lemma 6.1, we get

$$(\tilde{\underline{X}}^T\tilde{\underline{X}})^{-1} = (\underline{X}^T\underline{X} - \underline{x}_i\underline{x}_i^T)^{-1} = (\underline{X}^T\underline{X})^{-1} + \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i}, \tag{6.10}$$

and

$$(\tilde{\underline{X}}^T\tilde{\underline{Y}}) = (\underline{X}^T\underline{Y} - \underline{x}_i Y_i)$$

By definition

$$\begin{aligned}
\hat{\underline{\beta}}_{-i} &= (\tilde{\underline{X}}^T\tilde{\underline{X}})^{-1}\tilde{\underline{X}}^T\tilde{\underline{Y}} = (\underline{X}^T\underline{X} - \underline{x}_i\underline{x}_i^T)^{-1}(\underline{X}^T\underline{Y} - \underline{x}_i Y_i) \\
&= (\underline{X}^T\underline{X})^{-1} + \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i}(\underline{X}^T\underline{Y} - \underline{x}_i Y_i) \\
&= (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y} + \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{X}\underline{Y}}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i} - (\underline{X}^T\underline{X})^{-1}\underline{x}_i Y_i - \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i Y_i}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i} \\
&= \hat{\underline{\beta}} + (\underline{X}^T\underline{X})^{-1}\left(\frac{\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i} - \underline{x}_i Y_i - \frac{\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i Y_i}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i}\right) \\
&= \hat{\underline{\beta}} + (\underline{X}^T\underline{X})^{-1}\left(\frac{\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y} - \underline{x}_i Y_i(1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i - \underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i Y_i}{1 - \underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i}\right) \\
&= \hat{\underline{\beta}} + \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y} - (\underline{X}^T\underline{X})^{-1}\underline{x}_i Y_i}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i} \\
&= \hat{\underline{\beta}} - \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i(Y_i - \underline{x}_i^T\hat{\underline{\beta}})}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i}
\end{aligned} \tag{6.11}$$

Hence

$$\hat{\underline{\beta}} - \hat{\underline{\beta}}_{-i} = \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i(Y_i - \underline{x}_i^T\hat{\underline{\beta}})}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i}$$

$$= \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i}{1-w_i} \qquad (6.12)$$

where $R_i = Y_i - \underline{x}_i^T\hat{\underline{\beta}}$ is the ith residual. The expression $w_i = \underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i$ can be interpreted as the distance of the single design point from the centre of the design, see for instance, Hinkley (1977b). This interpretation of w_i is explained in the next section.

6.3 Interpretation of the Distance w_i

In this section, we attempt to explain in what sense the quantiy $w_i = \underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i$ can be interpreted to measure the distance.

Definition 6.1 (Distances)

Let P and Q be two points where these may represent measurements $\underline{x}$ and $\underline{y}$ on two objects. A real valued function d(P,Q) is a distance function if it has the following properties:

(i) Symmetry, $d(P,Q) = d(Q,P)$;

(ii) non-negativity, $d(P,Q) \geq 0$;

(iii) identification mark, $d(P,P) = 0$.

In addition, many distance functions hold the following properties:

(iv) definiteness, $d(P,Q) = 0$ if and only if $P = Q$;

(v) triangle inequality, $d(P,Q) \leq d(P,R) + d(R,Q)$.

Definition 6.2 (Euclidean Distance)

Let $\underline{X}$ be an (n x P) data matrix with rows $\underline{x}_1^T, \underline{x}_2^T, \ldots, \underline{x}_n^T$. Then the Euclidean distance between the points $\underline{x}_i$ and $\underline{x}_j$ is d_{ij}, where

$$d_{ij}^2 = \sum_{k=1}^{p} (x_{ik}-x_{jk})^2 = \| \underline{x}_i - \underline{x}_j \|^2 \qquad (6.13)$$

This distance function satisfies properties (i) - (v).

Definition 6.3 (Mahalanobis Distance)

The squared Mahalnobis distance between points $\underline{x}_i$ and the sample mean vector $\bar{\underline{x}}$ is defined as

$$D_i^2 = (\underline{x}_i - \bar{\underline{x}})^T S^{-1} (\underline{x}_i - \bar{\underline{x}}) \qquad (6.14)$$

where S is the estimate for the variance-covariance matrix Σ.

The distance measure defined in Lemma 6.4 gives a measure of how far the ith point is from the centre of the data. For example, in multiple regression, w_i, measures the distance from the point $\underline{x}_i$ to the centre of the data and the cases with unusual values for the independent variables will tend to have large values of w_i. This point could be illustrated through a simple example. Consider the simple linear regression model,

$$y_i = \alpha + \beta x_i + e_i \qquad i = 1,2,\ldots,n$$

where $E(e_i) = 0$, $Var(e_i) = \sigma^2$ and $Cov(e_i, e_j) = 0$ for $i \neq j$.
For this model

$$(\underline{X}^T\underline{X}) = \begin{pmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix}, \quad \text{and} \quad (\underline{X}^T\underline{X})^{-1} = \frac{1}{SXX} \begin{pmatrix} \frac{\Sigma x_i^2}{n} & -\frac{\Sigma x_i}{n} \\ -\frac{\Sigma x_i}{n} & 1 \end{pmatrix}$$

where $SXX = \Sigma x_i^2 - n\bar{x}^2$ and $\bar{x} = \sum_{i=1}^{n} x_i/n$.

Now, $w_i = \underline{x}_i^T (\underline{X}^T\underline{X})^{-1} \underline{x}_i$ and hence

$$w_i = (1 \quad x_i) \begin{pmatrix} \frac{\Sigma x_i^2}{nSXX} & -\frac{\Sigma x_i}{nSXX} \\ \frac{\Sigma x_i}{nSXX} & \frac{1}{SXX} \end{pmatrix} \begin{pmatrix} 1 \\ x_i \end{pmatrix}$$

$$= \frac{\Sigma x_i^2}{nSXX} - \frac{2x_i \Sigma x_i}{nSXX} + \frac{x_i^2}{SXX}$$

$$= \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SXX} \qquad (6.15)$$

It could be seen from this expression that w_i will be minimised when $x_i = \bar{x}$ whereas w_i will increase for the unusual cases, that is, the cases for which x_i is far from $\bar{x}$. If x_i is sufficiently far from $\bar{x}$, then $\frac{(x_i - \bar{x})^2}{SXX}$ will become larger and w_i may be closer to one.

The distance notion described above can be formalised in several ways. The one that works (at least for models with intercept) is best defined in terms of the deviations from means form of the model (6.8). Suppose that $\underline{\mathcal{X}}^T\underline{\mathcal{X}}$ is the corrected cross product matrix, $\underline{\bar{x}}$ is a $P' \times 1$ vector of sample means of the P' independent variables and we redefine $\underline{x}_i^T$ to be the ith row of $\underline{X}$ without the one for intercept. Then Weisberg (1980) has shown that

$$w_i = \frac{1}{n} + (\underline{x}_i - \underline{\bar{x}})^T (\underline{\mathcal{X}}^T\underline{\mathcal{X}})^{-1} (\underline{x}_i - \underline{\bar{x}}) \qquad (6.16)$$

If $\frac{1}{n}$ term is dropped from the right hand side of (6.16) and the remaining term is multiplied by $(n-1)$, then we get the quantity called the Mahalnobis distance from $\underline{x}_i$ to the cetnre of data $\underline{\bar{x}}$. Geometrically the second term on the right hand side of (6.16) gives the equation of an ellipsoid centred at $\underline{\bar{x}}$. The definition of distance, in fact, depends on the data. The points near the long or major axis of the ellipsoid need to be much further away from $\underline{\bar{x}}$, in the usual Euclidean

distance sense, than the points closer to the minor axis, so as to have the same values for w_i .

6.4 *Properties of the Unweighted Jack-knifed Regression Estimator for the Linear Model*

Consider the linear regression model (6.8) $\underline{Y} = \underline{X}\,\underline{\beta} + \underline{e}$, where $\underline{Y}$ is an nx1 vector of observations, $\underline{X}$ is n x P matrix of known constants, $\underline{\beta}$ is a P x 1 vector of unknown parameters and $\underline{e}$ is an n x 1 vector of randomly distributed errors. The error variables are assumed to be independently and identically distributed such that $E(\underline{e}) = 0$ and $Var(\underline{e}) = \sigma^2 I$. For this model, we consider the least squares estimator $\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$ based on the complete data set as an estimator of $\underline{\beta}$.

In the unweighted jack-knife procedure the parameter estimates are obtained by successively deleting the rows of $\underline{X}$ matrix and $\underline{Y}$ vector. Let $\hat{\underline{\beta}}_{-i}$, i = 1,2,...,n denote the corresponding estimator of $\underline{\beta}$ obtained by deleting the ith row $\underline{x}_i^T$ from $\underline{X}$ and the ith observation Y_i from $\underline{Y}$. Then under the unweighted jack-knife procedure the psuedovalues are defined as usual.

$$\tilde{\beta}_i = n\hat{\underline{\beta}} - (n-1)\hat{\underline{\beta}}_{-i} \tag{6.17}$$

Now from Lemma 6.4, we know that

$$\hat{\underline{\beta}}_{-i} = \hat{\underline{\beta}} - \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i(Y_i - \underline{x}_i^T\hat{\underline{\beta}})}{1-\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i} \quad ,$$

and substituting in (6.17) we get

$$\tilde{\beta}_i = n\hat{\beta} - (n-1)\left[\hat{\beta} - \frac{(X^TX)^{-1}x_i(Y_i - x_i^T\hat{\beta})}{1-x_i^T(X^TX)^{-1}x_i}\right]$$

$$= \hat{\beta} + \frac{(n-1)D_o^{-1}x_iR_i}{1-w_i} \quad , \; i = 1,2,3,\ldots,n \; , \tag{6.18}$$

where $D_o = X^TX$ is called the design matrix, $R_i = Y_i - x_i^T\hat{\beta}$ denotes the ith residual $w_i = x_i^TD_o^{-1}x_i$ measures the distances of single design points from the centre of the design and $\frac{R_i}{1-w_i} = Y_i - \hat{Y}_{-i}$ is the difference between Y_i and its least squares predictor from all other observations.

The jack-knifed estimator $\tilde{\beta}$ is the average of the pseudovalues give by (6.18),i.e.,

$$\tilde{\beta} = n^{-1}\sum_{i=1}^{n}\tilde{\beta}_i = n^{-1}\sum_{i=1}^{n}\left[n\hat{\beta} + \frac{(n-1)D_o^{-1}x_iR_i}{1-w_i}\right]$$

$$= \hat{\beta} + \frac{n-1}{n}D_o^{-1}\sum_{i=1}^{n}(1-w_i)^{-1}x_iR_i \tag{6.19}$$

6.4.1 Bias and Variance of the Jack-knifed Regression Estimator $\tilde{\beta}$

The expected value of the jack-knifed estimator $\tilde{\beta}$ is given by

$$E(\tilde{\beta}) = E(\hat{\beta}) + \frac{n-1}{n}D_o^{-1}\sum_{i=1}^{n}(1-w_i)^{-1}x_iE(R_i)$$

$$= \beta \qquad \because \; E(R_i) = 0, \; E(\hat{\beta}) = \beta \tag{6.20}$$

Thus $\tilde{\beta}$ is an unbiased estimator of β . Since $\tilde{\beta} \neq \hat{\beta}$ in general, we know by Gauss-Markov theorem that if the model is true,

$$\mathrm{Var}(\tilde{\beta}) \geq \mathrm{Var}(\hat{\beta}) \; . \tag{6.21}$$

However, when w_i = constant then $Var(\underline{\tilde{\beta}}) = Var(\underline{\hat{\beta}})$. Also under the conditions of Lemma 6.3, we note that $\tilde{\beta}$ is consistent.

Variance of $\tilde{\beta}$

The variance of the jack-knifed estimator $\tilde{\beta}$ is given by

$$Var(\underline{\tilde{\beta}}) = Var\left(\underline{\hat{\beta}} + \frac{n-1}{n} D_o^{-1} \sum_{i=1}^{n} (1-w_i)^{-1} \underline{x}_i R_i\right)$$

Since $\underline{\hat{\beta}}$ and $\underline{R}$ are uncorrelated, therefore

$$Var(\underline{\tilde{\beta}}) = Var(\underline{\hat{\beta}}) + \left(\frac{n-1}{n}\right)^2 D_o^{-1} Var\left(\sum_{i=1}^{n} (1-w_i)^{-1} \underline{x}_i R_i\right) D_o^{-1} \qquad (6.22)$$

Let $C_i = \dfrac{\underline{x}_i}{1-w_i}$ then $\underline{C}^T = (C_1, C_2, \ldots, C_n)$ and expression (6.22) can be written as

$$Var(\underline{\tilde{\beta}}) = Var(\underline{\hat{\beta}}) + \frac{(n-1)^2}{n^2} D_o^{-1} Var(\underline{C}^T \underline{R}) D_o^{-1}$$

$$= \sigma^2 D_o^{-1} + \left(\frac{n-1}{n}\right)^2 D_o^{-1} \underline{C}^T Var(\underline{R}) \underline{C} D_o^{-1} \quad , \quad \because Var(\underline{\hat{\beta}}) = \sigma^2 D_o^{-1}$$

$$= \sigma^2 D_o^{-1} + \frac{(n-1)^2}{n^2} D_o^{-1} \underline{C}^T (\underline{I} - \underline{X} D_o^{-1} \underline{X}^T) \underline{C} D_o^{-1} \sigma^2 \quad , \quad \because Var(\underline{R}) = (\underline{I} - \underline{X} D_o^{-1} \underline{X}^T)\sigma^2$$

$$= \sigma^2 D_o^{-1} + \frac{(n-1)^2}{n^2} D_o^{-1} (\underline{C}^T \underline{C} - \underline{C}^T \underline{X} D_o^{-1} \underline{X}^T \underline{C}) D_o^{-1} \sigma^2 \qquad (6.23)$$

Now, let

$$D_m = \sum_{i=1}^{n} (1-w_i)^{-m} \underline{x}_i \underline{x}_i^T \quad , \quad (m = 0,1,2) \quad . \qquad (6.24)$$

Then, we have

$$\underline{C}^T \underline{C} = \sum_{i=1}^{n} \frac{\underline{x}_i \underline{x}_i^T}{(1-w_i)^2} = D_2$$

$$\underline{C}^T\underline{X} = \left(\frac{\underline{x}_1}{1-w_1}, \frac{\underline{x}_2}{1-w_2}, \frac{\underline{x}_3}{1-w_3}, \ldots, \frac{\underline{x}_n}{1-w_n}\right) \begin{pmatrix} \underline{x}_1^T \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_n^T \end{pmatrix}$$

$$= \sum_{i=1}^{n} \frac{\underline{x}_i \underline{x}_i^T}{1-w_i} = D_1$$

Substituting $\underline{C}^T\underline{C} = D_2$ and $\underline{C}^T\underline{X} = D_1$ in expression (6.23), we get

$$\mathrm{Var}(\tilde{\underline{\beta}}) = \sigma^2 D_o^{-1} + \frac{(n-1)^2}{n^2} D_0^{-1}(D_2 - D_1 D_o^{-1} D_1) D_o^{-1}\sigma^2$$

$$= \sigma^2\left\{D_0^{-1} + \left(\frac{n-1}{n}\right)^2 D_o^{-1} (D_2 - D_1 D_0^{-1} D_1) D_o^{-1}\right\} \qquad (6.25)$$

Now supposing that w_i is of order n^{-1} and expanding $(1-w_i)^{-m}$ we find the order of $\mathrm{Var}(\tilde{\underline{\beta}}) - \mathrm{Var}(\hat{\beta})$.

$$(1-w_i)^{-m} = 1 + mw_i + \frac{m(m+1)}{2!} w_i^2 + \ldots + \frac{m(m+1)+\ldots+(m+n-1)}{n!} w_i^n ,$$

$$= 1 + mw_i + O(n^{-2}) \qquad (6.26)$$

and substituting (6.26) in (6.24), we get

$$D_m = \sum_{i=1}^{n} (1-w_i)^{-m} \underline{x}_i \underline{x}_i^T = \sum_{i=1}^{n} w_i (1+mw_i+O(n^{-2})) \underline{x}_i \underline{x}_i^T ,$$

$$= \sum_{i=1}^{n} \underline{x}_i \underline{x}_i^T + m \sum_{i=1}^{n} w_i \underline{x}_i \underline{x}_i^T + O(n^{-2}) \qquad (6.27)$$

Let $\sum_{i=1}^{n} \underline{x}_i \underline{x}_i^T = L$ and $\sum_{i=1}^{n} w_i \underline{x}_i \underline{x}_i^T = M$, then (6.27) reduces to

$$D_m = L + mM + O(n^{-2}) \qquad (6.28)$$

From (6.28), we have

$$D_0 = L \quad , \quad D_o^{-1} = L^{-1} \quad ,$$

$$D_1 = L + M + O(n^{-1}) \ ,$$

$$D_2 = L + 2M + O(n^{-1}) \ ,$$

and

$$D_1 D_o^{-1} D_1 = (L+M+O(n^{-1}))\ (L^{-1})\ (L+M+O(n^{-1})) \ ,$$

$$= LL^{-1}L + ML^{-1}L + LL^{-1}M + O(n^{-1}) \ ,$$

$$= L + 2M + O(n^{-1}) \ .$$

Hence

$$D_2 - D_1 D_o^{-1} D_1 = L + 2M + O(n^{-1}) - L - 2M + O(n^{-1}) = O(n^{-1}) \qquad (6.29)$$

Substituting (6.29) in (6.25), we get

$$\mathrm{Var}(\tilde{\underline{\beta}}) - \mathrm{Var}(\hat{\underline{\beta}}) = \sigma^2 \left(\frac{n-1}{n}\right)^2 [L^{-1}][O(n^{-1})][L^{-1}]$$

$$= O(n^{-3}) \qquad (6.30)$$

Thus the difference between the variances of the jack-knifed estimator $\tilde{\beta}$ and the ordinary least squares estimator $\hat{\beta}$ is of order n^{-3} .

6.5 Properties of the Unweighted Jack-knifed Variance Estimator

Another important aspect of the jack-knife technique is that it provides a distribution free estimation of the variance of the parameter estimator. The jack-knife variance estimate is given by

$$\hat{V}_J = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\tilde{\underline{\beta}}_i - \tilde{\underline{\beta}})(\tilde{\underline{\beta}}_i - \tilde{\underline{\beta}})^T \ . \qquad (6.30)$$

Now

$$\tilde{\underline{\beta}}_i - \tilde{\underline{\beta}} = \hat{\underline{\beta}} + (n-1)D_o^{-1}(1-w_i)^{-1}\underline{x}_i R_i - \hat{\underline{\beta}} - (n-1)n^{-1}D_o^{-1} \sum_{i=1}^{n} (1-w_i)\underline{x}_i R_i \ ,$$

$$= (n-1)D_o^{-1}\left[\frac{\underline{x}_i}{1-w_i} R_i - \frac{1}{n}\sum_{j=1}^{n} \frac{\underline{x}_j R_j}{1-w_j}\right] \ .$$

Substituting the expression for $\underline{\beta}_i - \tilde{\underline{\beta}}$ in (6.30), we get

$$\hat{V}_J = \frac{(n-1)^2}{n(n-1)} D_o^{-1} \sum_{i=1}^{n} \left(\frac{\underline{x}_i}{1-w_i} R_i - \frac{1}{n} \sum_j \frac{\underline{x}_j}{1-w_j} R_j \right) \left(\frac{\underline{x}_i}{1-w_i} R_i - \frac{1}{n} \sum_k \frac{\underline{x}_k}{1-w_k} R_k \right)^T D_o^{-1}$$

$$= \frac{n-1}{n} D_o^{-1} \sum_{i=1}^{n} \left(\underline{C}_i R_i - \frac{1}{n} \sum_{j=1}^{n} \underline{C}_j R_j \right) \left(\underline{C}_i R_i - \frac{1}{n} \sum_{k=1}^{n} \underline{C}_k R_k \right)^T D_o^{-1}$$

$$= \frac{n-1}{n} D_o^{-1} \sum_{i=1}^{n} \left(\underline{C}_i R_i^2 \underline{C}_i^T - \frac{1}{n} \sum_{j=1}^{n} \underline{C}_j R_j R_i \underline{C}_i^T - \frac{1}{n} \sum_{k=1}^{n} \underline{C}_i R_i R_k \underline{C}_k^T + \frac{1}{n^2} \sum_j \sum_k \underline{C}_j R_j R_k \underline{C}_k^T \right) D_o^{-1} \qquad (6.31)$$

$$E(\hat{V}_J) = \frac{n-1}{n} D_o^{-1} \sum_{i=1}^{n} \left(\underline{C}_i E(R_i^2) \underline{C}_i^T - \frac{1}{n} \sum_j \underline{C}_j E(R_i R_j) \underline{C}_i^T - \frac{1}{n} \sum_k \underline{C}_i E(R_i R_k) \underline{C}_k^T + \frac{1}{n^2} \sum_j \sum_k \underline{C}_j E(R_j R_k) \underline{C}_k^T \right) D_o^{-1}$$

$$= \frac{n-1}{n} D_o^{-1} \left(\sum_{i=1}^{n} \underline{C}_i \mathrm{Var}(R_i) \underline{C}_i^T - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \underline{C}_i \mathrm{Cov}(R_i R_j) \underline{C}_j^T - \frac{1}{n} \sum_j \sum_k \underline{C}_j \mathrm{Cov}(R_i R_k) \underline{C}_k^T + \frac{1}{n} \sum_j \sum_k \underline{C}_j \mathrm{Cov}(R_j R_k) \underline{C}_k^T \right) D_o^{-1}$$

$$= \frac{n-1}{n} D_o^{-1} \left(\sum_{i=1}^{n} \underline{C}_i \mathrm{Var}(R_i) \underline{C}_i^T - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \underline{C}_i \mathrm{Cov}(R_i R_j) \underline{C}_j^T \right) D_o^{-1} \qquad (6.32)$$

Now, $\underline{R}^T = (R_1, R_2, \ldots, R_n)$ and

$$\mathrm{Var}(\underline{R}) = \sigma^2 (\underline{I} - \underline{X} D_o^{-1} \underline{X}^T) = \sigma^2 \left(\underline{I} - \begin{pmatrix} \underline{x}_1^T D_o^{-1} \\ \underline{x}_2^T D_o^{-1} \\ \vdots \\ \underline{x}_n^T D_o^{-1} \end{pmatrix} (\underline{x}_1, \ldots, \underline{x}_n) \right)$$

$$= \sigma^2 \left(\underline{I} - \begin{pmatrix} \underline{x}_1^T D_o^{-1} \underline{x}_1, \ldots, \underline{x}_1^T D_o^{-1} \underline{x}_n \\ \cdots\cdots\cdots\cdots \\ \underline{x}_i^T D_o^{-1} \underline{x}_1, \ldots, \underline{x}_i^T D_o^{-1} \underline{x}_n \\ \cdots\cdots\cdots\cdots \\ \underline{x}_n^T D_o^{-1} \underline{x}_1, \ldots, \underline{x}_n^T D_o^{-1} \underline{x}_n \end{pmatrix} \right) \qquad (6.33)$$

Thus $\mathrm{Var}(R_i) = \sigma^2(1-\underline{x}_i^T D_o^{-1} \underline{x}_i)$ and $\mathrm{Cov}(R_i R_j) = -\sigma^2 \underline{x}_i^T D_o^{-1} \underline{x}_j$. Taking the first term in (6.32), we have

$$\sum_{i=1}^{n} \underline{C}_i \, \mathrm{Var}(R_i) \underline{C}_i^T = \sigma^2 \sum_{i=1}^{n} \left(\frac{\underline{x}_i}{1-w_i} (1-\underline{x}_i^T D_o^{-1} \underline{x}_i) \frac{\underline{x}_i^T}{1-w_i} \right)$$

$$= \sigma^2 \sum_{i=1}^{n} \frac{\underline{x}_i \underline{x}_i^T}{1-w_i}$$

$$= \sigma^2 D_1 \,, \tag{6.34}$$

and

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \underline{C}_i \, \mathrm{Cov}(R_i R_j) \underline{C}_j^T = \sigma^2 \sum_{i=1}^{n} \frac{\underline{x}_i \underline{x}_i^T}{(1-w_i)^2} - \sigma^2 \sum_i \sum_j \frac{\underline{x}_i}{1-w_i} \underline{x}_i^T D_o^{-1} \frac{\underline{x}_j}{1-w_j} \underline{x}_j^T$$

$$= \sigma^2 (D_2 - D_1 D_o^{-1} D_1) \tag{6.35}$$

Substituting (6.34) and (6.35) in (6.32), we get

$$E(\hat{V}_J) = \frac{n-1}{n} D_o^{-1} \left[\sigma^2 D_1 - \frac{1}{n} \sigma^2 (D_2 - D_1 D_o^{-1} D_1) \right] D_o^{-1}$$

$$= \frac{n-1}{n} D_o^{-1} \left[D_1 - \frac{1}{n} (D_2 - D_1 D_o^{-1} D_1) \right] D_o^{-1} \sigma^2 \tag{6.36}$$

We note that $E(\hat{V}_J) \neq \mathrm{Var}(\hat{\underline{\beta}}) = \sigma^2 D_o^{-1}$ and also $E(\hat{V}_J) \neq \mathrm{Var}(\tilde{\underline{\beta}})$. Thus the jackknifed variance estimator $\hat{V}_J$ which may be used to estimate both $\mathrm{Var}(\hat{\underline{\beta}})$ and $\mathrm{Var}(\tilde{\underline{\beta}})$ is biased in the balanced case. However, under the conditions of Lemma 3.4 (Miller, 1974a) it could be shown that $n\hat{V}_J \to n \, \mathrm{Var}(\hat{\underline{\beta}})$ which implies that $\hat{V}_J$ is an accurate large sample variance estimate.

Supposing that w_i are of order n^{-1} and defining D_m^* as in (6.37) we obtain an approximation of $E(\hat{V}_J)$ to $O(n^{-4})$. Let

$$D_m^* = \sum_{i=1}^{n} w_i^m \underline{x}_i \underline{x}_i^T \qquad (i = 1,2) \tag{6.37}$$

Then we have

$$D_o^* = \sum_{i=1}^{n} \underset{-i}{x}_i \underset{-i}{x}_i^T = D_o$$

$$D_1^* = \sum_{i=1}^{n} w_i \underset{-i}{x}_i \underset{-i}{x}_i^T$$

$$D_2^* = \sum_{i=1}^{n} w_i^2 \underset{-i}{x}_i \underset{-i}{x}_i^T \qquad (6.38)$$

From (6.28) and (6.38), we get

$$D_1 = D_0 + D_1^* + D_2^*$$

$$D_2 = D_o + 2D_1^* + 3D_2^*$$

Now,

$$D_1 - \frac{1}{n}(D_2 - D_1 D_o^{-1} D_1) \doteq (D_o + D_1^* + D_2^*) - \frac{1}{n}\{D_o + 2D_1^* + 3D_2^* - (D_o + D_1^* + D_2^*) D_o (D_o + D_1^* + D_2^*)\}$$

$$\doteq D_o + D_1^* + D_2^* - \frac{1}{n}[D_o + 2D_1^* + 3D_2^* - D_o - 2D_1^* - 2D_2^* - D_1^* D_o^{-1} D_1^* - D_2^* D_o^{-1} D_1^* - D_1 D_o^{-1} D_2^*]$$

$$\doteq D_o + D_1^* + D_2^* + O(n^{-3}) \ .$$

Hence

$$D_o^{-1}[D_1 - \frac{1}{n}(D_2 - D_1 D_o^{-1} D_1)] D_o^{-1} \doteq D_o^{-1}[D_o + D_1^* + D_2^*] D_0^{-1} + O(n^{-3}) \ .$$

$$\doteq D_o^{-1} + D_o^{-1} D_1^* D_o^{-1} + D_o^{-1} D_2^* D_o^{-1} + O(n^{-4}) \qquad (6.39)$$

Substituting (6.39) in (6.36), we get to $O(n^{-4})$

$$E(\hat{V}_J) \doteq \frac{n-1}{n} D_o^{-1}[D_o + D_1^* + D_2^*] D_o^{-1} \sigma^2 \qquad (6.40)$$

These results indicate that in case of an unweighted jack-knife procedure, the jack-knifed regression estimator $\tilde{\underline{\beta}}$ is unbiased but less efficient than the least squares estimator $\hat{\underline{\beta}}$. The jack-knifed variance estimate $\hat{V}_J$ also gives biased estimate of the variance.

6.6 Conclusions

In this chapter we have considered the application of the unweighted jack-knife to multiple linear regression model. We observe that the jack-knifed estimates of regression parameters are unbiased but less efficient $(VAR(\tilde{\underline{\beta}}) > Var(\hat{\underline{\beta}}))$ and the jack-knifed variance estimator $\hat{V}_J$ is biased. These failures (with respect to design matrix) are due to the fact that model is unbalanced whereas the jack-knife procedure applied is of the standard balanced form which assumes the regression errors to be independently and identically distributed. The lack of balance comes from D_o. The reason is that certain observations may be influential and may cause a severe non-uniform dispersion in the observed design, that is, w_i may differ widely. However, if w_i are constant then $Var(\tilde{\underline{\beta}}) = Var(\hat{\underline{\beta}})$. This shows that when the linear regression weights are fairly unequal, the unweighted jack-knife gives inaccurate results. Thus we note that the traditional estimation techniques generally applied to the balanced form of data are not suitable for the unbalanced data. We therefore require a procedure that may take into account the unbalanced structure of the data. This suggests the need for more robust jack-knife procedures which we discuss in subsequent chapters.

CHAPTER 7

INFLUENCE FUNCTION AND REGRESSION

7.1 Introduction

Consider the model

$$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{e}$$

Where $\underline{Y}$ is an $n \times 1$ vector of observations, $\underline{X}$ is an $n \times p$ full rank matrix of known constants, $\underline{\beta}$ is a $p \times 1$ vector of unknown parameters and $\underline{e}$ is an $n \times 1$ vector of randomly distributed errors such that $E(\underline{e}) = \underline{0}$ and $Var(\underline{e}) = \sigma^2\underline{I}$.

In fitting the model (7.1) we estimate $\underline{\beta}$ by $\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$ and the fitted values corresponding to the observed values are then given by

$$\hat{\underline{Y}} = \underline{X}\hat{\underline{\beta}}$$

$$= \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}\underline{Y} \qquad (7.2)$$

The vector of residuals is given by

$$\underline{R} = \underline{Y} - \hat{\underline{Y}}$$

$$= [\underline{I} - \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T]\underline{Y} \qquad (7.3)$$

The ith residual is denoted by $R_i = y_i - \underline{x}_i^T\,\hat{\underline{\beta}}$, where $\underline{x}_i^T$ is the ith row of $\underline{X}$ matrix. The covariance matrices of $\hat{\underline{Y}}$ and $\underline{R}$ are respectively,

$$V(\hat{\underline{Y}}) = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T\sigma^2 \ . \qquad (7.4)$$

and

$$V(\underline{R}) = [\underline{I} - \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T]\sigma^2 \quad . \tag{7.5}$$

The full sample estimate of σ^2 is

$$s^2 = \underline{Y}^T[\underline{I}-\underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}]\underline{Y}/(n-p) \quad . \tag{7.6}$$

The matrix defined by $\underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}$ is very important and we shall call it $\underline{W} = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}$. Another important set of quantities is the diagonal elements of $\underline{W}$, given by

$$w_i = \underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i \tag{7.7}$$

which gives a measure of how far the ith case is from the centre of the data.

Recall that $Var(R_i) = (1-w_i)\sigma^2$ will be small when w_i is large with the result that cases with $\underline{x}_i$ near $\bar{\underline{x}}$ will fit poorly whereas the cases with $\underline{x}_i$ far from $\bar{\underline{x}}$ will fit well. This is undesirable because the violations of the model are most likely to occur under unusual conditions and it may not be possible to find these violations by simply examining the residuals. For use in diagnostic procedures, several transformations of the ordinary residuals have been suggested to overcome some of their shortcomings. For example, an improved set of residuals could be obtained by scaling so that cases with large w_i get larger scaled residuals while cases with smaller w_i get smaller scaled residulas. Then all the residuals in an analysis can be compared directly. One method to do this scaling is to divide each of the residuals by an estimate of its standard deviation. These are then called *Studentised residuals* and are defined by

$$t_i = \frac{R_i}{s(1-w_i)^{\frac{1}{2}}}, \qquad i = 1,2,\ldots,n \quad . \tag{7.8}$$

The Studentised residuals are slightly correlated with Y_i's but in practice this correlation is negligible. The residuals and the Studentised residuals represent one attempt to measure the success or failure of the model. However, the important point to consider in any data analysis is the impact that each observation has on the estimates. For example, in some data sets one or more cases may have sufficient impact on the analysis such that if these cases are removed, the results based on the reduced data set may be quite different from those based on the complete data. Thus the outcome of an analysis would change dramatically. We shall term such cases influential and the study of the dependence of conclusions and inferences on various aspects of a problem formation as the study of influence.

The basic idea of influence analysis is quite simple. In this chapter we discuss the influence function and its role in regression analysis. In Section 7.2, we discuss some results concerning the influence function and the various sample versions of the influence function which provide the basis and the justification for the various techniques used for the detection of influential cases. It is well recognised that not all the cases in a set of data play an equal role in determining estimates, tests and other statistics. The idea of measuring the influence of each case on the regression leads us to look at the data in a slightly different way. Based on the empirical influence function we discuss the different methodologies for assessing the influence of individual or group of cases on regression analysis in Section 7.3. We, however, limit our discussion to aspects of analysis that can be summarised by the sample influence function. Finally, in Section 7.4 we give the concluding remarks.

7.2 The Influence Function

The influence function is a useful mathematical construction for studying the behaviour of estimates. Defined in terms of population parameters, it monitors the change in estimates when the data are slightly modified. The original use of the influence function and related notions, exploited by Von Mises (1947, 1964) and expanded by Reeds (1976), lies in determining asymptotic properties of an estimator. Andrew et al., (1972), Hampel (1974), Hinkley (1977b) and Cook and Weisberg (1980, 1982) use influence functions to compare estimators and suggest robust modifications of the existing estimation techniques.

Let R be the real line and T be a real valued functional defined on some subset of the set of all probability measures on R. Further, let F denote a probability measure on R for which T is defined and δ_z dnote the probability measure determined by the point mass 1 in any given point $z \in R$. Then the influence function $I_{T,F}(z)$ of the estimatør T at the underlying distribution F is defined pointwise by

$$I_{T,F}(z) = \lim_{\varepsilon \to 0} \frac{T[(1-\varepsilon)F+\varepsilon\delta z]-T(F)}{\varepsilon} \tag{7.9}$$

provided the limit exists for every point $z \in R$. Thus, influence function is just the ordinary right hand derivative, evaluated at $\varepsilon = 0$ of $T[(1-\varepsilon)F+\varepsilon\delta z]$. This point is illustrated through simple examples from Hampel (1974).

Example 7.1 (Mean)

The statistical functional is

$$T(F) = \int z dF(z) = \mu \ . \tag{7.10}$$

and

$$T[(1-\varepsilon)F+\varepsilon\delta z] = \int z d\{(1-\varepsilon)F+\varepsilon\delta z\}$$

$$= (1-\varepsilon)\mu + \varepsilon z \tag{7.11}$$

Now, using (7.9), (7.10) and (7.11), the influence function is given by

$$I_{T,F}(z) = \lim_{\varepsilon\to 0} \frac{[(1-\varepsilon)\mu+\varepsilon z-\mu]}{\varepsilon} = z-\mu \tag{7.12}$$

Example 7.2 (Variance)

The statistical functional in this case is

$$T(F) = \int (z-\mu)^2 dF(z) = \sigma^2 \ . \tag{7.13}$$

and

$$T[(1-\varepsilon)F+\varepsilon\delta z] = \int (z-\mu)^2 d\{(1-\varepsilon)F+\varepsilon\delta z\},$$

$$= (1-\varepsilon)\sigma^2+\varepsilon(z-\mu)^2 \tag{7.14}$$

Now, using (7.9), (7.13) and (7.14), the influence function is given by

$$I_{T,F}(z) = \lim_{\varepsilon\to 0} \frac{[(1-\varepsilon)\sigma^2+ (z-\mu)^2-\sigma^2]}{\varepsilon} \ .$$

$$= (z-\mu)^2 - \sigma^2 \tag{7.15}$$

The study of influence function not only serves to deepen our understanding of the estimators but also helps to derive new estimators with specific robustness properties. The influence function can be used in a number of ways to see how individual cases affect the sample average. One possibility is to study how does the mean change if an additional observation z were added, giving a sample of size $n + 1$.

The other possibility is to study the change in the mean when an observation is deleted. Both these possibilities are illustrated through a simple example from Cook and Weisberg (1982), which helps to understand the use of the influential function and suggests specific procedures for special purposes.

Example 7.3 (Sample Average)

Case I (Addition of an Observation)

Suppose that a single additional observation z is added, giving a sample of size $n + 1$. The new sample distribution function is given by

$$\hat{F}_+ = \frac{n\hat{F}}{n+1} + \frac{1}{n+1}\,\delta z$$

Now

$$T(\hat{F}_+) = T\left[\frac{n}{n+1}\,\hat{F} + \frac{1}{n+1}\,\delta z\right]$$

$$= T(\hat{F}) + \frac{1}{n+1}\,(z - T(\hat{F}))$$

$$\bar{z}_+ = \bar{z} + \frac{1}{n+1}\,(z - \bar{z}) \qquad (7.16)$$

$$T(\hat{F}) = \int z\,d\hat{F}(z) = \bar{z}$$

$$T(\hat{F}_+) = \bar{z}_+$$

It is clear from expression (7.16) that for a fixed sample size n, $\bar{z}_+ - \bar{z}$ increases linearly as z deviates from $\bar{z}$. This gives the influence of a single future case on the current sample average and only in an indirect way indicates the influence of z_i, $i = 1,2,\ldots,n$ on $\bar{z}$.

Case II (Deletion of an Observation)

The influence of the ith observation z_i on $\bar{z}_i$ may be determined by removing z_i from the sample. The new sample distribution function in this case is given by

$$\hat{F}_{-i} = \frac{n}{n-1}\hat{F} - \frac{1}{n-1}\delta z_i$$

The influence of z_i on $\bar{z}$ can be obtained by evaluating the ith function at $z = z_i$. Thus proceeding as before we have

$$T(\hat{F}_{-i}) = T\left(\frac{n}{n-1}\hat{F} - \frac{1}{n-1}\delta z_i\right)$$

$$= T(\hat{F}) - \frac{1}{n-1}[z_i - T(\hat{F})]$$

$$\bar{z}_{-i} = \bar{z} - \frac{1}{n-1}(z_i - \bar{z})$$

$$\bar{z} - \bar{z}_{-i} = \frac{1}{n-1}(z_i - \bar{z}) \qquad (7.17)$$

where $\bar{z}_{-i}$ denotes the sample mean with the ith case deleted. The expression (7.17) has been expressed in terms of the full sample average. This shows that the influence of a single observation depends on the sample size and the full sample residual. Thus any case with sufficiently large residual will be influential for the sample mean.

7.2.1 The Influence Function for the Linear Model

Consider the linear model described in (7.1). Let the design point $\underline{x}$ and the response variable Y have a joint distribution function F such that

$$E_F\left\{\begin{pmatrix}\underline{x}\\ y\end{pmatrix}(\underline{x}^T,\ y)\right\} = \begin{pmatrix}\underline{\Sigma}(F) & \gamma(F)\\ \gamma^T(F) & \Gamma(F)\end{pmatrix} \qquad (7.18)$$

Then assuming that $\underline{\Sigma}$ is non-singular and the functional corresponding to the least squares estimator of $\underline{\beta}$ is given by $T(F) = \underline{\Sigma}^{-1}(F)\gamma(F)$, Hinkley (1977b) defines the influence function of $\underline{\beta}$ at $(\underline{x}^T, y)$ as

$$I_{\beta,F}(\underline{x},y) = \sum^{-1}(F)\underline{x}(y-\underline{x}^T T(F)) \tag{7.19}$$

$$= \sum^{-1}\underline{x}(y-\underline{x}^T\underline{\beta}) \tag{7.20}$$

The derivational details could be found in Cook and Weisberg (1982).

7.2.2 Sample Versions of the Influence Function

The influence function, discussed earlier in this chpater, describes an estimation technique with respect to a theoretical sampling population F. In case of finite sample situations, more information appropriate for the specific problem can be obtained by using the finite sample versions of the influence function that correspond directly to the observed data. Several finite sample versions of the influence function that depend on an observed sample have been suggested. However, the empirical influence function (EIF) and the sample influence function (SIF) have received the greatest attention and will be briefly discussed in this Subsection.

In general, the EIF is obtained by substituting the sample distribution function $\hat{F}$ for F in the influence function. For linear models, using (7.19) and putting $\hat{\beta} = T(\hat{F})$ will yield

$$EIF(\underline{x},y) = n(\underline{X}^T\underline{X})^{-1}\underline{x}(y-\underline{x}^T\hat{\underline{\beta}}) \ , \tag{7.21}$$

and

$$EIF_i = EIF(\underline{x}_i, y_i) = n(\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i \tag{7.22}$$

where $R_i = y_i - \underline{x}_i^T\hat{\underline{\beta}}$. The empirical influence function (7.22) appears to be an exact analogy of the influence function. It was used by Hinkley (1977b) for the weighted jack-knifed regression estimation, which we will discuss in the next chapter. The empirical influence function is appealing in the sense that it measures the effect of an infinitesimal perturbation of $\hat{F}$ at $\underline{x}_i$. The EIF

assumes that an infinitely large sample has been used to obtain $\hat{F}$, and it measures the instantaneous rate of change in the estimator as a single case is added to the data.

A second sample version of the influence function which is more suitable for the jack-knife can be constructed to show the influence of the ith case on the computed estimate of β. The underlying idea is to substitute the sample distribution function with the ith case deleted for F in the influence function and then evaluate the resulting EIF at the ith case. Let $\hat{F}_{-i}$ denote the sample distribution function with the ith case deleted. For the least squares estimators of $\underline{\beta}$, substituting F_{-i} for F in (7.19) yields an empirical influence function with the ith case deleted.

$$EIF_{-i}(\underline{x},y) = (n-1)(\underline{X}^T_{-i}\underline{X}_{-i})^{-1}\underline{x}(y-\underline{x}^T\hat{\underline{\beta}}_{-i}) \ , \tag{7.23}$$

where $\hat{\underline{\beta}}_{-i} = T(\hat{F}_{-i})$ and $(\underline{X}^T_{-i}\underline{X}_{-i})/(n-1) = \int \underline{x}\underline{x}^T d\hat{F}_{-i}$. This gives n empirical influence functions, one for each $i = 1,2,\ldots,n$. The influence of the ith case is determined by evaluating (7.23) at $(\underline{x}_i, y_i)$.

$$EIF_{-i}(\underline{x}_i,y_i) = (n-1)(\underline{X}^T_{-i}\underline{X}_{-i})^{-1}\underline{x}_i(y_i-\underline{x}_i^T\hat{\underline{\beta}}_{-i}) \tag{7.24}$$

The expression (7.24) can be written in a more revealing form in terms of the full sample as

$$EIF_{-i}(\underline{x}_i,y_i) = (n-1)\frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i}{(1-w_i)^2} \tag{7.25}$$

The interpretation of EIF_{-i} is similar to that for EIF discussed above.

As pointed out earlier, in this Subsection, both EIF and EIF_{-i} have been developed under the assumption that infinitely large samples have been used to obtain $\hat{F}$ and $\hat{F}_{-i}$, $i = 1,2,\ldots,n$. The sample size n in expressions (7.22) and (7.25) appears as a result of the covariance structure and does not necessarily reflect the effects of a finite sample. Therefore, when investigating the influence of individual cases on the computed statistics, a more explicit dependence on n is desirable, otherwise important finite sample characteristics can be obscured. A more desirable sample version of the influence function can be obtained by setting $F = \hat{F}$ and taking $\varepsilon = -\frac{1}{n-1}$ in the definition of the influence function (7.9), see for instance, Chapter 3, expression (3.85). The sample influence function is given by

$$SIF_i = -(n-1)(T(\hat{F}_{-i}) - T(\hat{F})) , \tag{7.26}$$

$$= (n-1)(\underline{\hat{\beta}}-\underline{\hat{\beta}}_{-i}) .$$

$$= (n-1)\frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i}{(1-w_i)} \tag{7.27}$$

which is proportional to the change in the estimate of $\underline{\beta}$ when a case is deleted.

It could be seen from the above formulations that the essential difference between these three versions of the influence function lies in the power of $(1-w_i)$ term in the denominator, while the numerators are essentially the same when evaluated at the sample points. Recall that remote rows of $\underline{X}$ will tend to have $(1-w_i)$ small. The EIF_{-i} will be most sensitive to cases with large w_i whereas EIF will be least sensitive. The SIF lies between these versions in terms of the relative weight given to w_i .

The sample influence function (SIF) defined as (7.26) has desirable properties. It is computed from the observed data and apart from constants it is interpreted as a change in a statistic when a case is deleted. It has the additional advantage that for many problems including the linear least squares regression it can be easily computed.

In the remainder of this chapter, we discuss the methods for finding influential cases based on the sample influence function. In order to be most useful such methods should allow the cases to be ordered on the basis of influence. For the linear least squares, the SIF for $\hat{\beta}$ is a p-dimensional vector and there is no natural way of ordering the multi-dimensional vectors. Even in the case $p = 2$, where a scatter plot of SIF can be constructed and interpreted there is no natural way to construct complete ordering of the points on the basis of influence. To overcome these problems Cook (1977, 1979) and Cook and Weisberg (1980), among others, have developed distance measures for detecting influential cases in regression. We discuss these distance measures in the next section.

7.3 Applications of the Influence Function for Assessment of Influential Observations in Regression

It is generally recognised that the overall summary statistic such as R^2 or $\hat{\beta}$, etc., computed from data based on full regression model can lead to distorted and misleading results. These considerations have lead to the use of a number of procedures such as plots of residuals so as to isolate peculiarities in the data.

The impact that an observation or a case can have on the analysis of data has received considerable attention during the recent years. A case here means a response y_i along with the associated design points or row $\underline{x}_i$ of $\underline{X}$ matrix. Recall that an observation or a case

is called influential if the important features of analysis are changed when they are deleted from the data. Cases can be influential because they correspond to outlying responses, remote points in a factor space or perhaps a combination of the two. If a case is influential because it is remote in the factor space, then it should be the most important case in the data since it may provide the only information in a region where the ability to take observation is limited. Alternatively, such a case might be deleted if it is believed that the model fit to most of the data is not appropriate in the presence of such a case.

An important class of measures of influence can be based on sample influence function. Let $\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$ be the usual least squares estimator of $\underline{\beta}$ based on the full data and let $\hat{\underline{\beta}}_{-i} = (\underline{X}^T_{-i}\underline{X}_{-i})^{-1}\underline{X}^T_{-i}\underline{Y}_{-i}$ be the corresponding estimate with the ith case deleted. Then the sample influence function for $\hat{\underline{\beta}}$ is given by

$$(n-1)^{-1}SIF_i = (\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i}) \tag{7.28}$$

The sample influence function (SIF_i) is one of the several influence functions described by Mallows (1975). Recall that the values of SIF_i are comparable only within the given data set and model. Also since SIF_i is p-vector, its routine use as a diagnostic for isolating influential cases may be laborious. Alternatively, for a given positive (semi) definite matrix M and a non-zero scale factor c, SIF_i can be characterised by the distance, $D_i(M,C)$ between $\hat{\underline{\beta}}$ and $\hat{\underline{\beta}}_{-i}$ defined by

$$D_i(M.C) = (n-1)^2 \frac{(SIF_i)^T M(SIF_i)}{C} \tag{7.29}$$

Clearly, the character of $D_i(M,C)$ is determined by M and C which may be chosen to reflect specific interests, see for instance, Cook and Weisberg (1980).

7.3.1 Distance Measures

For the linear model, described earlier in this chapter, Cook (1977, 1979) and Cook and Weisberg (1980) have discussed the use of $\hat{\underline{\beta}} - \hat{\underline{\beta}}_{-i}$ in pointing out the large residuals with the main purpose of isolating data points that might be ignored or deleted in a refit of the model. Cook's measure is based on the confidence ellipsoids for judging the contribution of each data point to the determination of the least squares estimate of $\underline{\beta}$. Thus in order to determine the influence which the ith data point has on the estimate $\hat{\underline{\beta}}$, a general distance measure using (7.29) is given, for some $\underline{M}$ and C, by

$$D_i(M,C) = \frac{(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i})^T \underline{M} (\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i})}{C} \tag{7.30}$$

This general distance measure was proposed by Cook and Weisberg (1980). A large value of $D_i(\underline{M}, C)$ would correspond to cases which when deleted would result in a large movement in the estimate of $\underline{\beta}$. We shall call a case with a large value of $D_i(\underline{M}, C)$ influential for estimating $\underline{\beta}$ relative to $D_i(\underline{M}, C)$. The natural choice for $\underline{M}$ and C are $(\underline{X}^T\underline{X})$ and ps^2 respectively. The resulting statistic suggested by Cook (1977) is

$$D_i(\underline{X}^T\underline{X}, ps^2) = \frac{(\hat{\underline{\beta}}_{-i}-\hat{\underline{\beta}})^T(\underline{X}^T\underline{X})(\hat{\underline{\beta}}_{-i}-\hat{\underline{\beta}})}{ps^2} \quad , i = 1,2,\ldots,n \; . \tag{7.31}$$

The magnitude of the distance between $\hat{\underline{\beta}}$ and $\hat{\underline{\beta}}_{-i}$ may be assessed by comparing $D_i(\underline{X}^T\underline{X}, ps^2)$ to the probability points of the central F-distribution with p and $n-p$ degrees of freedom. This is equivalent to studying the least squares confidence ellipsoids for $\underline{\beta}$ based on the full data and finding the ellipsoid that passes through $\hat{\underline{\beta}}_{-i}$, the F-distribution is only used to transform $D_i(\underline{X}^T\underline{X}, ps^2)$ to a more familiar scale. The distance measure (7.31) may also be written in an alternative form proposed by Bingham (1977).

$$D_i(\underline{X}^T\underline{X}, ps^2) = \frac{(\hat{\underline{Y}}_{-i}-\hat{\underline{Y}})^T(\hat{\underline{Y}}_{-i}-\hat{\underline{Y}})}{ps^2}, \tag{7.32}$$

suggesting that $D_i(\underline{X}^T\underline{X}, ps^2)$, aside form the scale factor ps^2, is the ordinary squared Euclidean distance that the fitted vector moves when the ith case is deleted from the data.

The distance measure (7.31) can be written in computationally convenient form as under:

$$\begin{aligned} D_i(\underline{X}^T\underline{X}, ps^2) &= \left(\frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i}{1-w_i}\right)^T \underline{X}^T\underline{X}\left(\frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i)}{1-w_i}\right)/ps^2 \\ &= \frac{\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i}{(1-w_i)p}\,\frac{R_i^2}{s^2(1-w_i)} \qquad \because \hat{\underline{\beta}}-\underline{\beta}_i = (\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i \\ &\qquad\qquad w_i = \underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i \\ &= \frac{w_i}{(1-w_i)}\,\frac{t_i^2}{p}, \end{aligned} \tag{7.33}$$

$$= \frac{w_i}{(1-w_i)p}\cdot\frac{R_i}{s(1-w_i)^{\frac{1}{2}}},$$

$$= \frac{\mathrm{Var}(\hat{\underline{Y}}_i)}{\mathrm{Var}(R_i)}\cdot\frac{t_i^2}{p} \tag{7.34}$$

$\mathrm{Var}(R_i) = \sigma^2(1-w_i)$, $\mathrm{Var}(\hat{Y}_i) = \sigma^2 w_i$, $t_i = \dfrac{R_i}{s(1-w_i)^{\frac{1}{2}}}$ is the ith studentised residual and $s^2 = \underline{Y}^T(\underline{I}-\underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T)\underline{Y}/n-p$.

We note from expression (7.34) that the quantities controlling $D_i(\underline{X}^T\underline{X}, \rho s^2)$ are the number of parameters p, the ith Studentised residual t_i and the ratio of the variances of the ith predicted value $\hat{Y}_i$ to the variance of the ith residual. It may be pointed out that t_i^2 is a measure of the degree to which the ith observation can be regarded as an outlier whereas the ratio $\mathrm{Var}(\hat{Y}_i)/\mathrm{Var}(R_i)$ measures the relative sensitivity of the estimate $\hat{\beta}$ to potential outlying values at each data point. Thus $D_i(\underline{X}^T\underline{X}, \rho s^2)$ can be large if either t_i^2 or w_i is large.

The ratio $\dfrac{w_i}{1-w_i}$ can be given several interesting interpretations. Cook (1977) noted that $\dfrac{w_i}{1-w_i} = \dfrac{\mathrm{Var}(\hat{Y}_i)}{\mathrm{Var}(R_i)}$. Weisberg (1980) noticed that $\underline{x}_i^T(\underline{X}_{-i}^T\underline{X}_{-i})^{-1}\underline{x}_i = \dfrac{w_i}{1-w_i}$, implying that it is a distance relative to the ellipsoid defined by $(\underline{X}_{-i}^T\underline{X}_{-i})$. Huber (1981) noted the relationship $\hat{Y}_i = (1-w_i)\underline{x}_i^T\hat{\beta}_{-i}+w_iY_i$ so that the ratio $\dfrac{w_i}{1-w_i}$ can be interpreted as a function of the relative weight of Y_i in determining $\hat{Y}_i$. Finally, $\dfrac{w_i}{1-w_i}$ is proportional to the total change in the variance of the prediction at $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ when $\underline{x}_i$ is deleted,

$$\frac{w_i}{1-w_i} = [\sum_j \mathrm{Var}(\underline{x}_j^T\hat{\beta}_{-i}) - \sum_j \mathrm{Var}(\underline{x}_j^T\hat{\beta})]/\sigma^2 .$$

The ith case will be called large if D_i is large; the exact defnition of large, however, depends on the problem, but D_i greater than 1, corresponding to distances between $\hat{\beta}$ and $\hat{\beta}_{-i}$ beyond 50% confidence region, usually provides a basis for comparison, see, for instance Cook and Weisberg (1982).

We have discussed above one choice of $\underline{M}$ and C but several other choices could be made. Though no detailed comparison regarding these alternative choices exist but it is hoped that any choice of $\underline{M}$ and C with $D_i(\underline{M},C)$ location/scale invariant would give approximately the same information, see for example, Cook and Weisberg (1980). Some of these choices are detailed as Table 7.1 for illustration.

Table 7.1 $D_i(\underline{M},C) = (\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i})^T \underline{M} (\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i})/C$ for different $\underline{M}$ and C .

$\underline{M}$	C	Reduced form	Source	Source
$\underline{X}^T\underline{X}$	ps^2	$\frac{1}{p} t_i^2 \frac{w_i}{1-w_i}$		Cook (1977)
$\underline{X}^T\underline{X}$	ps^2_{-i}	$\frac{n-p}{p} F_i \frac{w_i}{1-w_i}$, where $F_i = \frac{t_i^2(n-p-1)}{(n-p-t_i^2)}$		Welsch and Kuh (1977)
$\underline{X}_{-i}^T\underline{X}_{-i}$	ps^2	$\frac{1}{p} t_i^2 w_i$		
$\underline{X}_{-i}^T\underline{X}_{-i}$	ps^2_{-i}	$\frac{1}{p} F_i w_i$		
$[\text{Diag}(\underline{X}^T\underline{X})^{-1}]^{-1}$	ps^2_{-i}	$[\frac{n-p}{p} F_i \underline{x}_i^T(\underline{X}^T\underline{X})^{-1} \underline{M}(\underline{X}^T\underline{X})^{-1}\underline{x}_i]/1-w_i$		Welsch and Kuh (1977)
$\underline{I}$	ps^2	$t_i^2 \underline{x}_i^T(\underline{X}^T\underline{X})^{-2}\underline{x}_i/1-w_i$		

Source: Cook and Weisberg (1980).

The importance and usefulness of t_i^2 and w_i's in the analysis of data have been studied by many researchers, for example, Srikantan (1961), Welsch and Kuh (1977), Hoaglin and Welsch (1978), Belsley, Kuh and Welsch (1980) and Weisberg (1980). Huber (1975) notes that the large values of w_i correspond to "outlying" design points and suggests that it will be difficult to spot outlying observations if max w_i is not considerably small. Davies and Hutton (1975) note that if $w_i > 0.2$ it is possible for a moderate error in the corresponding observation to affect the estimates significantly and yet go undetected when the residuals are checked. Box and Draper (1975) suggest that for a design to be insenstive to outliers, the w_i should be constant.

7.3.1.1 Reasons for Large w_i and its Implications

The characteristic of $\underline{x}_i^T$ which cause w_i to be large were studied by Cook and Weisberg (1980). Assuming the model with the intercept, let $\mu_1 \leq \mu_2 \leq,\ldots,\leq\mu_{p-1}$ denote the eigenvalues of the corrected cross products matrix for the data and $\underline{P}_1,\ldots,\underline{P}_{p-1}$ denote thecorresponding eigenvectors. Then by the spectral decomposition of the corrected cross product matrix,

$$w_i = \frac{1}{n} + \sum_{\ell=1}^{p-1} \left(\frac{\underline{P}_\ell^T(\underline{x}_\ell - \bar{\underline{x}})}{\sqrt{u_\ell}} \right)^2 \qquad (7.35)$$

where $\bar{\underline{x}}$ is the vector of sample averages. Further let $\theta_{\ell i}$ denote the angle between $\underline{P}_\ell$ and $(\underline{x}_i - \bar{\underline{x}})$, then

$$\cos(\theta_{\ell i}) = \frac{\underline{P}_\ell^T(\underline{x}_i - \bar{\underline{x}})}{[(\underline{x}_i - \bar{\underline{x}})^T(\underline{x}_i - \bar{\underline{x}})]^{\frac{1}{2}}} \qquad (7.36)$$

and

$$w_i = \frac{1}{n} + (\underline{x}_i - \bar{\underline{x}})^T (\underline{x}_i - \bar{\underline{x}}) \sum_{\ell=1}^{k-1} \frac{\cos^2 \theta_{\ell i}}{u_\ell} \qquad (7.37)$$

Thus, we note that w_i is large if $\underline{x}_i$ is far from $\bar{\underline{x}}$, that is, it is sufficiently away from most of the cases and also if $\underline{x}_i$ is substantially in a direction of an eigenvector corresponding to a small eigenvalue of the corrected cross product. On the other hand, if $\underline{x}_i - \bar{\underline{x}}$ is small, w_i will be small regardless of its direction. Contours of $\underline{x}_i$ are ellipsoid centered at $\bar{\underline{x}}$ with axis given by the eigen structure of $(\underline{X}^T\underline{X})^{-1}$.

7.3.2 Distance Measures Based on Internal Scaling

We have so far discussed the distance measures that correspond to using confidence contours to order the values of the sample influence function. These distance measures are based on fixed metrics that do not depend on the observed behaviour of sample versions of the influence function. Of course these measures do depend on the expected behaviour of the data so far as $(\underline{X}^T\underline{X})^{-1}$ or other related inner product matrices accurately represent the variance of $\hat{\underline{\beta}}$. In contrast, distance measures that are based on a matrix derived from the observed values of the appropriate sample version of the influence function may also be constructed. These distance measures are robust with respect to variations in the model or methods of analysis that would necessitate different distance measures. For example, if the model were altered to have $\mathrm{Var}(\underline{e}) = \sigma^2 \underline{V}^{-1}$, where $\underline{V}$ is known, then to be consistent with previous rationale the inner product matrix $\underline{X}^T\underline{X}$ for the distance measures discussed earlier should be changed to $\underline{X}^T\underline{V}\underline{X}$.

In this section we discuss the distance measures that use the multivariate outlier technique to order the values and the jack-knife method to choose $\underline{M}$ and C.

7.3.2.1 Ordering Using a Multivariate Outlier Statistic

The method particularly useful for study of the n values of SIF is Wilk's (1963) criterion for detecting a single outlier in a multivariate sample. Let $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$ be p-vectors and define $\bar{\underline{b}} = n^{-1} \sum_{i=1}^{n} \underline{b}_i$ and $\Delta = \Sigma(\underline{b}_i - \bar{\underline{b}})(\underline{b}_i - \bar{\underline{b}})^T$. Wilk's criterion selects $\underline{b}_i$ as possible outlier if i minimises

$$\frac{|\Delta_{-i}|}{|\Delta|} \tag{7.38}$$

where Δ_{-i} denotes that the ith case has been deleted. Since $|\Delta|$ is proportional to the square of the volume of a p-dimensional ellipsoid, minimising this ratio is equivalent to choosing $\underline{b}_i$ to minimise the volume remaining after $\underline{b}_i$ is deleted. This implies that in some sense $\underline{b}_i$ must be far from the other vecotrs. Using the results on determinants from Henderson and Searle (1981), it could be shown that minimising (7.38) is equivalent to maximising the distance

$$\delta_i = (\underline{b}_i - \bar{\underline{b}})^T \Delta^{-1} (\underline{b}_i - \bar{\underline{b}}) , \quad i = 1,2,\ldots,n \tag{7.39}$$

In case of linear least squares regression, explicit formulae for δ_i can be obtained for any of the empirical versions of the influence function discussed earlier in this chapter.

Consider, for example, the empirical influence function (7.22) it is sufficient to take $\underline{b}_i = (\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i$ and thus $\bar{\underline{b}} = 0$. The inner product matrix, say Δ_o, is given by

$$\Delta_o = \sum_j \underline{b}_j \underline{b}_j^T = (\underline{X}^T\underline{X})^{-1} [\sum_j R_j^2 \underline{x}_j \underline{x}_j^T] (\underline{X}^T\underline{X})^{-1} \tag{7.40}$$

Now $\frac{n}{n-p} \Delta_o$ is the robust variance estimator $\hat{V}_w$ of $Var(\hat{\underline{\beta}})$ obtained by Hinkley (1977b) using the weighted jack-knife (see Chapter 8 for details). Substituting (7.40) in (7.39), the corresponding distance measure is

$$\delta_i^o = (\underline{b}_i - \bar{\underline{b}})^T \Delta_o^{-1} (\underline{b}_i - \bar{\underline{b}})$$

$$= R_i^2 \underline{x}_i^T [\sum_j R_j^2 \underline{x}_j \underline{x}_j^T]^{-1} , \quad i = 1,2,\ldots,n \tag{7.41}$$

For the sample influence function defined in (7.27), we can take $\underline{b}_i = (\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i/(1-w_i)$. Since in this case $\bar{\underline{b}}$ is not zero, the form (7.39) does not simplify. The cross product matrix, say Δ_1 , is

$$\Delta_1 = (\underline{b}_j - \bar{\underline{b}})(\underline{b}_j - \bar{\underline{b}})^T$$

$$= (\underline{X}^T\underline{X})^{-1} \left(\sum_j \frac{R_j^2}{(1-w_j)^2} \underline{x}_j \underline{x}_j^T - n\bar{\underline{z}}\bar{\underline{z}}^T \right) (\underline{X}^T\underline{X})^{-1} \tag{7.42}$$

where $n\bar{\underline{z}} = \Sigma \underline{x}_j R_j/(1-w_j)$. Recall that matrix $(n-1)\frac{\Delta_1}{n}$ is the estimate of $\mathrm{Var}(\hat{\underline{\beta}})$ obtained from the usual unweighted jack-knife (see Chapter 6, Section 6.4). The corresponding distance measure δ_i^1 can be obtained exactly, but some algebraic simplification is achieved if usually small correction for the centre $\bar{\underline{z}}$ is ignored. Hence putting $\bar{\underline{z}} = 0$ and substituting (7.42) in (7.39)

$$\delta_i^1 = \frac{R_i^2}{(1-w_i)^2} \underline{x}_i^T \left[\sum \frac{R_j^2}{(1-w_j)^2} \underline{x}_j \underline{x}_j^T \right]^{-1} \underline{x}_i \tag{7.43}$$

This measure can be computed as the diagonal elements of the projection on the columns of $\underline{V}^{\frac{1}{2}}\underline{X}$, where the n x n diagonal matrix $\underline{V}$ has diagonal elements $\frac{R_i^2}{(1-w_i)^2}$.

7.3.2.2 The Jack-knife Method

As pointed out earlier in this section, the jack-knife method can also be used to provide an alternative distance measure. Recall that in the usual unweighted jack-knife method, estimates are obtained by

averaging n pseudo-estimates each obtained by deleting one case at a time. In most cases, the jack-knifed estiamtes of parameters and variances are known to have desirable properties. For example, Hinkley (1977b) suggests that the weighted jack-knifed variance estimate $\tilde{V}_{WJ} = \frac{n}{n-p}\Delta_o$ of $Var(\hat{\underline{\beta}})$ is robust against the non-homogeneity of error variances. This suggests the use of $J_i \equiv D_i(\Delta_o^{-1}, \frac{pn}{n-p})$ as an alternative to $D_i(\underline{X}^T\underline{X}, ps^2)$. The interpretation of J_i is the same as that of D_i except that this adjustment provides a metric that should be more robust.

$$J_i \equiv D_i(\Delta_0^{-1}, \frac{pn}{n-p}) = \frac{n-p}{np}\frac{R_{i\cdot i}^2 \underline{x}_i^T[\sum_j R_{j\cdot j}^2 \underline{x}_j\underline{x}_j^T]^{-1}\underline{x}_i}{(1-w_i)^2} \tag{7.44}$$

$$= \frac{n-p}{np}\frac{\delta_i^o}{(1-w_i)^2}$$

The distance measure J_i provides a compromise between δ_i^o and δ_i^1. Besides, the interpretation of J_i as a robust version of D_i has some appeal.

7.4 Concluding Remarks

In this chapter, we have discussed the influence function and its role in the assessment of influential observations in regression analysis. We note that the influence function serves as a useful tool in exhibiting the sensitivity of an estimator to the observed values. It allows us to modify estimation procedures so that they do not depend on influential observations or any other specific features of the observations. The different distance measures help to provide information concerning the reliability of the conclusions and their dependence on the model. The ability to find influential cases is helpful for the analysis in making decisions regarding their usefulness in a particular problem under study.

It is also interesting to note that the jack-knife method provides a robust version of the distance measure D_i . Further investigation in this direction may be interesting.

CHAPTER 8

THE WEIGHTED JACK-KNIFED ESTIMATION FOR THE LINEAR MODEL

8.1 Introduction

Miller (1974a) used the standard jack-knifed technique for his results on regression models which we have discussed under the title of "Unweighted Jack-knife for Regression Estimation" in Chapter 6. The pseudovalues in the unweighted jack-knife are defined symmetrically with respect to the observations whereas in the linear model situations the observations do not affect the least squares estimates in a symmetric manner. For example, heterogeneity of error variances renders the standard jack-knife technique inadequate in such stiuations. More importantly, as we have seen in Chapter 7, a non-homogeneous spacings of the observations cause a wide variation in the variances of the residuals. This draws attention to the imbalance in the design matrix $D_o = \underline{X}^T\underline{X}$. This situation demands a more general definition of the jack-knife and it was in order to adjust for such imbalances that Hinkley (1977b) proposed weighted pseudovalues and hence the weighted jack-knife.

In this chapter, we discuss the application of the weighted jack-knife to regression estimation. First, in Section 8.2, we give a suitable definition and justification for the weighted jack-knife. Earlier, in Chapter 6 we have seen that the unweighted jack-knife regression estimator $\tilde{\beta}$ is unbiased but less efficient than the least squares estimator $\hat{\beta}$. In Section 8.3, we examine the properties of the weighted jack-knifed regression estimator $\tilde{\beta}_{WJ}$ and the weighted jack-knifed variance estimator $\tilde{V}_{WJ}$. An important aspect

of a statistical procedure is its insensitivity to departures from the usual assumptions. This important aspect of the weighted jack-knife is also discussed in this section. Miller (1974a) and Hinkley (1977b) assume constant variances for the random variables. This, however, may not be true in many situations. In Section 8.4 we consider a simple generalisation of the weighted jack-knife for a more general variance structure $\sigma^2\underline{V}$. A theorem is presented on the bias of the weighted jack-knife variance estimator V^*_{WJ} under this more general model. Finally, we give a summary and conclusions in Section 8.5.

8.2 The Weighted Jack-knife

The unweighted jack-knife, as discussed in Chapter 6, assumes equal sample sizes and equal variances and deals with symmetric statistics. The pseudo-values are defined symmetrically with respect to the observations $(\underline{x}_i, Y_i)$ whereas the model considered may be unbalanced. The lack of balance is reflected in the distances w_i, see for example, Chapter 6 (expression 6.18).

In situations where $\underline{x}_i$ are samples from a multivariate normal population, the estimated likelihood of the value $\underline{x}_i$ is a decreasing function of w_i. Moreover, the $\mathrm{Tr}\{\mathrm{Var}(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i})\}$ is an increasing function of w_i, see for example, Hinkley (1977b). This last result is expressed as Lemma 8.1.

Lemma 8.1

Consider the model $\underline{Y} = \underline{X}\underline{\beta} + \underline{e}$ described in (6.8). Let $\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$ be the estimator of $\underline{\beta}$ based on the complete data set and $\hat{\underline{\beta}}_{-i}$ be the corresponding estimator of $\underline{\beta}$ obtained by successively deleting the ith row $\underline{x}_i^T$ from the matrix $\underline{X}$ and Y_i from the vecotr $\underline{Y}$. Then $\mathrm{Tr}\{\mathrm{Var}(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i})\}$ is an increasing function of w_i.

Proof

From expression (6.12) we have

$$\hat{\underline{\beta}} - \hat{\underline{\beta}}_{-i} = \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i R_i}{1-w_i} \quad ,$$

$$\text{Var}\,(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i}) = \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i\ \text{Var}(R_i)\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}}{(1-w_i)} \quad ,$$

$$= \frac{(\underline{X}^T\underline{X})^{-1}\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}}{(1-w_i)}\sigma^2, \quad \because \text{Var}(R_i) = (1-w_i)\sigma^2$$

$$\text{Tr}\{\text{Var}(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i})\} = \text{Tr}\ \frac{[(\underline{X}^T\underline{X})^{-1}\underline{x}_i\underline{x}_i^T(\underline{X}^T\underline{X})^{-1}]}{(1-w_i)}\sigma^2 \quad , \qquad (8.1)$$

which is an increasing function of w_i .

Thus we observe that the contribution of the ith observation to the jack-knifed estimator $\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i}$ has a weight decreasing in w_i. Hinkley (1977b) suggests that a specific choice of wieght is indicated by the fact that

$$n(1-w_i)(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i}) = nD_o^{-1}\underline{x}_i R_i = \text{EIF}(\underline{\beta}:\underline{x}_i^T, Y_i) \qquad (8.2)$$

is the empirical influence function of $\underline{\beta}$ at $(\underline{x}_i^T, Y_i)$, see for example, Lemma 1 (Hinkley, 1977b) which says that the influence function at $(\underline{x}, y)$ is approximately proportional to the incremental change in $\underline{\beta}$ when a very small fraction of the measurement population is moved to $(\underline{x}, y)$. Further details on this subject could be seen in Chapter 7, Section 7.2.

Thus, it follows from the above formulation that the contribution of the ith observation can be very important in estimating parameters.

Recall that in multiple regression w_i measures the distance from the point $\underline{x}_i$ to the centre of the data and the cases with the unusual values of $\underline{x}_i$ will tend to have large w_i. Since $\text{Var}(R_i) = (1-w_i)\sigma^2$, therefore, a wide variation will exist in the variance of the residuals. This wide variation in the variance of residuals indicates a peculiarity of the $\underline{X}$ matrix, that is, a non-homogeneous spacing of the observations and thus directs attention to data deficiencies. The standard jack-knife technique is inadequate in such situations because it does not take into consideration the unbalanced structure of the data and the heterogeneity of the variances.

In order to tackle such situations, Hinkley (1977b) proposed a weighted jack-knife procedure for the regression estimation. The basic idea of the weighted jack-knife is essentially due to Quenouille. The essence of the weighted jack-knife lies in the weighted pseudovalues which are defined in terms of the weights w_i. These weighted pseudovalues take into consideration the non-homogeneous spacings between the design points. The weighted jack-knife may, therefore, be defined as that dealing with the unbalanced or non-symmetric statistics.

8.3 Applications to Regression Estimation

In Chapter 6 (Section 6.4), we discussed the application of the unweighted jack-knife technique to regression estimation. In this section, we study the consequences of the weighted jack-knifed procedure on regression estimation. Following Hinkley (1977b) we define the weighted pseudovalues

$$\begin{aligned}\tilde{Q}_i &= \hat{\underline{\beta}} + n(1-w_i)(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{-i}) \\ &= \hat{\underline{\beta}} + nD_o^{-1}\underline{x}_i R_i \qquad (8.3)\end{aligned}$$

where, as before, $D_o = (\underline{X}^T\underline{X})$, $R_i = (Y_i - \underline{x}_i^T\hat{\underline{\beta}})$ and $w_i = \underline{x}_i^T(\underline{X}^T\underline{X})^{-1}\underline{x}_i$.

The weighted jack-knifed regression estimator $\tilde{\beta}_{WJ}$ is the average of the weighted pseudovalues $\tilde{Q}_i$ and is given by

$$\tilde{\beta}_{WJ} = n^{-1}\sum_{i=1}^{n}\tilde{Q}_i = n^{-1}\sum_{i=1}^{n}(\hat{\underline{\beta}}+nD_o^{-1}\underline{x}_iR_i)$$

$$= \hat{\underline{\beta}} + D_o^{-1}\sum_{i=1}^{n}\underline{x}_iR_i$$

$$= \hat{\underline{\beta}} \qquad \because \sum_{i=1}^{n}\underline{x}_iR_i = 0 \qquad (8.4)$$

$$E(\tilde{\beta}_{WJ}) = E(\hat{\underline{\beta}}) = \underline{\beta} \qquad (8.5)$$

$$Var(\tilde{\beta}_{WJ}) = Var(\tilde{\underline{\beta}}) = (\underline{X}^T\underline{X})^{-1}\sigma^2 = \sigma^2D_o^{-1} \qquad (8.6)$$

Thus we note that the weighted jack-knifed estimator $\tilde{\beta}_{WJ}$ possesses the reproducing property as given by (8.4) and (8.6). In comparison to this it could be seen from expressions (6.21) and (6.24) that the unweighted jack-knifed estimator $\tilde{\underline{\beta}}$ does not possess these properties. The reproduction properties of $\tilde{\beta}_{WJ}$ may result in a better performance of $\tilde{\beta}_{WJ}$ in the non-linear cases.

8.3.1 Variance Estimation

The weighted jack-knifed variance estimate is given by

$$\tilde{V}_{WJ} = \frac{1}{n(n-p)}\sum_{i=1}^{n}(\tilde{Q}_i-\tilde{\beta}_{WJ})(\tilde{Q}_i-\tilde{\beta}_{WJ})^T$$

$$= \frac{1}{n(n-p)}\sum_{i=1}^{n}\left\{\hat{\underline{\beta}}+nD_o^{-1}\underline{x}_iR_i-\tilde{\beta}_{WJ}\right\}\left\{\hat{\underline{\beta}}+nD_o^{-1}\underline{x}_iR_i-\tilde{\beta}_{WJ}\right\}$$

$$= \frac{1}{n(n-p)}\left[\sum_{i=1}^{n}(nD_o^{-1}\underline{x}_iR_i)(nD_o^{-1}\underline{x}_iR_i)^T\right]$$

$$= \frac{n}{n-p}D_o^{-1}\left(\sum_{i=1}^{n}R_i^2\underline{x}_i\underline{x}_i^T\right)D_o^{-1} \qquad (8.7)$$

where $n-p$ indicates the degrees of freedom in the residual vector.

We now work out the expected value of $\tilde{V}_{WJ}$ to examine its unbiasedness. This result by Hinkley is expressed as Theorem 8.1.

Theorem 8.1

For the linear model (6.8) the weighted jack-knifed variance estimator $\tilde{V}_{WJ}$ is biased. The magnitude of this bias is of $O(n^{-2})$, that is, $E(\tilde{V}_{WJ}) - Var(\tilde{\beta}_{WJ}) = O(n^{-2})$.

Proof

Using expression (8.7), we have

$$E(\tilde{V}_{WJ}) = \frac{n}{n-p} D_o^{-1} E\left[\sum_{i=1}^{n} R_{ii}^2 \underline{x}_i \underline{x}_i^T\right] D_o^{-1} ,$$

$$= \frac{n}{n-p} D_o^{-1} \sigma^2 \left[\sum_{i=1}^{n} (1-\underline{x}_i^T D_o^{-1} \underline{x}_i) \underline{x}_i \underline{x}_i^T\right] D_o^{-1} ,$$

$$= \frac{n}{n-p} \sigma^2 (D_o^{-1} D_o D_o^{-1} - D_o^{-1} D_1^* D_o^{-1}) ,$$

$$= \frac{n}{n-p} (D_o^{-1} - D_o^{-1} D_1^* D_o^{-1}) \sigma^2 \qquad (8.8)$$

where $\sum_{i=1}^{n} \underline{x}_i \underline{x}_i^T = \underline{X}^T \underline{X} = D_o$ and D_m^* is defined in (6.40). Thus, we observe that $\tilde{V}_{JW}$ is biased in unbalanced cases. The magnitude of the bias of $\tilde{V}_{WJ}$ is as given under:

$$E(\tilde{V}_{WJ}) - Var(\tilde{\beta}_{WJ}) = \frac{n}{n-p} \sigma^2 (D_o^{-1} - D_o^{-1} D_1^* D_o^{-1}) - \sigma^2 D_o^{-1} ,$$

$$\doteq -\sigma^2 D_o^{-1} D_1^* D_o^{-1} , \text{ for large } n ,$$

$$= L^{-1} (O(1)) L^{-1}$$

$$= O(n^{-2}) . \qquad (8.9)$$

This shows that the bias decreases at a faster rate than $1/n$.

However, when $w_i = pn^{-1}$, that is, in the balanced case $\tilde{V}_{WJ}$ is exactly unbiased. We illustrate this property as Corollary 8.1.

Corollary 8.1

If $w_i = \frac{p}{n}$, then $E(\tilde{V}_{WJ}) = Var(\tilde{\beta}_{WJ}) = Var(\hat{\underline{\beta}})$.

Proof

From(8.8),we have

$$E(\tilde{V}_{WJ}) = \frac{n}{n-p}\sigma^2 D_o^{-1}(D_o - D_1^*)D_o^{-1} ,$$

$$= \frac{n}{n-p}\sigma^2 D_o^{-1}\left[\sum_{i=1}^{n} \underline{x}_i\underline{x}_i^T - \sum_{i}^{n} w_i \underline{x}_i\underline{x}_i^T\right]D_o^{-1} ,$$

$$= \frac{n}{n-p}\sigma^2 D_o^{-1}\left[\sum_{i=1}^{n} \underline{x}_i\underline{x}_i^T - \sum_{i=1}^{n} \frac{p}{n}\underline{x}_i\underline{x}_i^T\right]D_o^{-1} ,$$

$$= \frac{n}{n-p}\sigma^2 D_o^{-1}\left[\frac{n-p}{n}\sum_{i=1}^{n} \underline{x}_i\underline{x}_i^T\right]D_o^{-1} ,$$

$$= \sigma^2 D_o^{-1} \qquad (8.10)$$

$$= Var(\tilde{\beta}_{WJ}) \qquad (8.11)$$

$$= Var(\hat{\underline{\beta}}) \qquad (8.12)$$

Thus we note that, for regression estimation, the weighted jack-knife approach gives better performance than the unweighted jack-knife approach discussed earlier in Chapter 6, Section 6.4.

8.3.2 Robustness Properties of the Weighted Jack-knife Variance Estimator $\tilde{V}_{WJ}$

In a general sense, a statistical procedure is described as robust if it is not very sensitive to departures from assumptions on which

it depends. Robustness is thus concerned with the general applicability of statistical procedures. The underlying idea being to see how large is the domain of applicability of a given statistical procedure. Recall that the basic assumptions of the model (6.8) are that the error variables e_i are independently and identically distributed with mean zero and constant variance σ^2. This assumption of common variances is known as homoscedasticity. In practice, however, departures from this assumption may occur and the error variances may not be homogeneous. The consequences of the heterogeneity of the variances are:

(i) that the ordinary least squares estimates of regression parameters are still unbiased but inefficient.

(ii) the ordinary least squares estimates of variance are biased.

We illustrate these points through Lemma 8.2 and then following Hinkley (1977b) we demonstrate the robustness of the weighted jack-knifed variance estimator $\tilde{V}_{WJ}$ through Lemma 8.3.

Lemma 8.2

Under the model $\underline{Y} = \underline{X}\beta + \underline{e}$, $E(\underline{e}) = 0$ and $Var(\underline{e}) = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2) = \Lambda$, the variance of the least squares estimator $\hat{\underline{\beta}}$ is given by

$$Var(\hat{\underline{\beta}}) = D_o^{-1}\underline{X}^T\Lambda\underline{X}D_o^{-1} \quad . \tag{8.13}$$

where $D_o = (\underline{X}^T\underline{X})$.

Proof

$$\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$$

$$E(\hat{\underline{\beta}}) = \underline{\beta} + (\underline{X}^T\underline{X})^{-1}\underline{X}^TE(\underline{e})$$

$$= \underline{\beta} \quad . \tag{8.14}$$

$$\text{Var}(\hat{\underline{\beta}}) = E(\hat{\underline{\beta}}-\underline{\beta})(\hat{\underline{\beta}}-\underline{\beta})^T ,$$

$$= E[(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{e}][(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{e}]^T ,$$

$$= (\underline{X}^T\underline{X})^{-1}\underline{X}^T E(\underline{e}\,\underline{e}^T)\underline{X}(\underline{X}^T\underline{X})^{-1} ,$$

$$= (\underline{X}^T\underline{X})^{-1}\underline{X}^T \Lambda\, \underline{X}(\underline{X}^T\underline{X})^{-1} ,$$

$$= D_o^{-1}\underline{X}^T \Lambda\, \underline{X} D_o^{-1} \qquad (8.15)$$

If we ignore the heteroscedasticity, then the usual estimate is given by

$$\hat{V} = \frac{1}{n-p} \sum_{j=1}^{n} R_j^2 D_o^{-1} \qquad (8.16)$$

Since $E(R_j^2) \doteq \sigma_j^2$, the expected value of $\hat{V}$ is given by

$$E(\hat{V}) = \frac{1}{n-p} E \sum_{j=1}^{n} R_j^2 D_o^{-1} ,$$

$$= \frac{1}{n-p} \sum_{j=1}^{n} \sigma_j^2 D_o^{-1} ,$$

$$= \frac{1}{n-p} (\text{Tr } \Lambda) D_o^{-1} , \quad \because \text{Tr } \Lambda = \sum_{j=1}^{n} \sigma_j^2$$

$$\doteq \frac{1}{n} (\text{Tr } \Lambda) D_o^{-1} . \qquad (8.17)$$

Thus we would be estimating the variance of $\hat{\underline{\beta}}$ by an expression whose expected value is $\frac{1}{n}(\text{Tr } \Lambda)D_o^{-1}$ whereas the true variance is given by (8.15). This implies that in the case of heteroscedasticity even the estaimted variances are biased.

The weighted jack-knifed variance estimate $\tilde{V}_{WJ}$ is superior to the usual estimate $\hat{V}$ in the sense that it is robust against the non-homogeneity of the error variances. This property is described as Lemma 8.3.

Lemma 8.3

Under the conditions of model in Lemma 8.2, the expected value of the weighted jack-knifed variance estimate $\tilde{V}_{WJ}$ is given by

$$E(\tilde{V}_{WJ}) = \frac{n}{n-p} D_o^{-1} \underline{X}^T \Lambda \underline{X} D_o^{-1} \quad . \tag{8.18}$$

Proof

$$E(\tilde{V}_{WJ}) = \frac{n}{n-p} D_o^{-1} (E \sum_{j=1}^{n} R_j^2 \underline{x}_j \underline{x}_j^T) D_o^{-1} \quad ,$$

$$\doteq \frac{n}{n-p} D_o^{-1} \sum_{j=1}^{n} \sigma_j^2 \underline{x}_j \underline{x}_j^T D_o^{-1} \quad , \quad \because E(R_j^2) \doteq \sigma_j^2 \quad .$$

$$\doteq D_o^{-1} \underline{X}^T \Lambda \underline{X} D_o^{-1} \quad , \tag{8.19}$$

$$= \mathrm{Var}(\hat{\underline{\beta}}) \tag{8.20}$$

It is apparent from expression (8.19) that $\tilde{V}_{WJ}$ approximates the true variance of $\hat{\underline{\beta}}$ given by (8.15) when the error variances are unequal whereas the usual estimate $\hat{V}$ lacks this property. This reflects the superior performance of the weighted jack-knifed variance estimate $\tilde{V}_{WJ}$.

8.4 Generalization of Weighted Jack-knife

We have so far considered the model $\underline{Y} = \underline{X}\underline{\beta} + \underline{e}$, where $\underline{X}$ is an $n \times p$ matrix with rank p . For this model we assumed that the error variables e_i are uncorrelated. We used the ordinary least squares method as the estimation procedure which under the conditions of this model is considered to be the appropriate one. However, in practice, the above conditions may not always hold. For example, it may happen that the variances of the observations are not all equal, that is, $\mathrm{Var}(\underline{e})$ is not of the form $\sigma^2 I$ but may have the form $\mathrm{Var}(\underline{e}) = \Lambda = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. We have seen, earlier in Lemma 8.2, that the

least squares estimator $\hat{\underline{\beta}}$ still gives an unbiased estimate of $\underline{\beta}$ but $Var(\hat{\underline{\beta}})$ is biased. We also come across situations where the off-diagonal elements of $Var(\underline{e})$ are not zero, that is, the observations may be correlated. When either of these situations occur, the ordinary least squares method of estimation may not be appropriate and hence it becomes necessary to modify the procedure for obtaining estimates. Under such situations the generalised least squares may be the appropriate procedure. In this section, we study the behaviour of the weighted jackknife under the generalised least squares estimation procedure. Consider the model,

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{e} \ , \tag{8.21}$$

with $E(\underline{Y}) = \underline{X}\underline{\beta}$ and $Cov(\underline{Y}) = \sigma^2\underline{V}$ where $\underline{V}$ is a known $n \times n$ positive definite matrix.

The pseudovalues are now given by

$$\underline{\beta}^*_i = \underline{\beta}^* + n(1-K_i)(\underline{\beta}^* - \underline{\beta}^*_{-i}) \ . \tag{8.22}$$

where $K_i = \dfrac{\underline{x}_i^T \underline{C}_o^{-1} \underline{x}_i}{V_i}$, $\underline{C}_o = (\underline{X}^T\underline{V}^{-1}\underline{X})$, $\underline{\beta}^* = (\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{V}^{-1}\underline{Y}$

$\underline{\beta}^*_{-i} = (\tilde{\underline{X}}^T\tilde{\underline{V}}^{-1}\tilde{\underline{X}})^{-1}\tilde{\underline{X}}^T\tilde{\underline{V}}^{-1}\tilde{\underline{Y}}$, $(\tilde{\underline{X}}^T\tilde{\underline{V}}^{-1}\underline{X})^{-1} = (\underline{X}^T\underline{V}^{-1}\underline{X} - \underline{x}_i V_i^{-1}\underline{x}_i)^{-1}$ and

$(\tilde{\underline{X}}^T\tilde{\underline{V}}^{-1}\tilde{\underline{Y}}) = (\underline{X}^T\underline{V}^{-1}\underline{Y} - \underline{x}_i V_i^{-1} Y_i)$

We can write

$$\underline{\beta}^*_{-i} = (\underline{X}^T\underline{V}^{-1}\underline{X} - \underline{x}_i V_i^{-1}\underline{x}_i^T)(\underline{X}^T\underline{V}^{-1}\underline{Y} - \underline{x}_i V_i^{-1} Y_i) \ ,$$

$$= (\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{V}^{-1}\underline{Y} + \frac{(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{x}_i V_i^{-1}\underline{x}_i^T(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{V}^{-1}\underline{Y}}{1-\underline{x}_i^T(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{x}_i V_i^{-1}}$$

$$- (\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{x}_i V_i^{-1} Y_i - \frac{(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{x}_i V_i^{-1}\underline{x}_i^T(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{x}_i V_i^{-1} Y_i}{1-\underline{x}_i^T(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{x}_i V_i^{-1}} \ ,$$

$$= \underline{\beta}^* - \frac{(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{x}_i V_i^{-1}(Y_i - \underline{x}_i^T\hat{\underline{\beta}}^*)}{1-K_i} \ ,$$

$$= \underline{\beta}^* - \frac{\underline{C}_o^{-1}\underline{x}_i V_i^{-1} R_i^*}{1-K_i} \ .$$

Hence

$$\underline{\beta}^* - \underline{\beta}^*_{-i} = \frac{\underline{C}_o^{-1}\underline{x}_i V_i^{-1} R_i^*}{1-K_i}, \quad \text{where} \quad R_i^* = Y_i - \underline{x}_i^T \underline{\beta}^* \tag{8.23}$$

The weighted jack-knifed estimator $\underline{\beta}^*_{WJ}$ is the average of the pseudovalues and is given by

$$\beta^*_{WJ} = \frac{1}{n}\sum_{i=1}^{n} \underline{\beta}^*_{-i} = \frac{1}{n}\sum_{i=1}^{n}\left(\underline{\beta}^* + n(1-K_i)\frac{\underline{C}_o^{-1}\underline{x}_i V_i^{-1} R_i^*}{1-K_i}\right)$$

$$= \beta^* + n\underline{C}_o^{-1}\sum_{i=1}^{n}\underline{x}_i V_i^{-1} R_i^*$$

$$= \underline{\beta}^* \tag{8.24}$$

$$E(\underline{\beta}^*_{WJ}) = E(\underline{\beta}^*)$$

$$= (\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{V}^{-1}\underline{X}\underline{\beta} \qquad \because E(\underline{Y}) = \underline{X}\underline{\beta}$$

$$= \underline{\beta} \tag{8.25}$$

$$\text{Var}(\underline{\beta}^*_{WJ}) = \text{Var}(\underline{\beta}^*)$$

$$= (\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\sigma^2$$

$$= \underline{C}_o^{-1}\sigma^2 \tag{8.26}$$

It follows from the above results that the weighted jack-knifed estimate of $\underline{\beta}$ under the conditions of the model (8.21) is identical to the generalised least squares estimator $\underline{\beta}^*$. It also possesses the reproducing properties which may result in the better performance of $\underline{\beta}^*_{WJ}$.

8.4.1 Variance Estimation

The weighted jack-knifed variance estimate is now given by

$$V^*_{WJ} = \frac{1}{n(n-p)} \sum_{i=1}^{n} (\underline{\beta}^*_i - \underline{\beta}^*_{WJ})(\underline{\beta}^*_i - \underline{\beta}^*_{WJ})^T .$$

$$= \frac{n^2}{n(n-p)} \sum_{i=1}^{n} (\underline{C}_o^{-1} \underline{x}_i V_i^{-1} R^*_i)(\underline{C}_o^{-1} \underline{x}_i V_i^{-1} R^*_i)^T .$$

$$= \frac{n}{n-p} \underline{C}_o^{-1} \left(\sum_{i=1}^{n} \underline{x}_i V_i^{-1} R^{*2}_i V_i^{-1} \underline{x}_i^T \right) \underline{C}_o^{-1} .$$

$$= \frac{n}{n-p} \underline{C}_o^{-1} \left(\sum_{i=1}^{n} \frac{\underline{x}_i R^* \underline{x}_i^T}{V_i^2} \right) \underline{C}_o^{-1} \qquad (8.27)$$

where $n-p$ indicate the degrees of freedom in the residual vector. We now examine the bias of V^*_{WJ} . This result is presented as Theorem 8.2.

Theorem 8.2

Under the model (8.21) the bias of the weighted jack-knifed variance estimator V^*_{WJ} is of $O(n^{-1})$, that is,

$$E(V^*_{WJ}) - Var(\underline{\beta}^*_{WJ}) = O(n^{-1}) \qquad (8.28)$$

Proof

$$E(V^*_{WJ}) = \frac{n}{n-p} \underline{C}_o^{-1} \sum_{i=1}^{n} \left(\frac{\underline{x}_i E(R^{2*}_i) \underline{x}_i^T}{V_i^2} \right) \underline{C}_o^{-1} \qquad (8.29)$$

Now

$$Var(\underline{R}^*) = Var[\underline{Y} - \underline{X}\underline{\beta}^*] = Var[\underline{Y} - \underline{X}(\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \underline{X}^T \underline{V}^{-1} \underline{Y}],$$

$$= (\underline{I} - \underline{X}(\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} \underline{X} \underline{V}^{-1}) \sigma^2 \underline{V} .$$

Therefore,

$$\begin{aligned} \mathrm{Var}(R_i^*) &= (1-\underline{x}_i(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{x}_i^T\underline{V}_i^{-1})\sigma^2 V_i \\ &= (1-K_i)\sigma^2 V_i \end{aligned} \tag{8.30}$$

Now substitution of (8.30) in (8.29) gives

$$\begin{aligned} E(V_{WJ}^*) &= \frac{n}{n-p}\underline{C}_o^{-1}\left(\sum_{i=1}^{n}\frac{\underline{x}_i(1-K_i)\sigma^2 V_i\underline{x}_i^T}{V_i^2}\right)\underline{C}_o^{-1} \\ &= \frac{n}{n-p}\sigma^2\underline{C}_o^{-1}\left(\sum_{i=1}^{n}\frac{(1-K_i)\underline{x}_i\underline{x}_i^T}{V_i}\right)\underline{C}_o^{-1} \\ &= \frac{n}{n-p}\sigma^2\,\underline{C}_o^{-1}\left(\sum_{i=1}^{n}\underline{x}_i V_i^{-1}\underline{x}_i^T - \sum_{i=1}^{n}K_i\underline{x}_i V_i^{-1}\underline{x}_i^T\right)\underline{C}_o^{-1} \end{aligned} \tag{8.31}$$

Let

$$\underline{C}_m = \sum_{i=1}^{n}(1-K_i)^{-m}\underline{x}_i V_i^{-1}\underline{x}_i^T \tag{8.32}$$

Now supposing that K_i's are of $O(n^{-1})$ and expanding $(1-K_i)^{-m}$, we get to $O(n^{-2})$,

$$\underline{C}_o = \sum_{i=1}^{n}\underline{x}_i V_i^{-1}\underline{x}_i^T\,,\quad \underline{C}_1 = \sum_{i=1}^{n}(1-K_i)^{-1}\underline{x}_i V_i^{-1}\underline{x}_i^T \text{ and } \underline{C}_2 = \sum_{i=1}^{n}(1-K_i)^{-2}\underline{x}_i V_i^{-1}\underline{x}_i^T$$

Further, defining

$$\underline{C}_m^* = \sum_{i=1}^{n}K_i^m\underline{x}_i V_i^{-1}\underline{x}_i^T\,, \tag{8.33}$$

we have

$$\underline{C}_1^* = \sum_{i=1}^{n}K_i\underline{x}_i V_i^{-1}\underline{x}_i^T \text{ and } \underline{C}_2^* = \sum_{i=1}^{n}K_i^2\underline{x}_i V_i^{-1}\underline{x}_i^T$$

Substituting these values from (8.32) and (8.33) in (8.31), it follows that

$$E(V^*_{WJ}) = \frac{n}{n-p} \underline{C}_o^{-1} (\underline{C}_o - \underline{C}^*_1) \underline{C}_o^{-1}\sigma^2$$

$$= \frac{n}{n-p} (\underline{C}_o^{-1} - \underline{C}_o^{-1}\underline{C}^*_1\underline{C}_o^{-1})\sigma^2 \qquad (8.34)$$

Thus, we observe that V^*_{WJ} is biased in general and the bias is given by

$$E(V^*_{WJ}) - Var(\underline{\beta}^*_{WJ}) = \frac{n}{n-p} (\underline{C}_o^{-1} - \underline{C}_o^{-1}\underline{C}^*_1\underline{C}_o^{-1})\sigma^2 - \underline{C}_o^{-1}\sigma^2$$

$$= O(n^{-2}) \qquad (8.35)$$

This expression shows that the bias decreases at a faster rate than 1/n.

8.5 Summary and Conclusions

In this chapter we have mainly concentrated on the application of the weighted jack-knife procedure to regression estimation. We observe that:

(i) The weighted jack-knifed estimator $\tilde{\underline{\beta}}_{WJ}$ based on the weighted pseudovalues takes into consideration the unequal spacings between the observations. We also note that $\tilde{\underline{\beta}}_{WJ}$ is unbiased and $Var(\tilde{\underline{\beta}}_{WJ}) = Var(\hat{\underline{\beta}})$. The weighted jack-knifed variance estimator $\tilde{V}_{WJ}$ gives a biased estimate for the $Var(\tilde{\underline{\beta}}_{WJ})$ but the bias decreases at a faster rate than 1/n. When $w_i = \frac{p}{n}$, that is, in the balanced case $\tilde{V}_{WJ}$ is exactly unbiased.

(ii) The weighted jack-knifed variance estimator $\tilde{V}_{WJ}$ is robust, that is, it approximates the true variance of $\hat{\underline{\beta}}$ when the error variances are heterogeneous. Thus the weighted jack-knifed procedure leads to improved variance estimation.

(iii) Under the more general model with variance structure $\sigma^2 \underline{V}$, where $\underline{V}$ is known, the weighted jack-knifed estimator $\underline{\beta}^*_{WJ}$ is unbiased, that is, it is identical to the generalised least squares estimator $\underline{\beta}^*$. Also $\text{Var}(\underline{\beta}^*_{WJ}) = \text{Var}(\underline{\beta}^*)$. Under these conditions the weighted jack-knifed variance estimator V^*_{WJ} gives a biased estimate of the variance of $\underline{\beta}^*$ but this bias decreases at a faster rate than $1/n$.

In a nutshell we conclude that, in general, the weighted jack-knife procedure gives a more robust performance than the unweighted jack-knife procedure. This conclusion is supported by a small simulation study by Hinkley (1977b). However, further numerical investigations are required to study the behaviour of the weighted jack-knifed variance estimator $\tilde{V}_{WJ}$ using real data. In subsequent chapters we extend the weighted jack-knife method to ratio estimation.

CHAPTER 9

THE WEIGHTED JACK-KNIFED RATIO ESTIMATION

9.1 Introduction

Ratio estimation occupies an important place in sample surveys. Since the simple ratio estimator $\hat{R} = \frac{\bar{y}_s}{\bar{x}_s}$ is biased, it has become an area of application of the jack-knife method. In sample surveys, the auxiliary population mean $\bar{x}$ may be assumed known or at least estimated from a much larger sample in which case the ratio estimate $\frac{\bar{y}_s}{\bar{x}_s}\sum_{i=1}^{N} x_i$ based on (y_i,x_i), $i = 1,2,\ldots,n$ is often a more precise estimator of the population total T than the less sophisticated expansion estimator $N\bar{y}_s$. The ratio method is considered to be most effective where the relationship between the response y and the auxiliary variate x is linear through the origin.

Durbin (1959) pioneered the application of the jack-knife technique to ratio estimation by studying the behaviour of the jack-knifed ratio estimator $\tilde{R}_J = \frac{1}{g}\sum_{i=1}^{g} \tilde{R}_i$, under the model $y_i = \alpha+\beta x_i+e_i$, where e_i are independently and identically distributed with either a normal or gamma distribution and $\tilde{R}_i = g\hat{R}-(g-1)\hat{R}_{-i}$, $i = 1,2,\ldots,g$ are the pseudovalues. He considered the case with $g = 2$ and established that neglecting terms of $O(n^{-4})$, the jack-knifed ratio estimator $\tilde{R}_J$ has both smaller bias and smaller variance than the simple ratio estimate $\hat{R} = \frac{\bar{y}_s}{\bar{x}_s}$. Durbin also showed that for a gamma distribution, with coefficient of variation less than 1/4, the jack-knife method reduces the bias, increases the variance but reduces the mean square

error. Rao (1965) showed that the optimum choice of g with the jack-knife method for the normal distribution of the auxiliary variable is $g = n$. Rao and Webster (1966), through a combination of theoretical and numerical work, demonstrated that the optimum choice $g = n$ also holds when the auxiliary variable has the gamma distribution. Chakarbarty and Rao (1968) also obtained similar results on the estimates of variance, etc.

A considerable amount of research has been devoted to investigate the properties of the jack-knifed and other competing ratio estimators under superpopulation models with different variance structures. However, we note that all of these studies have used superpopulation models as an evaluation technique whereas the basic criteria regarding bias, variances and standard errors, etc., are referred to the probability sampling models. For example, Rao and Kuzik (1974) through a semi-empirical study compared the average performanc of $\hat{V}_1 c$ and $\hat{V}_2 c$ under the model $y_i = \beta x_i + e_i$ with x_i having a gamma distribution. These authors followed the rules of conventional theory in studying the performance of these estimators and by averaging over all possible samples could not detect flaws in $\hat{V}_1 c$ as a variance estimator.

The jack-knife is a distribution free method for bias reduction and variance estimation. In a wide class of problems it is known that the jack-knife produces consistent results. It may, however, be pointed out that the standard jack-knife procedure is not insensitive to extreme data points. Thus, when there are certain deviant points in the data, the standard jack-knife estimator may give misleading results. This has also been noted by Hinkley (1977b) and Wainer and Thissen (1975) in connection with the estimation of correlation.

We note that the jack-knife technqiue so far applied in sample surveys for ratio estimation deals with the balanced cases, that is, the pseudovalues are defined symmetrically with respect to the observations. However, there may be situations when the data points may not be well balanced. Such deviant points may have serious implications on the behaviour of the ratio estimate and the estimator of its variance. Under such circumstances it is necessary to modify the existing jack-knife procedure so as to obtain estimators which are less sensitive to extreme data points or the unbalanced structure of the data.

In this Chapter we propose weighted jack-knife procedures as an alternative to the unweighted jack-knife for ratio estimation. In Section 9.2, using the idea of empirical influence function, we propose the weighted pseudovalues, the weighted jack-knifed ratio estimator and the weighed jack-knifed variance estimator as modifications to their standard jack-knifed counterparts. We observe that the weighted jack-knifed ratio estimator and the weighted jack-knifed variance estimator reproduce the original estimator $\hat{R}$ and $\hat{V}_{2c}$ respectively. We discuss the properties of the weighted jack-knifed variance estimator in Section 9.3. This section contains theorems on the bias and the stability of the weighted jack-knifed variance estimator $\hat{V}_{WJ}$. It is shown that this estimator is approximately unbiased under the model ξ for all variance functions with certain restrictions on n and f. Theorem 9.3 shows that the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ is quite stable. A violation of the model assumptions may affect the behaviour of an estimator. In Section 9.4, we discuss the effect of changing the regression function in our model on the behaviour of the variance estimator $\hat{V}_{WJ}$. Another important point in

connection with the jack-knife technique is to ascertain whether a given jack-knife procedure removes the first order bias. Using Taylor's Series expansion, we examine this important aspect of the weighted and the unweighted jack-knife procedures in Section 9.5. Theorems are presented on the bias of the weighted and the unweighted jack-knife procedures. In Section 9.6, we discuss these results and summarise our conclusions. Finally, in Section 9.7 we consider some problems for further research work.

9.2 The Weighted Jack-knife for Ratio Estimation

The unweighted jack-knife lacks in one important aspect that it does not take into consideration the influence which the ith data point may have on the behaviour of the ratio estimate and the estimator of its variance. Under the unweighted jack-knife method the pseudovalues are defined symmetrically with respect to the observations whereas the data may be unbalanced. The unbalance is apparent in the distances w_i. The dinstances w_i provide a mesure of how far the ith point is from the centre of the data. The contribution of the ith observation to the jack-knifed estimate, $\hat{R} - \hat{R}_{-i}$ have a weight decreasing in w_i. We express this last statement as Lemma 9.1.

Lemma 9.1

Consider a sample of n observations (x_i, y_i), $i = 1,2,\ldots,n$ Let $\hat{R}$ denote the ratio estimate based on the full sample and $\hat{R}_{-i}$ be the corresponding ratio estimate obtained from the sample when the ith case has been deleted. Then

$$\hat{R} - \hat{R}_{-i} = \frac{y_i - \hat{R}x_i}{n\bar{x}_s(1-w_i)}, \quad \text{where} \quad w_i = \frac{x_i}{\sum_{i=1}^{n} x_i}. \tag{9.1}$$

Proof

We have $\hat{R} = \left(\sum_{j=1}^{n} y_j\right) / \left(\sum_{j=1}^{n} x_j\right)$ and $\hat{R}_{-i} = \dfrac{\sum_{j=1}^{n} y_j - y_i}{\sum_{i=1}^{n} x_j - x_i}$.

Thus

$$\hat{R} - \hat{R}_{-i} = \frac{\sum_{j=1}^{n} y_j}{\sum_{j=1}^{n} x_j} - \frac{\sum_{j=1}^{n} y_j - y_i}{\sum_{j=1}^{n} x_j - x_i},$$

$$= \frac{\hat{R}\left(\sum_{j=1}^{n} x_j - x_i\right) - \left(\sum_{j=1}^{n} y_j - y_i\right)}{\left(\sum_{j=1}^{n} x_j - x_i\right)},$$

$$= \frac{y_i - \hat{R}x_i}{n\bar{x}_s(1-w_i)}, \quad \text{where} \quad w_i = \frac{x_i}{\Sigma x_i} \tag{9.2}$$

This indicates that the contribution of the ith observation to the jack-knifed estimate $\hat{R} - \hat{R}_{-i}$ have a weight decreasing in w_i .

We note from Lemma 9.1 that a specific choice of weight could be given by the fact that

$$n(1-w_i)(\hat{R}-\hat{R}_{-i}) = \frac{n(y_i - \hat{R}x_i)}{\sum_{i=1}^{n} x_i} = \frac{n\hat{e}_i}{\sum_{i=1}^{n} x_i} = \hat{I}\ (R:x_i,y_i) \tag{9.3}$$

is the empirical influence function (defined at 7.22) of R at (x_i,y_i) where $\hat{e}_i = (y_i - \hat{R}x_i)$.

The proof of this result appears in Lemma 9.2.

Lemma 9.2

Consider a smaple of n observations $(x_1,y_1), (x_2,y_2),\ldots,(x_n,y_n)$. Let $\hat{R}$ denote the ratio estimate based on the full sample and $\hat{R}_{-i}$ be the ratio estimate obtained from the sample when the ith observation has been deleted. Then the empirical influence function of R at (x_i,y_i) is given by

$$n(1-w_i)(\hat{R}-\hat{R}_{-i}) = \frac{n\hat{e}_i}{\sum_{i=1}^{n} x_i} = \hat{I}(R:x_i,y_i)$$

Proof

From Lemma 9.1, we have

$$\hat{R} - \hat{R}_{-i} = \frac{y_i - \hat{R}x_i}{(1-w_i)\sum_{i=1}^{n} x_i}, \quad \text{where} \quad w_i = \frac{x_i}{\Sigma x_i}.$$

Let R_i^* denote the ith weighted pseudovalue and R^* the weighted jack-knifed ratio estimate, then

$$\hat{I}(R:x_i,y_i) = R_i^* - R^* \tag{9.4}$$

Now, the ith weighted pseudovalue R_i^* is given by

$$R_i^* = \hat{R} + n(1-w_i)(\hat{R}-\hat{R}_{-i})$$

$$= \hat{R} + \frac{n(y_i - \hat{R}x_i)}{\sum_{i=1}^{n} x_i} \tag{9.5}$$

The weighted jack-knifed estimate R^* is the average of the weighted pseudovalues R_i^* given by (9.5), that is,

$$R^* = \frac{1}{n}\sum_{i=1}^{n} R_i^* = \frac{1}{n}\sum_{i=1}^{n} [\hat{R}+n(1-w_i)(\hat{R}-\hat{R}_{-i})]$$

$$= \hat{R} + \sum_{i=1}^{n} \frac{(y_i-\hat{R}x_i)}{\sum_{i=1}^{n} x_i}$$

$$= \hat{R} \qquad \because \sum_{i=1}^{n} (y_i-\hat{R}x_i) = 0 \qquad (9.6)$$

Using (9.4), (9.5) and (9.6), the empirical influence function can now be given by

$$\hat{I}(R:x_i,y_i) = R_i^* - R^*$$

$$= \hat{R} + \frac{n(y_i-\hat{R}x_i)}{\sum_{i=1}^{n} x_i} - \hat{R}$$

$$= \frac{n(y_i-\hat{R}x_i)}{\sum_{i=1}^{n} x_i}$$

$$= \frac{n\hat{e}_i}{\sum_{i=1}^{n} x_i} \qquad (9.7)$$

We therefore propose the weighted pseudovalues, the weighted jack-knife ratio estimator and the weighted jack-knife variance estimator for ratio estimation in lieu of their unweighted jack-knife counterparts.

(i) the weighted pseudovalues

$$\tilde{R}_i = \hat{R} + n(1-w_i)(\hat{R}-\hat{R}_{-i}) \qquad (9.8)$$

(ii) the weighted jack-knifed ratio estimator is

$$\tilde{R}_{WJ} = \frac{1}{n}\sum_{i=1}^{n}\tilde{R}_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\hat{R}+n(1-w_i)(\hat{R}-\hat{R}_{-i})]$$

$$= \hat{R} + \sum_{i=1}^{n}\frac{(y_i-\hat{R}x_i)}{\sum_{i=1}^{n}x_i}$$

$$= \hat{R} . \qquad \because \sum_{i=1}^{n}(y_i-\hat{R}\,x_i) = 0 \qquad (9.9)$$

(iii) the weighted jack-knifed variance estimator is

$$\tilde{V}_{WJ} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\tilde{R}_i-\tilde{R}_{WJ})^2 ,$$

$$= \frac{1}{n(n-1)}\sum_{i=1}^{n}[\hat{R}+n(1-w_i)(\hat{R}-\hat{R}_{-i})-\tilde{R}_{WJ}]^2 ,$$

$$= \frac{1}{n(n-1)}\sum_{i=1}^{n}(y_i-\hat{R}x_i)^2/\bar{x}_s^2 , \qquad (9.10)$$

$$= \hat{V}_{2c} , \qquad (9.11)$$

$$= \frac{\bar{x}}{\bar{x}_s^2}\hat{V}_{1c} \qquad (9.12)$$

Where $\hat{V}_{2c}$ and $\hat{V}_{1c}$ are the two commonly used conventional variance estimators for $Var(\hat{R})$, see for example, Cochran (1977, p.155). Thus we note that $\tilde{R}_{WJ}$ and $\tilde{V}_{WJ}$ possess reproducing properties indicated by the expressions (9.9) and (9.11). This implies that the weighted jack-knifed ratio estimator $\tilde{R}_{WJ}$ is exactly equal to the conventional ratio estimator $\hat{R}$ and the weighted jack-knifed variance estimator $\tilde{V}_{WJ}$ is exactly equal to the conventional variance estimator $\hat{V}_{2c}$. The variance estimator $\hat{V}_{2c}$ is a good estimator of the conditional variance $V(\hat{R})$, see for example, Hájek (1958) and Konijn (1973, p.353).

Royall and Eberhardt (1975) noted the bias of $\hat{V}_{1c}$ when the finite population is a random sample from a superpopulation in which $y_i = \beta x_i + e_i$, where e_i are independent with zero mean and variance $\sigma^2 x_i^t$. They suggested a simple modification,

$$\hat{V}_H = \hat{V}_{1c} \frac{\bar{x}_{\tilde{s}} \bar{x}}{\bar{x}_s} \left(1 - \frac{C^2_{x(s)}}{n}\right),$$

where $C^2_{x(s)} = \frac{1}{n-1} \sum_s \left[\frac{x_i - \bar{x}_s}{\bar{x}_s}\right]^2$, $\bar{x}_s$ is the mean of the sample units and $\bar{x}_{\tilde{s}}$ is the mean of the $(N-n)$ non-sampled units. The estimator $\hat{V}_H \doteq \hat{V}_{2c}$ for large n and $N >>> n$ which justifies the use of $\hat{V}_{2c}$ from a superpopulation point of view. The estimator $\hat{V}_H$ is model unbiased when $t = 1$ and remains approximately ξ-unbiased when the variance function in the model described above is not proportional to x_i.

In the next section, we examine the behaviour of the weighted jack-knife variance estimator $\tilde{V}_{WJ}$ and see how does it compare with the randomisation based estimator $\hat{V}_{1c}$, the prediction theory estimator $\hat{V}_H$ and the unweighted jack-knife variance estimator $\hat{V}_J$, etc.

9.3 Properties of the Weighted Jack-knifed Variance Estimator Under the Model ξ

In this section we derive the expected values and the bias of the weighted jack-knifed variance estimator under the general linear regression model considered to be most suitable for evaluating the ratio estimators. Consider the model,

$$y_i = \beta x_i + e_i \qquad (9.13)$$

where $E(y_i) = \beta x_i$, $Var(y_i) = v_i$ and $Cov(y_i, y_j) = 0$ for $i \neq j$. We consider the ratio estimator $\hat{T}_R = \frac{\sum_s y_i}{\sum_s x_i}\left(\sum_{i=1}^N x_i\right)$ for the finite population total $T = \sum_{i=1}^N y_i$. The expressions for the bias are derived with respect to the actual mean square error $E_\xi(\hat{T}_R - T)^2$.

Three other variants of the model (9.13) considered as special cases are those with $v_i = \sigma^2$, $v_i = \sigma^2 x_i$ and $v_i = \sigma^2 x_i^2$. The weighted jack-knifed variance estimator for the finite population total can be written as

$$\hat{V}_{WJ} = \frac{N^2}{n}(1-f)\frac{\bar{x}^2}{\bar{x}_s^2}\sum_{i=1}^{n}(y_i-\hat{R}x_i)^2/(n-1) \qquad (9.14)$$

9.3.1a Expected Values of $\hat{V}_{WJ}$ Under the Model ξ with Different Variance Functions

We consider the general case as Theorem 9.1 and then give special cases.

Theorem 9.1

Consider the general linear regression model,

$$y_i = \beta x_i + e_i,\ E(e_i) = 0\ ,\ Var(e_i)=v_i,\ Cov(e_i,e_j) = 0 \text{ for } i\neq j \qquad (9.15)$$

Under this model the expected value of the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ is given by

$$E_\xi(\hat{V}_{WJ}) = \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \bar{v}_s[1+\frac{1}{n}(C^2_{x(s)}-2C_{xv(s)})]\ . \qquad (9.16)$$

Proof

The weighted jack-knife variance estimator for the finite population total can be written as

$$\hat{V}_{WJ} = \frac{N^2}{n}(1-\frac{n}{N})\frac{\bar{x}^2}{\bar{x}_s^2}\sum_{i=1}^{n}(y_i-\hat{R}x_i)^2/(n-1)$$

Under the model (9.15), the expected value of $\hat{V}_{WJ}$ is given by

$$E_\xi(\hat{V}_{WJ}) = \frac{N^2}{n}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 E_\xi\left[\sum_{i=1}^{n}(y_i-\hat{R}x_i)^2/(n-1)\right], \quad \text{where} \quad f = \frac{n}{N}.$$

$$= \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 E_\xi\left[\sum_{i=1}^{n}\left\{(\hat{\beta}-\beta)^2x_i^2-2e_i(\hat{\beta}-\beta)x_i+e_i^2\right\}/(n-1)\right]. \tag{9.17}$$

Now, we have

$$E_\xi(\hat{\beta}-\beta)^2 = E_\xi\left[\frac{\sum_s y_i}{\sum_s x_i} - \beta\right]^2 = E_\xi\left(\frac{\sum_s e_i}{\sum_s x_i}\right)^2 = \frac{\sum_s v_i}{(\sum_s x_i)^2} \tag{9.18}$$

$$E_\xi(e_i\hat{\beta}) = E_\xi\left[e_i\frac{\sum_s(\beta x_i+e_i)}{\sum_s x_i}\right] = E_\xi\left(\frac{e_i\sum_s e_i}{\sum_s x_i}\right) = \frac{v_i}{\sum_s x_i}. \tag{9.19}$$

Since the x's are known constants after the sample is taken, and hence does not have expectation under the model. Substituting (9.18) and (9.19) in (9.17) we get

$$E_\xi(\hat{V}_{WJ}) = \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2\left[\frac{(\sum_s x_i^2)(\sum_s v_i)}{(\sum_s x_i)^2} + \sum_s v_i - \frac{2\sum_s x_i v_i}{\sum_s x_i}\right]/(n-1),$$

$$= \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \bar{v}_s\left[\frac{\sum x_i^2}{n\bar{x}_s^2} + n - \frac{2\sum_s x_i v_i}{\bar{v}_s(\sum_s x_i)}\right]/(n-1), \quad \text{where} \quad \bar{v}_s = \frac{1}{n}\sum_s v_i.$$

$$= \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \bar{v}_s\left[1+\frac{1}{n}(C^2_{x(s)}-2C_{xv(s)}\right], \tag{9.20}$$

$$= \frac{N}{f}(1-f)\frac{\bar{x}\tilde{x}_s}{\bar{x}_s^2}\bar{v}_s\frac{\bar{x}}{\tilde{x}_s}\left[1+\frac{1}{n}(C^2_{x(s)}-2C_{xv(s)})\right], \tag{9.21}$$

$$= A \frac{\bar{x}}{\bar{x}_{\tilde{s}}} \left[1 + \frac{1}{n}(C^2_{x(s)} - 2C_{xv(s)})\right] \quad , \tag{9.22}$$

where

$$A = \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2}\bar{v}_s \quad , \quad C^2_{x(s)} = \sum_{i=1}^{n} \frac{(x_i - \bar{x}_s)^2}{(n-1)\bar{x}_s^2} \quad \text{and} \quad C_{xv(s)} = \frac{\sum_s (x_i - \bar{x}_s)(v_i - \bar{v}_s)}{(n-1)\bar{x}_s \bar{v}_s} \, .$$

Corollary 9.1

If we consider the model with constant variance function and put $v_i = \sigma^2$, then $\bar{v}_s = \sigma^2$ and $C_{xv}(s) = \frac{\Sigma(x_i - \bar{x}_s)(v_i - \bar{v}_s)}{(n-1)\bar{x}_s \bar{v}_s} = 0$.

Substituting these values in (9.20), we get

$$E_\xi(\hat{V}_{WJ}) = \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2}\sigma^2 \frac{\bar{x}}{\bar{x}_{\tilde{s}}}\left[1 + \frac{1}{n} C^2_{x(s)}\right] ,$$

$$= A^* \frac{\bar{x}}{\bar{x}_{\tilde{s}}}\left[1 + \frac{1}{n} C^2_{x(s)}\right] , \tag{9.23}$$

where $A^* = \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2}$.

For large n , expression (9.23) reduces to

$$E_\xi(\hat{V}_{WJ}) = A^* \frac{\bar{x}}{\bar{x}_{\tilde{s}}} \, . \tag{9.24}$$

Corollary 9.2

If we consider the model with $v_i = \sigma^2 x_i$, then $\bar{v}_s = \sigma^2 \bar{x}_s$

and $C^2_{xv(s)} = \frac{\Sigma(x_i - \bar{x}_s)(v_i - \bar{v}_s)}{(n-1)\bar{x}_s \bar{v}_s} = \sum_{i=1}^{n} \frac{(x_i - \bar{x}_s)^2}{(n-1)\bar{x}_s^2} = C^2_{x(s)}$.

Under this situation, we have

$$E_\xi(\hat{V}_{WJ}) = \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_s}{\bar{x}_s^2}\sigma^2\bar{x}_s\frac{\bar{x}}{\bar{x}_s}\left(1-\frac{1}{n}C^2_x(s)\right),$$

$$= A^*\frac{\bar{x}_s\bar{x}}{\bar{x}_s}\left(1-\frac{1}{n}C^2_x(s)\right). \qquad (9.25)$$

$$= A^*\frac{\bar{x}_s\bar{x}}{\bar{x}_s}, \quad \text{when } n \text{ is large.}$$

Corollary 9.3

If we consider the model with $v_i = \sigma^2 x_i^2$, then $\bar{v}_s = \sigma^2\bar{\bar{x}}_s$,

where $\bar{\bar{x}}_s = \frac{1}{n}\sum_s x_i^2$; $C^2_x(s) = \frac{\sum_s(x_i-\bar{x}_s)^2}{(n-1)\bar{x}_s^2} = \frac{n}{n-1}\frac{(\bar{\bar{x}}_s-\bar{x}_s^2)}{\bar{x}_s^2}$ and

$$C_{xv}(s) = \frac{\sum_s(x_i-\bar{x}_s)(x_i^2-\bar{\bar{x}}_s)}{(n-1)\bar{x}_s\bar{\bar{x}}_s} = \frac{n}{n-1}\frac{\sum_s(\bar{\bar{x}}_s-\bar{x}_s\bar{\bar{x}}_s)}{\bar{\bar{x}}_s\bar{\bar{x}}_s}, \quad \text{where } \bar{\bar{x}}_s^k = \frac{1}{n}\sum_s x_i^k.$$

Substituting these in (9.24), gives

$$E_\xi(\hat{V}_{WJ}) = \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_s}{\bar{x}_s^2}\sigma^2\frac{\bar{\bar{x}}_s\bar{x}}{\bar{x}_s}\left[1+\frac{1}{n-1}\left(\frac{\bar{\bar{x}}_s}{\bar{x}_s^2}-1\right)-\frac{2}{n-1}\left(\frac{\bar{\bar{x}}_s-\bar{x}_s\bar{\bar{x}}_s}{\bar{x}_s\bar{\bar{x}}_s}\right)\right],$$

$$= A^*\frac{\bar{\bar{x}}_s\bar{x}}{\bar{x}_s}\left[1-\frac{1}{n-1}\left\{\frac{2\bar{\bar{x}}_s}{\bar{x}_s\bar{\bar{x}}_s}-\frac{\bar{\bar{x}}_s}{\bar{x}_s^2}-1\right\}\right]. \qquad (9.26)$$

$$= A^*\frac{\bar{\bar{x}}_s\bar{x}}{\bar{x}_s}, \quad \text{for large } n. \qquad (9.27)$$

9.3.1b Error Variance $E_\xi(\hat{T}_R-T)^2$

In order to find the bias of weighted jack-knifed variance estimator $\hat{V}_{WJ}$, we require to find expressions for $E_\xi(\hat{T}_R-T)^2$ under the model (9.15) for different variance functions. This result is presented as Theorem 9.2. The derivation procedure is the same as in Royall and Eberhardt (1975).

Theorem 9.2

Under the model (9.15) the error variance is given by

$$E_\xi(\hat{T}_R-T)^2 = \frac{N}{f}(1-f)\frac{\bar{x}\tilde{\bar{x}}_s}{\bar{x}_s^2}\bar{v}_s\left[1+f\frac{\bar{x}_s}{\bar{x}}\frac{(\bar{x}_s\tilde{\bar{v}}_s-\tilde{\bar{x}}_s\bar{v}_s)}{\tilde{\bar{x}}_s\bar{v}_s}\right]. \tag{9.28}$$

Proof

Under the model (9.15) the error variance is

$$E_\xi(\hat{T}_R-T)^2 = E_\xi\left[\frac{\sum_s y_i}{\sum_s x_i}\left(\sum_{i=1}^N x_i\right) - \sum_{i=1}^N y_i\right]^2,$$

$$= E_\xi\left[\frac{\sum_s (\beta x_i+e_i)}{\sum_s x_i}\left(\sum_{i=1}^N x_i\right) - \sum_{i=1}^N (\beta x_i+e_i)\right]^2,$$

$$= E_\xi\left[\frac{\sum_s e_i}{\sum_s x_i}\left(\sum_{i=1}^N x_i\right) - \sum_{i=1}^N e_i\right]^2,$$

$$= E_\xi\left[\frac{\left(\sum_{i=1}^N x_i\right)^2}{\bar{x}_s^2}\bar{e}_s^2 - 2N\left(\sum_{i=1}^N x_i\right)\frac{\bar{e}_s}{\bar{x}_s}\bar{e}_N+N^2\bar{e}_N^2\right]. \tag{9.29}$$

Now under the model (9.15), we have

$$E_\xi(\bar{e}_s^2) = \mathrm{Var}\left(\frac{1}{n}\sum_s e_i\right) = \frac{1}{n^2}\sum_s v_i = \frac{1}{n}\bar{v}_s .$$

$$E_\xi(\bar{e}_N^2) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^N e_i\right) = \frac{1}{N^2}\sum_{i=1}^N v_i = \frac{1}{N}\bar{v}$$

$$\mathrm{Cov}(\bar{e}_s,\bar{e}_N) = E_\xi(\bar{e}_s\bar{e}_N) - E_\xi(\bar{e}_s)E_\xi(\bar{e}_N)$$

$$= \frac{1}{nN}E_\xi\left(\sum_s e_i\right)\left(\sum^N e_j\right) = \frac{1}{Nn}E_\xi\sum_{i\in s} e_i^2) = \frac{1}{Nn}\sum_{i\in s} v_i$$

$$= \frac{1}{N}\bar{v}_s .$$

Substituting these expressions in (9.29), we get

$$E_\xi(\hat{T}_R-T)^2 = \frac{\left(\sum_{i=1}^{N} x_i\right)}{n\bar{x}_s^2}\bar{v}_s - 2\frac{\sum_{i=1}^{N} x_i}{\bar{x}_s}\bar{v}_s + \sum_{i=1}^{N} v_i \quad ,$$

$$= \frac{\left(\sum_{i=1}^{N} x_i\right)\bar{v}_s}{n\bar{x}_s^2}\left(\sum_{i=1}^{N} x_i - n\bar{x}_s\right) + \sum_{i=1}^{N} v_i - \frac{\sum_{i=1}^{N} x_i}{\bar{x}_s}\bar{v}_s \quad . \qquad (9.30)$$

Since $N\bar{x} = n\bar{x}_s + (N-n)\bar{x}_{\tilde{s}}$ and $N\bar{v} = n\bar{v}_s + (N-n)\bar{v}_{\tilde{s}}$, therefore (9.30) reduces to

$$E_\xi(\hat{T}_R-T)^2 = \frac{N\bar{x}\bar{v}_s}{n\bar{x}_s^2}(N-n)\bar{x}_{\tilde{s}} + (n\bar{v}_s+(N-n)\bar{v}_{\tilde{s}}) - \frac{(n\bar{x}_s+(N-n)\bar{x}_{\tilde{s}})\bar{v}_s}{\bar{x}_s} \quad ,$$

$$= \frac{N\bar{x}\bar{v}_s}{n\bar{x}_s^2}(N-n)\bar{x}_{\tilde{s}} + (N-n)\bar{v}_{\tilde{s}} - (N-n)\frac{\bar{x}_{\tilde{s}}\bar{v}_s}{\bar{x}_s} \quad ,$$

$$= \frac{N\bar{x}\bar{v}_s}{n\bar{x}_s^2}(N-n)\bar{x}_{\tilde{s}}\left[1 + \frac{\bar{x}_s^2}{\bar{x}\bar{v}_s}\frac{n}{N}\frac{1}{\bar{x}_{\tilde{s}}}\left(\frac{\bar{x}_s\bar{v}_{\tilde{s}}-\bar{x}_{\tilde{s}}\bar{v}_s}{\bar{x}_s}\right)\right] \quad ,$$

$$= \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2}\bar{v}_s\left[1+f\frac{\bar{x}_s}{\bar{x}}\left(\frac{\bar{x}_s\bar{v}_{\tilde{s}}-\bar{x}_{\tilde{s}}\bar{v}_s}{\bar{x}_{\tilde{s}}\bar{v}_s}\right)\right] \quad , \qquad (9.31)$$

$$= A\left[1+f\frac{\bar{x}_s}{\bar{x}}\frac{(\bar{x}_s\bar{v}_{\tilde{s}}-\bar{x}_{\tilde{s}}\bar{v}_s)}{\bar{x}_{\tilde{s}}\bar{v}_s}\right] \quad . \qquad (9.32)$$

Corollary 9.4

If $v_i = \sigma^2$, then $\bar{v}_s = \bar{v}_{\tilde{s}} = \sigma^2$ and

$$E_\xi(\hat{T}_R-T)^2 = \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2}\sigma^2\left[1+\frac{n}{N}\frac{\bar{x}_s}{\bar{x}}\left(\frac{(\bar{x}_s-\bar{x}_{\tilde{s}})}{\bar{x}_{\tilde{s}}}\right)\right] \quad ,$$

$$= A^*\frac{\bar{x}}{\bar{x}_{\tilde{s}}}\left[1+\frac{f}{1-f}\frac{(\bar{x}_s-\bar{x})^2}{\bar{x}^2}\right] \quad . \qquad (9.33)$$

Corollary 9.5

If $v_i = \sigma^2 x_i$, then $\bar{v}_s = \sigma^2 \bar{x}_s$, $\bar{v}_{\tilde{s}} = \sigma^2 \bar{x}_{\tilde{s}}$ and $\bar{x}_s \bar{v}_{\tilde{s}} - \bar{x}_{\tilde{s}} \bar{v}_s = 0$.

Thus substituting in (9.31), we get

$$E_\xi (\hat{T}_R - T)^2 = \frac{N}{f} (1-f) \frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2} \sigma^2 \bar{x}_s ,$$

$$= A^* \bar{x}_s . \tag{9.34}$$

Corollary 9.6

If $v_i = \sigma^2 x_i^2$, then $\bar{v}_s = \sigma^2 \bar{\bar{x}}_s$, $\bar{v}_{\tilde{s}} = \sigma^2 \bar{\bar{x}}_{\tilde{s}}$ and

$$E_\xi (\hat{T}_R - T)^2 = \frac{N}{f} (1-f) \frac{\bar{x}\bar{x}_{\tilde{s}}}{\bar{x}_s^2} \sigma^2 \bar{\bar{x}}_s \left[1 + f \frac{\bar{x}_s}{\bar{x}} \left(\frac{\bar{\bar{x}}_s \bar{x}_{\tilde{s}} - \bar{x}_{\tilde{s}} \bar{\bar{x}}_s}{\bar{x}_{\tilde{s}} \bar{\bar{x}}_s} \right) \right]$$

$$= A^* \bar{\bar{x}}_s \left[1 + f \frac{\bar{x}_s}{\bar{x}} \left(\frac{\bar{\bar{x}}_s \bar{x}_{\tilde{s}}}{\bar{x}_{\tilde{s}} \bar{\bar{x}}_s} - 1 \right) \right] . \tag{9.35}$$

$$= \frac{N}{f} (1-f) \sigma^2 \bar{\bar{x}}_s \left[(1-f) \frac{\bar{x}_{\tilde{s}}^2}{\bar{x}_s^2} + f \frac{\bar{\bar{x}}_{\tilde{s}}}{\bar{\bar{x}}_s} \right] . \tag{9.36}$$

9.3.2 Bias of the Weighted Jack-knifed Variance Estimator

We now work out the expressions for the bias of the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ under the model (9.15) for different variance functions.

(i) Bias Under the Model with General Variance Function: $\mathrm{Var}(e_i) = v_i$

$$\text{Bias}(\hat{V}_{WJ}) = E_\xi(\hat{V}_{WJ}) - E_\xi(\hat{T}_R - T)^2,$$

$$= A\frac{\bar{x}}{\bar{x}_{\tilde{s}}}\left[1+\frac{1}{n}\left\{C^2_x(s)-2C_{xv}(s)\right\}\right] - A\left[1+f\frac{\bar{x}_s}{\bar{x}}\frac{(\bar{x}_s\bar{v}_{\tilde{s}}-\bar{x}_s\bar{v}_s)}{\bar{x}_{\tilde{s}}\bar{v}_s}\right],$$

$$= A\left[\left(\frac{\bar{x}}{\bar{x}_{\tilde{s}}}-1\right)+\frac{1}{n}\left\{C^2_x(s)-2C_{xv}(s)\right\}-f\frac{\bar{x}_s}{\bar{x}}\left(\frac{\bar{x}_s\bar{v}_{\tilde{s}}}{\bar{x}_{\tilde{s}}\bar{v}_s}-1\right)\right]. \qquad (9.37)$$

However, if we assume that n is large and f is small, then it could be easily seen from (9.37) that the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ is approximately unbiased.

(ii) Bias Under the Model (9.15) with Constant Variance Function

When $v_i = \sigma^2$, then from expressions (9.23) and (9.33), we have

$$\text{Bias}(\hat{V}_{WJ}) = A^*\frac{\bar{x}}{\bar{x}_{\tilde{s}}}\left[\frac{1}{n}C^2_x(s) - \frac{f}{1-f}\frac{(\bar{x}_s-\bar{x})}{\bar{x}^2}\right]. \qquad (9.38)$$

From expression (9.38), we note that the bias reduces to zero if we assume that n is large and f is small.

(iii) Bias Under Model (9.15) with Variance Function Proportion to x

When $v_i = \sigma^2 x_i$, then from expression (9.25) and (9.34), we get

$$\text{Bias}(\hat{V}_{WJ}) = A^*\frac{\bar{x}_s\bar{x}}{\bar{x}_{\tilde{s}}}\left[1-\frac{1}{n}C^2_x(s)\right] - A^*\bar{x}_s,$$

$$= A^*\left[\frac{\bar{x}_s(\bar{x}-\bar{x}_{\tilde{s}})}{\bar{x}_{\tilde{s}}} - \frac{1}{n}C^2_x(s)\right]. \qquad (9.39)$$

It could be easily seen from expression (9.39) that for large n and small f, the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ is approximately unbiased under Model (9.15) when the variance function is proportional to x.

(iv) Bias Under Model (9.15) with Variance Function Quadratic in x.

When $v_i = \sigma^2 x_i^2$, then from expressions (9.26) and (9.35), we have

$$\text{Bias}\,(\hat{V}_{WJ}) = A^* \bar{\bar{x}}_s \left[\left(\frac{\bar{x}}{\bar{x}_{\tilde{s}}} - 1 \right) - \frac{1}{n-1} \left(\frac{2\bar{\bar{x}}_s}{\bar{x}_s \bar{\bar{x}}_s} - \frac{\bar{\bar{x}}_s}{\bar{x}_s^2} - 1 \right) - f \frac{\bar{x}_s}{\bar{x}} \left(\frac{\bar{x}_s \bar{\bar{x}}_{\tilde{s}}}{\bar{x}_{\tilde{s}} \bar{\bar{x}}_s} - 1 \right) \right] . \quad (9.40)$$

It is quite evident from expression (9.40) that, for large n and small f, the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ is approximately unbiased when the variance function is quadratic in x.

Thus, we note from these results that under the conditions that, n is large and f is small, the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ is robust. This implies that this estimator does not depend on the variance function v_i. This is a desirable property of an estimator.

9.3.3 Stability of the Weighted Jack-knifed Variance Estimator $\hat{V}_{WJ}$

In the previous subsection we have examined the bias of the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ and found that it is approximately unbiased under the model (9.15) for the variance functions considered, subject to the condition that f is small and n is large. Now under the same conditions and also assuming that the error distribution is normal, we examine the stability of the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ under the model (9.15) with different variance functions. This result is presented as Theorem 9.3.

Theorem 9.3

Consider the model $y_i = \beta x_i + e_i$, $E(e_i) = 0$, $E(e_i^2) = v_i$, $E(e_i,e_j) = 0$ for $i \neq j$ and $e_i \sim NID(0,v_i)$. Then the variance of the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ is approximately given by

$$\mathrm{Var}(\hat{V}_{WJ}) \doteq \frac{2A^2}{n} \frac{\bar{\bar{v}}_s^2 \bar{x}^2}{\bar{\bar{v}}_s \bar{\bar{x}}_s^2}, \quad \text{where} \quad A = \frac{N}{f}(1-f) \frac{\bar{\bar{x}}\bar{\bar{x}}_s\bar{\bar{v}}_s}{\bar{\bar{x}}_s^2}, \tag{9.41}$$

$$\bar{\bar{v}}_s = \frac{1}{n}\sum_s v_i, \quad \bar{\bar{v}}_s^2 = \frac{1}{n}\sum_s v_i^2 \quad \text{and} \quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i .$$

Proof

We have from (9.14)

$$\hat{V}_{WJ} = \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sum_s (y_i - \hat{R}x_i)^2/n-1 .$$

This expression can be written as

$$\hat{V}_{WJ} = \frac{N}{f}(1-f)\left(\frac{\bar{\bar{x}}\bar{\bar{x}}_s\bar{\bar{v}}_s}{\bar{x}_s^2}\right)\left(\frac{\bar{x}}{\bar{\bar{x}}_s\bar{\bar{v}}_s}\right)\sum_s (y_i - \hat{R}x_i)^2/n-1 ,$$

$$= \frac{N}{f}(1-f)\left(\frac{\bar{\bar{x}}\bar{\bar{x}}_s\bar{\bar{v}}_s}{\bar{x}_s^2}\right)\left(\frac{\bar{x}}{\bar{\bar{x}}_s\bar{\bar{v}}_s}\right)\sum_s \left(\beta x_i + e_i - \left(\frac{\sum_s y_i}{\sum_s x_i}\right)x_i\right)^2 / (n-1),$$

$$= A\left(\frac{\bar{x}}{\bar{\bar{x}}_s\bar{\bar{v}}_s}\right)\sum_s \left[e_i - \frac{\left(\sum_s e_i\right)}{\left(\sum_s x_i\right)}x_i\right]^2 / (n-1) . \tag{9.42}$$

Now,

$$\mathrm{Var}[\hat{V}_{WJ}] = A^2\left(\frac{\bar{x}}{\bar{x}_s\bar{v}_s}\right)^2 \mathrm{Var}\left[\sum_s \left\{e_i - \frac{\left(\sum_s e_i\right)}{\left(\sum_s x_i\right)} x_i\right\}^2 / n-1\right],$$

$$= \frac{A^2}{(n-1)^2}\left(\frac{\bar{x}}{\bar{x}_s\bar{v}_s}\right)^2 \mathrm{Var}\left[\sum_s e_i^2 - 2\sum_s \frac{e_i x_i \sum_s e_i}{\sum_s x_i} + \sum_s \frac{\left(\sum_s e_i\right)^2 x_i^2}{\left(\sum_s x_i\right)^2}\right],$$

$$= \frac{A^2}{(n-1)^2}\left(\frac{\bar{x}}{\bar{x}_s\bar{v}_s}\right)^2 \left[\mathrm{Var}\left(\sum_s e_i^2\right) + 4\,\mathrm{Var}\left(\sum_s \frac{e_i x_i \sum_s e_i}{\sum_s x_i}\right) + \mathrm{Var}\left(\sum_s \frac{\left(\sum_s e_i\right)^2 x_i^2}{\left(\sum_s x_i\right)^2}\right) + \text{Covariance terms}\right],$$

$$\doteq \frac{A^2}{(n-1)^2}\left(\frac{\bar{x}}{\bar{x}_s\bar{v}_s}\right)^2 \left[\mathrm{Var}\left(\sum_s e_i^2\right) + 4\,\mathrm{Var}\left(\sum_s \frac{e_i x_i \sum_s e_i}{\sum_s x_i}\right) + \mathrm{Var}\left(\sum_s \frac{\left(\sum_s e_i\right)^2 x_i^2}{\left(\sum_s x_i\right)^2}\right)\right]. \quad (9.43)$$

The covariance terms are of $O(n^{-3})$ or greater, hence we ignore these terms. Now taking the first term in (9.43) we get

$$\mathrm{Var}\left(\sum_s e_i^2\right) = \sum_s \mathrm{Var}(e_i^2) + \sum_{i\neq j}\sum \mathrm{Cov}(e_i^2, e_j^2) .$$

$$\mathrm{Var}(e_i^2) = E_\xi(e_i^4) - (E_\xi e_i^2)^2 = \mu_{4i} - v_i^2$$

$$\mathrm{Cov}(e_i^2, e_j^2) = E_\xi(e_i^2 e_j^2) - E_\xi(e_i^2)E_\xi(e_j^2) = 0 \quad \because \; e_i \text{ and } e_j \text{ are independent}$$

Therefore, we have

$$\mathrm{Var}\left(\sum_s e_i^2\right) = \sum_s (\mu_{4i} - v_i^2) \tag{9.44}$$

Taking the second term in (9.43), we get

$$\mathrm{Var}\left[\sum_s \frac{e_i(\sum_s e_i)}{\sum_s x_i} x_i\right] = \sum_s \mathrm{Var}\left[\frac{e_i(\sum_s e_i)x_i}{\sum_s x_i}\right]$$

$$+ \sum_{i\neq j}\sum \mathrm{Cov}\left[\frac{e_i(\sum_s e_i)x_i}{\sum_s x_i}, \frac{e_j(\sum_s e_j)x_j}{\sum_s x_j}\right] .$$

Now $E_\xi e_i^2\left(\sum_s e_i\right)^2 - \left[E_\xi e_i \sum_s e_i\right]^2 = \mu_{4i} - v_i^2$ and covariance term reduces to zero. Hence we have

$$\mathrm{Var}\left[\sum_s \frac{e_i \sum_s e_i x_i}{\sum_s x_i}\right] = \sum_s \frac{(\mu_{4i} - v_i^2)x_i^2}{(\sum_s x_i)^2} . \tag{9.45}$$

Taking the third term in (9.43), we get

$$\mathrm{Var}\left[\frac{(\sum_s e_i)^2 x_i^2}{(\sum_s x_i)^2}\right] = \sum_s \mathrm{Var}\left[\frac{(\sum_s e_i)^2 x_i^2}{(\sum_s x_i)^2}\right] + \sum_{i\neq j}\sum \mathrm{Cov}\left[\frac{(\sum_s e_i)^2 x_i^2}{(\sum_s x_i)^2}, \frac{(\sum_s e_j)^2 x_j^2}{(\sum_s x_j)^2}\right] .$$

Now

$$\mathrm{Var}\left[\frac{(\sum_s e_i)^2 x_i^2}{(\sum_s x_i)^2}\right] = C_i^2 \mathrm{Var}(\sum_s e_i)^2, \text{ where } C_i = \frac{x_i^2}{(\Sigma x_i)^2} ,$$

$$= C_i^2 [E_\xi (\sum_s e_i)^4 - \left\{E_\xi (\sum_s e_i)^2\right\}^2] . \tag{9.46}$$

$$E_\xi\left(\sum_s e_i\right)^2 = E_\xi\left[\sum_s e_i^2 + \sum_{i\neq j}\sum e_i e_j\right] = \sum_s v_i \ ,$$

and

$$E_\xi\left[\left(\sum_s e_i\right)^4\right] = E_\xi\left[\sum_s e_i^4 + 4\sum_{i\neq j}\sum e_i^3 e_j + 3\sum_{i\neq j}\sum e_i^2 e_j^2 + 6\sum_{i\neq j\neq k}\sum\sum e_i^2 e_j e_k + \sum_{i\neq j\neq k\neq m}\sum\sum\sum\sum e_i e_j e_k e_m\right],$$

$$= \sum_s \mu_{4i} + 3\sum_{i\neq j}\sum v_i v_j \ .$$

Substituting these in (9.46) and simplifying we get

$$\mathrm{Var}\left(\frac{\left(\sum_s e_i\right)^2 x_i^2}{\left(\sum_s x_i\right)^2}\right) = \sum_s\left[\left(\sum_s \mu_{4i} + 3\sum_{i\neq j}\sum v_i v_j\right)\frac{x_i^4}{\left(\sum_s x_i\right)^4} - \frac{\left(\sum_s v_i\right)^2 x_i^4}{\left(\sum_s x_i\right)^4}\right],$$

$$= \sum_s\left[\frac{\left(\sum_s \mu_{4i}\right)x_i^4}{\left(\sum_s x_i\right)^4} + \frac{3\left(\sum_{i\neq j}\sum v_i v_j\right)x_i^4}{\left(\sum_s x_i\right)^4} - \frac{\left(\sum_s v_i^2\right)x_i^4}{\left(\sum_s x_i\right)^4} - \frac{\left(\sum_{i\neq j}\sum v_i v_j\right)x_i^4}{\left(\sum_s x_i\right)^4}\right],$$

$$= \frac{\left(\sum_s \mu_{4i}\right)\sum_s x_i^4}{n^4\bar{x}_s^4} + \frac{2\left(\sum_{i\neq j}\sum v_i v_j\right)\sum_s x_i^4}{n^4\bar{x}_s^4} - \frac{\bar{v}_s\sum_s\sum x_i^4}{n^3\bar{x}_s^4} \ . \qquad (9.47)$$

Substituting (9.44), (9.45) and (9.47) in (9.43), we get

$$\mathrm{Var}(\hat{V}_{WJ}) \doteq \frac{A^2}{(n-1)^2}\left(\frac{\bar{x}}{\bar{x}_s\bar{v}_s}\right)^2\left[\sum_s(\mu_{4i}-v_i^2) + 4\sum\frac{(\mu_{4i}-v_i^2)x_i^2}{n^2\bar{x}_s^2} + \right.$$

$$\left.\frac{\left(\sum_s \mu_{4i}\right)\sum_s x_i^4}{n^2\bar{x}_s^4} + 2\,\frac{\left(\sum_{i\neq j}\sum v_i v_j\right)\sum_s x_i^4}{n\bar{x}_s^4} - \frac{\bar{v}_s\sum_s\sum x_i^4}{n^3\bar{x}_s^4}\right],$$

$$\doteq \frac{A^2}{(n-1)^2}\left(\frac{\bar{x}}{\bar{x}_s\bar{v}_s}\right)^2\sum_s(\mu_{4i}-v_i^2) \ , \quad \text{ignoring terms of } O(n^{-3}) \text{ and higher.}$$

Under the assumption of normality $\mu_{4i} = 3v_i^2$ and hence

$$\text{Var}(\hat{V}_{WJ}) \doteq \frac{2A^2}{n}\left(\frac{\bar{x}}{\bar{x}_s \bar{v}_s}\right)^2 \sum_s v_i^2$$

$$\doteq \frac{2A^2}{n}\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \frac{\bar{v}^2}{\bar{v}_s^2} \qquad (9.48)$$

Now, if we assume that both n and $1/f$ are large, then expression (9.48) reduces to zero.

We now consider some special cases of the above general result with different variance functions.

Corollary 9.7

When $v_i = \sigma^2$, we have $\sum_s v_i^2 = \sum_s \sigma^4$ and $\sum_s v_i = \sum_s \sigma^2$, so that expression (9.48) reduces to

$$\text{Var}(\hat{V}_{WJ}) \doteq \frac{2A^{*2}}{n}\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sigma^4, \text{ where } A^* = \frac{N}{f}(1-f)\frac{\bar{x}\bar{x}_s}{\bar{x}_s^2}\bar{\sigma}^2$$

$$\doteq \frac{2A^{*2}}{n}\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 . \qquad (9.49)$$

Corollary 9.8

When $v_i^2 = \sigma^2 x_i$, then we have $\sum_s v_i^2 = \sum_s \sigma^4 x_i^2 = n\sigma^4 \bar{x}_s^2$, and $\sum_s v_i = \sum_s \sigma^2 x_i = n\bar{\sigma}^2 \bar{x}_s$.

Substituting these values in (9.48), we get

$$\text{Var}(\hat{V}_{WJ}) \doteq \frac{2A^{*2}}{n-1}\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sigma^4 \bar{x}_s^2 . \qquad (9.50)$$

Corollary 9.9

When $v_i = \sigma^2 x_i^2$, then we have

$$\sum_s v_i^2 = \sum_s \sigma^4 x_i^4 = n\sigma^4 \bar{x}_s^4 .$$

Under this situation the variance of the jack-knifed variance estimator $\hat{V}_{WJ}$ is given by

$$\mathrm{Var}(\hat{V}_{WJ}) \doteq \frac{2A^{*2}}{n}\left(\frac{\bar{x}}{\bar{x}_{\tilde{s}}}\right)^2 \bar{x}_s^4 . \tag{9.51}$$

We are now well equipped to find the relative mean square error in $\hat{V}_{WJ}$ which is approximately given by,

$$\text{Relative Mean Square Error}(\hat{V}_{WJ}) = \frac{\mathrm{Var}(\hat{V}_{WJ}) + (\mathrm{Bias}(\hat{V}_{WJ}))^2}{[E_\xi(\hat{T}_R - T)^2]^2} .$$

Now using (9.48), (9.37) and (9.32), the relative mean square error for $\hat{V}_{WJ}$ in case of general variance function is given by

$$\doteq \frac{\left(\frac{2A^2}{n}\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \frac{\bar{v}_s^2}{\bar{v}^2} + A^2\left(\left(\frac{\bar{x}}{\bar{x}_{\tilde{s}}} - 1\right) + \frac{1}{n}\left\{C_{x(s)}^2 - 2C_{xv(s)}\right\} - f\,\frac{\bar{x}_s}{\bar{x}}\left(\frac{\bar{x}_s \bar{v}_{\tilde{s}}}{\bar{x}_{\tilde{s}} \bar{v}_s} - 1\right)\right)\right)^2}{A^2} .$$

$$\doteq \frac{2}{n}\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \frac{\bar{v}_s^2}{\bar{v}^2} + \left(\frac{\bar{x}}{\bar{x}_s} - 1\right) + \frac{1}{n}\left(C_{x(s)}^2 - 2C_{xv(s)}\right) - f\,\frac{\bar{x}_s}{\bar{x}}\left(\frac{\bar{x}_s \bar{v}_{\tilde{s}}}{\bar{x}_{\tilde{s}} \bar{v}_s} - 1\right)\Bigg\}^2 , \tag{9.52}$$

$\doteq 0$, for large n and small f .

Similarly,for other variance functions, the relative mean square error reduces to zero provided n and $1/f$ are large. Thus, we note that the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ is almost stable under the model (9.15) subject to the above conditions and under the assumption that e_i's are normally distributed.

9.4 Implications of the Failure of the Regression Model $E(y_i) = \beta x_i$ on the Weighted Jack-knifed Variance Estimator $\hat{V}_{WJ}$ and the Actual Mean Square Error

In Section (9.3) we considered the effects of the variants of the variance functions on the bias and variance of the weighted jack-knifed variance estimator $\hat{V}_{WJ}$. In this section, we study the implications of changing the regression function in our model (9.15) on the actual mean square error and $\hat{V}_{WJ}$. We consider the general case when $E(Y_i) = h(x_i)$ and then illustrate the consequences through examples.

The weighted jack-knifed variance estimator $\hat{V}_{WJ}$ can be conveniently be expressed in the form $C \sum_{i=1}^{n} \ell_j^2(\underline{y})$, where $C = \frac{N}{f}\frac{(1-f)}{n-1}\left(\frac{\bar{x}}{\bar{x}_s}\right)^2$ is a constant and $\ell_j(\underline{y}) = y_j - \hat{R}_s x_j$. Since, we have

$$E[\ell(\underline{Y})^2] = Var\ell(\underline{Y}) + [E(\ell(\underline{Y}))]^2$$

therefore, we can write

$$E[C \sum_j \ell_j^2(\underline{Y})] = C \sum_j \left\{Var\ell_j(\underline{Y}) + [E\ell_j(\underline{Y})]^2\right\}$$

$$= C \sum_j Var\ell_j(\underline{Y}) + C \sum_j [\ell_j E(\underline{Y})]^2 \qquad (9.53)$$

Thus, we note that the expectation is the sum of two parts, the first part depends on the variance function but not on the regression function whereas the second is obtained by substituting $E(y_i)$ for y_i in the estimator. Since the variance function appears only in the first part

and the regression function appears only in the second part, we can examine separately the effects of changing these two elements in our model. As pointed out above we have already studied the behaviour of weighted jack-knifed variance estimator $\hat{V}_{WJ}$ under different variance functions and now examine the effects of changing the regression function in our model on the behaviour of $\hat{V}_{WJ}$. We will therefore consider the second part only.

Consider the model $y_i = h(x_i) + e_i$,

$E(y_i) = h(x_i)$, $Var(y_i) = v_i$ and $Cov(y_i, y_j) = 0$ for $i \neq j$. (9.54)

(a) Implications on the Variance Estimator $\hat{V}_{WJ}$

We have $\hat{V}_{WJ} = \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sum_s (y_i - \hat{R}x_i)^2/(n-1)$. Now, we can write

$$E_\xi(\hat{V}_{WJ}) = C \sum_i^n \left\{Var\ell_i(\underline{y})\right\} + C \sum_i^n \left\{\ell_i E(\underline{y})\right\}^2 .$$

$$= C \sum_i^n Var(y_i - \hat{R}x_i) + C \sum_i^n [E(y_i - \hat{R}x_i)]^2 . \quad (9.55)$$

Now if the model (9.15) is correct then second term is zero and $E_\xi(\hat{V}_{WJ})$ in this case is given by (9.22). However, when the model (9.15) is not correct, i.e. $E(y_i) = h(x_i)$, then the second term in (9.55) is not zero and it is this term with which we are mainly concerned in this part. Thus under the model (9.54) we have

$$E_\xi(\hat{V}_{WJ}) = C \sum_{i=1}^n \left[E_\xi(y_i) - E\left(\frac{\sum_s y_i}{\sum_s x_i}\right)x_i\right]^2 + A\frac{\bar{x}}{\bar{x}_s}\left[1 + \frac{1}{n}\left(C^2_{x(s)} - 2C_{xv(s)}\right)\right]$$

The second term is simply (9.22) and we denote the first term by b_{WJ} .

$$b_{WJ} = C \sum_{i=1}^n \left[h(x_i) - \frac{\sum_s h(x_i)}{\sum x_i} x_i\right]^2 ,$$

$$= C \sum_{i=1}^n \left[h(x_i) - \frac{\bar{h}_s}{\bar{x}_s} x_i\right]^2 , \quad \text{where } \bar{h}_s = \frac{1}{n}\sum_{i=1}^n h(x_i) .$$

$$= b_{WJ} . \quad (9.56)$$

where

$$b_{WJ} = \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sum_{i=1}^{n}\left(h(x_i) - \frac{\bar{h}_s}{\bar{x}_s}x_i\right)^2/(n-1) .$$

The quantity b_{WJ} measures the difference between the fitted values $(\hat{\beta}x)$ and the true regression function for the sample points. It thus represents the implications of departure from the simple regression model $E(y_i) = \beta x_i$ on the estimate $\hat{V}_{WJ}$. The ratio $\frac{\bar{h}_s}{\bar{x}_s}$ implies that if a straight line, through the origin, is fitted to the true regression function $h(x)$ at the sample points by minimising $\sum_s (h(x_i)-\beta x_i)^2/x_i$, we get the fitted line $\hat{\beta}x$, where $\hat{\beta} = \frac{\bar{h}_s}{\bar{x}_s}$.

(b) Implications on the Actual MSE

Now we examine the implications of the failure of regression model $E(y_i) = \beta x_i$ on the actual mean square error in order to facilitate comparison. Under the model (9.54) and following the same procedure as above, we have,

$$\left[E_\xi(\hat{T}_R - T)\right]^2 = \left(\sum_{i=1}^{N} x_i\right)^2 \left[E_\xi \frac{\sum_s y_i}{\sum_s x_i} - E_\xi \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i}\right]^2 , \qquad (9.57)$$

$$= N^2 \frac{2}{x}\left[\frac{\sum_s h(x_i)}{\sum_s x_i} - \frac{\sum_{i=1}^{N} h(x_i)}{\sum_{i=1}^{N} x_i}\right]^2 ,$$

$$= N^2 \frac{2}{x}\left(\frac{\bar{h}_s}{\bar{x}_s} - \frac{\bar{h}}{\bar{x}}\right)^2 , \quad \text{where } \bar{h} = \frac{1}{N}\sum_{i=1}^{N} h(x_i) .$$

$$= b_{MSE} . \qquad (9.58)$$

where

$$b_{MSE} = N^2 \frac{2}{x}\left(\frac{\bar{h}_s}{\bar{x}_s} - \frac{\bar{h}}{\bar{x}}\right)^2 = N^2(\bar{h} - \hat{\beta}\bar{x})^2 , \quad \text{where } \hat{\beta} = \frac{\bar{h}_s}{\bar{x}_s} \qquad (9.59)$$

We note that the quantity $b_{MSE} = N^2(\bar{h}-\hat{\beta}\bar{x})^2$ measures the closeness of the average of the values on the fitted line $\hat{\beta}x$ at the points $x_1, x_2, \ldots x_N$ and the average values of the true regression function at these points. It may be noted that a poor fit point by point may result in a small value for b_{MSE} and not b_{WJ}. Also we observe that in the situation when the regression function in our model changes b_{WJ} does not appear to have any clear advantage in approximating b_{MSE}. This latter statement is illustrated through examples 9.1 and 9.2 below.

Example 9.1

Consider the model $y_i = \beta_o+\beta_1 x_i+e_i$, $E(y_i) = \beta_o+\beta_1 x_i$, $Var(y_i) = v_i$, $Cov(y_i,y_j) = 0$ for $i \neq j$. Under this model we work out the bias for the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ and the actual mean square error $E_\xi(\hat{T}_R-T)^2$. First we consider the actual mean square error.

$$\text{(a)} \quad E_\xi(\hat{T}_R-T)^2 = E_\xi\left[\frac{\sum_s y_i}{\sum_s x_i}\sum_{i=1}^N x_i - \sum_{i=1}^N y_i\right]^2,$$

$$= E_\xi\left[\frac{\sum_s (\beta_o+\beta_1 x_i+e_i)}{\sum_s x_i}\left(\sum_{i=1}^N x_i\right) - \sum_{i=1}^N (\beta_o+\beta_1 x_i+e_i)\right]^2,$$

$$= E_\xi\left[\frac{\sum_s \beta_o}{\sum_s x_i}\sum_{i=1}^N x_i + \frac{\sum_s e_i}{\sum_s x_i}\sum_{i=1}^N x_j - \sum_{i=1}^N \beta_o - \sum_{i=1}^N e_i\right]^2,$$

$$= E_\xi\left[\frac{\beta_o}{\bar{x}_s} N\bar{x} - N\beta_o + \frac{\sum_s e_i}{\sum_s x_i}\sum_{i=1}^N x_i - \sum_{i=1}^N e_i\right]^2,$$

$$= E_\xi \left[N\beta_o \frac{(\bar{x}-\bar{x}_s)}{\bar{x}_s} + \frac{\sum_s e_i}{\sum_s x_i} \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} e_i \right]^2 ,$$

$$= E_\xi \left[N^2\beta_o^2 \frac{(\bar{x}-\bar{x}_s)^2}{\bar{x}_s^2} + \frac{(\sum e_i)^2}{(\sum_s x_i)^2} \left(\sum_{i=1}^{N} x_i \right)^2 + \left(\sum_{i=1}^{N} e_i \right)^2 + 2N\beta_o \frac{(\bar{x}-\bar{x}_s)}{\bar{x}_s} \left(\frac{\sum_s e_i}{\sum_s x_i} \right) \sum_{i=1}^{N} x_i \right.$$

$$\left. - 2N\beta_o \frac{(\bar{x}-\bar{x}_s)}{\bar{x}_s} \sum_{i=1}^{N} e_i - 2 \left(\frac{\sum_s e_i}{\sum_s x_i} \sum_{i=1}^{N} x_i \right) \left(\sum_{i=1}^{N} e_i \right) \right] ,$$

$$= N^2\beta_o^2 \frac{(\bar{x}-\bar{x}_s)^2}{\bar{x}_s^2} + A \left[1 + f \frac{\bar{x}_s}{\bar{x}} \left(\frac{\bar{x}_s \tilde{v}_s - \tilde{x}_s \bar{v}_s}{\bar{x}_s \bar{v}_s} \right) \right] . \qquad (9.60)$$

We note from (9.32) and (9.60) that when the model is not correct, that is when the regression function is changed, the bias of the actual mean square error is $N^2\beta_o^2 \frac{(\bar{x}-\bar{x}_s)^2}{\bar{x}_s^2}$. It is clear that this bias depends on $\bar{x}_s$, that is, if $\bar{x}_s$ is much larger or smaller than $\bar{x}$ then the bias will be large. This will, however, vanish if the sample is well balanced on x and x^2 (i.e. $\bar{x} = \bar{x}_s$; $\bar{x}_s^2 = \bar{x}^2$) . Similar expressions are obtained for other variance functions.

Now we examine the behaviour of the weighted jack-knifed variance estimator $\hat{V}_{WJ}$. Under the above model, we have

$$\hat{V}_{WJ} = \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sum_{i=1}^{n} (y_i - \hat{R}_s x_i)^2/n-1 ,$$

$$= \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sum_{i=1}^{n} \left[\beta_o + \beta_1 x_i + e_i - \frac{\sum_s (\beta_o + \beta_1 x_i + e_i)}{\sum_s x_i} x_i \right]^2 /n-1 ,$$

$$= \frac{N}{1}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sum_{i=1}^{n} \left[\beta_o - \frac{\beta_o x_i}{\bar{x}_s} + e_i - \frac{(\sum_s e_i) x_i}{\sum_s x_i} \right]^2 /n-1 ,$$

$$= \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \sum_{i=1}^{n} \left[\beta_o \frac{(\bar{x}_s - x_i)}{\bar{x}_s} + e_i - \frac{(\sum_s e_i) x_i}{\sum_s x_i} \right]^2 /n-1 ,$$

$$E_\xi(\hat{V}_{WJ}) = \frac{N}{f}(1-f)\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 E_\xi\left[\frac{\beta_o^2(\bar{x}_s-x_i)^2}{\bar{x}_s^2} + e_i^2 + \frac{(\sum_s e_i)^2 x_i^2}{(\sum_s x_i)^2} + \frac{2\beta_o(\bar{x}_s-x_i)}{\bar{x}_s} e_i \right.$$

$$\left. - \frac{2\beta_o(\bar{x}_s-x_i)}{\bar{x}_s}\frac{(\sum_s e_i)x_i}{(\sum_s x_i)} - \frac{2e_i(\sum_s e_i)x_i}{\sum_s x_i}\right] /n-1 ,$$

$$= \frac{N}{f}\frac{(1-f)}{n-1}\left(\frac{\bar{x}}{\bar{x}_s}\right)^2 \frac{\beta_o^2(\bar{x}_s-x_i)^2}{\bar{x}_s^2} + A\frac{\bar{x}}{\bar{x}_s}\left[1 + \frac{1}{n}(C_{x(s)}^2 - 2C_{xv(s)})\right] . \quad (9.61)$$

we observe from expressions (9.22) and (9.61) that the bias in $\hat{V}_{WJ}$ due to departure from the regression assumption $E(y_i) = \beta x_i$ is $\frac{\beta_o^2(\bar{x}_s-x_i)^2}{\bar{x}_s^2}$ which is a positive quantity. Also, we note from (9.61) and (9.60) that b_{WJ} does not approximate b_{MSE}. This failure of the simple regression model has a conservative effect on $\hat{V}_{WJ}$. This also holds true for other variance functions under the example of this model.

Example 9.2

Consider the model $y_i = \beta_o+\beta_1 x_i+\beta_2 x_i^2+e_i$, $E(y_i)=\beta_o+\beta_1 x_i+\beta_2 x_i^2$, $Var(y_i) = v_i$, $Cov(y_i,y_j) = 0$ for $i\neq j$. Since the other expressions both for $\hat{V}_{WJ}$ and $E_\xi(\hat{T}_R-T)^2$, are the same as in Section 9.3, we only derive expressions for b_{MSE} and b_{WJ} which is our main concern here. Under the above model and from (9.57), we have

(i) $$b_{MSE} = \left(\sum_{i=1}^N x_i\right)^2 \left[\frac{\sum_s h(x_i)}{\sum_s x_i} - \frac{\sum_{i=1}^N h(x_i)}{\sum_{i=1}^N x_i}\right]^2 ,$$

$$= N^2\bar{x}^2\left[\frac{\sum_s (\beta_o+\beta_1 x_i+\beta_2 x_i^2)}{\sum_s x_i} - \frac{\sum_{i=1}^N (\beta_o+\beta_1 x_i+\beta_2 x_i^2)}{\sum_{i=1}^N x_i}\right]^2 , \because h(x_i)=\beta_o+\beta_1 x_i+\beta_2 x_i^2$$

$$= N^2\bar{x}^2\left[\frac{\beta_o}{\bar{x}_s} - \frac{\beta_o}{\bar{x}} + \beta_2\frac{\sum_s x_i^2}{\sum_s x_i} - \beta_2\frac{\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} x_i}\right]^2$$

$$= N^2\bar{x}^2\left[\beta_o\frac{(\bar{x}-\bar{x}_s)}{\bar{x}\bar{x}_s} + \beta_2\frac{(\bar{\bar{x}}\bar{x}_s - \bar{x}_s\bar{\bar{x}})}{\bar{x}\bar{x}_s}\right]^2, \quad \text{where} \quad \bar{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N} x_i^2,$$

$$= N^2\left[\beta_o\frac{(\bar{x}-\bar{x}_s)}{\bar{x}_s} + \beta_2\frac{(\bar{\bar{x}}\bar{x}_s - \bar{x}_s\bar{\bar{x}})}{\bar{x}_s}\right]^2, \quad \text{where} \quad \bar{\bar{x}}_s = \frac{1}{n}\sum_{i=1}^{n} x_i^2. \qquad (9.62)$$

Thus we note that when $E(y_i) = \beta_o+\beta_1 x_i+\beta_2 x_i^2$, the bias in the actual mean square error is $N^2\left[\beta_o\frac{(\bar{x}-\bar{x}_s)}{\bar{x}_s} + \beta_2\frac{(\bar{\bar{x}}\bar{x}_s - \bar{x}_s\bar{\bar{x}})}{\bar{x}_s}\right]^2$ which becomes zero if the sample is balanced both on x and x^2.

(ii) Under the same model we have

$$b_{WJ} = C\sum_{i=1}^{n}\left(h(x_i) - \frac{\bar{h}_s}{\bar{x}_s}x_i\right)^2,$$

$$= C\sum_{i=1}^{n}\left[\beta_o+\beta_1 x_i+\beta_2 x_i^2 - \frac{1}{n}\frac{\sum_s(\beta_o+\beta_1 x_i+\beta_2 x_i^2)}{\sum_s x_i}\right]^2,$$

$$= C\sum_{i=1}^{n}\left[\beta_o - \frac{\beta_o x_i}{\bar{x}_s} + \beta_2 x_i^2 - \beta_2\frac{\sum_s x_i^3}{\sum_s x_i}\right]^2,$$

$$= C\sum_{i=1}^{n}\left[\frac{\beta_o(\bar{x}_s - x_i)}{\bar{x}_s} + \beta_2\frac{(x_i^2\bar{x}_s - \bar{\bar{x}}_s)}{\bar{x}_s}\right]^2, \qquad (9.63)$$

Thus we note that b_{WJ} does not reduce to zero even if the sample is well balanced on x and x^2. This implies that when the regression function changes in our model the weighted jack-knifed variance estimator $\hat{V}_{WJ}$ has a positive bias. Thus we conclude from these results that the failure of the simple regression model $E(y_i) = \beta x_i$ has a conservative effect on the weighted jack-knifed variance estimator $\hat{V}_{WJ}$.

9.5 Comparison of the Bias Reduction Properties of the Weighted and Unweighted Jack-knifed Ratio Estimators Using Taylor's Series Expansion

In this section we compare the weighted and the unweighted jack-knifed ratio estimators in terms of their bias reduction properties under the model ξ.

Lemma 9.3

Let $\theta = f(R)$ where $f(.)$ is a smooth function of the ratio parameter $R = \frac{\bar{Y}}{\bar{X}}$ and f has a continuous first and second derivative f' and f'' respectively. Further, let the estimator of R be $\hat{R} = \frac{\sum_s y_i}{\sum_s x_i}$ and $\hat{\theta} = f(\hat{R})$. Then

$$\hat{\theta}-\hat{\theta}_{-i} = (\hat{R}-\hat{R}_{-i})f'(\hat{R}) - \frac{1}{2}(\hat{R}-\hat{R}_{-i})^2 f''(\hat{R}) + \frac{1}{6}(\hat{R}-\hat{R}_{-i})^3 f'''(\hat{R}) + \ldots \qquad (9.64)$$

where $\hat{R}-\hat{R}_{-i} = \frac{y_i - \hat{R}x_i}{(1-w_i)\sum_s x_i}$ and $w_i = \frac{x_i}{\sum_s x_i}$ (see Lemma 9.1).

Proof

Using Taylor's Series expansion, we have

$$f(\hat{R}_{-i}) = f(\hat{R}) + (\hat{R}_{-i}-\hat{R})f'(\hat{R}) + \frac{1}{2}(\hat{R}_{-i}-\hat{R}))^2 f''(\hat{R}) + \frac{1}{6}(\hat{R}_{-i}-\hat{R})^3 f'''(\hat{R}) + \ldots$$

Since $f(\hat{R}_{-i}) = \hat{\theta}_{-i}$, we can write

$$\hat{\theta}_{-i} = \hat{\theta} + (\hat{R}_{-i}-\hat{R})f'(\hat{R}) + \frac{1}{2}(\hat{R}_{-i}-\hat{R})^2 f''(\hat{R}) + \frac{1}{6}(\hat{R}_{-i}-\hat{R})^3 f'''(\hat{R}) + \ldots$$

Hence

$$\hat{\theta}-\hat{\theta}_{-i} = (\hat{R}-\hat{R}_{-i})f'(\hat{R}) - \frac{1}{2}(\hat{R}-\hat{R}_{-i})^2 f''(\hat{R}) + \frac{1}{6}(\hat{R}-\hat{R}_{-i})^3 f'''(\hat{R}) + \ldots \quad (9.65)$$

9.5.1 The Weighted Jack-knifed Ratio Estimator

We now apply (9.65) to derive an expression for the weighted jack-knifed ratio estimator. This result is expressed as Lemma 9.4.

Lemma 9.4

Under the conditions of Lemma 9.3, the weighted jack-knifed ratio estimator $\tilde{\theta}_{WJ}$ can be expressed as

$$\tilde{\theta}_{WJ} = f(R)+(\hat{R}-R)f'(R) + \frac{1}{2}(\hat{R}-R)^2 f''(R) - \frac{1}{2}\sum_{i=1}^{n} \frac{(1-w_i)^{-1}e_i^{*2}f''(\hat{R})}{(\sum_s x_i)^2} + \frac{1}{6}\sum_{i=1}^{n} \frac{(1-w_i)^{-2}e_i^{*3}}{(\sum_s x_i)^3} f'''(\hat{R}) \quad , \quad (9.66)$$

where $e_i^* = y_i - \hat{R}x_i$

Proof

For the weighted jack-knife the pseudovalues are defined as

$$\tilde{Q}_i = \hat{\theta} + n(1-w_i)(\hat{\theta}-\hat{\theta}_{-i}) \quad (9.67)$$

Substituting from (9.65) for $\hat{\theta} - \hat{\theta}_{-i}$ in (9.67), we have

$$\tilde{Q}_i = \hat{\theta} + n(1-w_i)[(\hat{R}-\hat{R}_{-i})f'(\hat{R}) - \frac{1}{2}(\hat{R}-\hat{R}_{-i})^2 f''(\hat{R}) + \frac{1}{6}(\hat{R}-\hat{R}_{-i})^3 f'''(\hat{R}) + \ldots] \quad ,$$

$$= \hat{\theta} + n(1-w_i)\left[\frac{e_i^*}{(1-w_i)\sum_s x_i} f'(\hat{R}) - \frac{1}{2}\frac{e_i^{*2}}{(1-w_i)^2(\sum_s x_i)^2} f''(\hat{R}) + \frac{1}{6}\frac{e_i^{*3}}{(1-w_i)^3(\sum_s x_i)^3} f'''(\hat{R}) + \ldots\right] \quad ,$$

$$= \hat{\theta} + n\frac{e_i^*}{\sum_s x_i} f'(\hat{R}) - \frac{n}{2}\frac{(1-w_i)^{-1}e_i^{*2}}{(\sum_s x_i)^2} f''(\hat{R}) + \frac{n}{6}(1-w_i)^{-2}\frac{e_i^{*3}}{(\sum_s x_i)^3} f'''(\hat{R}) \quad . \quad (9.68)$$

The weighted jack-knifed ratio estimator is the average of the pseudo values $\tilde{Q}_i$ and is given by

$$\tilde{\theta}_{WJ} = \frac{1}{n}\sum_{i=1}^{n}\tilde{Q}_i = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\theta} + \frac{ne_i^*}{\sum_s x_i} f'(\hat{R}) - \frac{n}{2}\frac{(1-w_i)^{-1}e_i^{*2}}{(\sum_s x_i)^2} f''(\hat{R}) + \frac{n}{6}\frac{(1-w_i)^{-2}e_i^{*3} f'''(\hat{R})}{(\sum_s x_i)^3}\right]$$

$$= \hat{\theta} + \sum_{i=1}^{n}\frac{e_i^*}{\sum_s x_i} f'(\hat{R}) - \frac{1}{2}\sum_{i=1}^{n}\frac{(1-w_i)^{-1}e_i^{*2}}{(\sum_s x_i)^2} f''(\hat{R}) + \frac{1}{6}\sum_{i=1}^{n}(1-w_i)^{-2}\frac{e_i^{*3}}{(\sum_s x_i)^3} f'''(\hat{R}) .$$

Now

$$\sum_{i=1}^{n} e_i^* = \sum_{i=1}^{n}(y_i - \hat{R}x_i) = n(\bar{y}_s - \frac{\bar{y}_s}{\bar{x}_s}\bar{x}_s) = 0 .$$

Thus

$$\tilde{\theta}_{WJ} = \hat{\theta} - \frac{1}{2}\sum_{i=1}^{n}\frac{(1-w_i)^{-1}e_i^{*2}}{(\sum_s x_i)^2} f''(\hat{R}) + \frac{1}{6}\sum_{i=1}^{n}(1-w_i)^{-2}\frac{e_i^{*3}}{(\sum_s x_i)^3} f'''(\hat{R}) ,$$

$$= f(\hat{R}) - \frac{1}{2}\sum_{i=1}^{n}\frac{(1-w_i)^{-1}e_i^{*2}}{(\sum_s x_i)^2} f''(\hat{R}) + \frac{1}{6}\sum_{i=1}^{n}(1-w_i)^{-2}\frac{e_i^{*3}}{(\sum_s x_i)^3} f'''(\hat{R}) \quad \because \hat{\theta} = f(\hat{R}) . \tag{9.69}$$

Now expanding $f(\hat{R})$, we have $f(\hat{R}) = f(R) + (\hat{R}-R)f'(R) + \frac{1}{2}(\hat{R}-R)^2 f''(R)$ and substituting this in (9.69), we get

$$\tilde{\theta}_{WJ} = f(R) + (\hat{R}-R)f'(R) + \frac{1}{2}(\hat{R}-R)^2 f''(R) - \frac{1}{2}\sum_{i=1}^{n}\frac{(1-w_i)^{-1}e_i^{*2} f''(\hat{R})}{(\sum_s x_i)^2}$$

$$+ \frac{1}{6}\sum_{i=1}^{n}\frac{(1-w_i)^{-2}e_i^{*3}}{(\sum_s x_i)^3} f'''(\hat{R}) . \tag{9.70}$$

9.5.1a Bias of the Weighted Jack-knifed Ratio Estimator

Now we find the bias of the weighted jack-knifed ratio estimator under the model ξ with different variance functions. These results are expressed as Theorem 9.4.

Theorem 9.4

If the conditions of Lemma 9.3 hold then, under the model $y_i = \beta x_i + e_i$ with $E(e_i) = 0$, $Var(e_i) = v_i$ and $Cov(e_i, e_j) = 0$ for $i \neq j$, the bias of the weighted jack-knife estimator $\tilde{\theta}_{WJ}$ is of $O(n^{-2})$.

Proof

From (9.70), we have

$$\tilde{\theta}_{WJ} = f(R)+(\hat{R}-R)f'(R)+ \frac{1}{2}(\hat{R}-R)^2 f''(R)- \frac{1}{2} \sum_{i=1}^{n} \frac{(1-w_i)^{-1} e_i^{*2}}{(\sum_s x_i)^2} f''(\hat{R})$$

$$+ \frac{1}{6} \sum_{i=1}^{n} \frac{(1-w_i)^{-2} e_i^{*3}}{(\sum_s x_i)^3} f'''(\hat{R}) \ .$$

Under the above model the expected value of $\tilde{\theta}_{WJ}$ is

$$E_\xi(\tilde{\theta}_{WJ}) = f(R)+E_\xi(\hat{R}-R)f'(R)+ \frac{1}{2} E_\xi(\hat{R}-R)^2 f''(R)- \frac{1}{2} \sum_{i=1}^{n} \frac{(1-w_i)^{-1} E_\xi(e_i^{*2})}{(\sum_s x_i)^2} f''(\hat{R})$$

$$+ \frac{1}{6} \sum_{i=1}^{n} \frac{(1-w_i)^{-2} E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(\hat{R}) \ . \qquad (9.71)$$

Now, $E_\xi(e_i^{*2}) = E_\xi(y_i - \hat{R}x_i)^2 = E_\xi[(\hat{\beta}-\beta)^2 x_i^2 - 2e_i(\hat{\beta}-\beta)x_i + e_i^2]$. Using (9.18) and (9.19), we get

$$E_\xi(e_i^{*2}) = v_i - 2 \frac{x_i v_i}{(\sum_s x_i)} + \frac{x_i^2 \sum_s v_i}{(\sum_s x_i)^2} \ . \qquad (9.72)$$

and

$$E_\xi(\hat{R}-R)^2 = \frac{\sum_s v_i}{(\sum_s x_i)^2} \tag{9.73}$$

Substituting (9.72) and (9.73) in (9.71), we get

$$E_\xi(\tilde{\theta}_{WJ}) = f(R) + \frac{1}{2}\frac{\sum_s v_i}{(\sum_s x_i)^2} f''(R) - \frac{1}{2}\sum_{i=1}^{n}(1-w_i)^{-1}\left\{v_i - 2\frac{x_i v_i}{(\sum_s x_i)} + \frac{x_i^2 \sum_s v_i}{(\sum_s x_i)^2}\right\}\frac{f''(R)}{(\sum_s x_i)^2}$$

$$+ \frac{1}{6}\sum_{i=1}^{n}\frac{(1-w_i)^{-2}E_\xi(e_i^{*3})f'''(R)}{(\sum_s x_i)^3},$$

$$= \theta + \frac{1}{2}\frac{\sum_s v_i}{(\sum_s x_i)^2} f''(R) - \frac{1}{2}\sum_{i=1}^{n}(1+w_i+w_i^2)\left(\frac{v_i - 2v_i w_i + w_i^2 \sum_s v_i}{(\sum_s x_i)^2}\right) f''(R)$$

$$+ \frac{1}{6}\sum_{i=1}^{n}\frac{(1+2w_i)E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R),$$

$$= \theta + \sum_{i=1}^{n}\frac{x_i v_i}{(\sum_s x_i)^3} f''(R) - \frac{1}{2}\sum_{i=1}^{n}\frac{x_i^2 \sum_s v_i}{(\sum_s x_i)^4} f''(R) + \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R). \tag{9.74}$$

Hence

$$\text{Bias}(\tilde{\theta}_{WJ}) = E_\xi(\tilde{\theta}_{WJ}) - \theta$$

$$= \sum_{i=1}^{n}\frac{x_i v_i}{(\sum_s x_i)^3} f''(R) - \frac{1}{2}\sum_{i=1}^{n}\frac{x_i^2 \sum_s v_i}{(\sum_s x_i)^4} f''(R) + \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R).$$

$$= O(n^{-2}). \tag{9.75}$$

Corollary 9.10

If in the model of Theorem 9.4, $v_i = \sigma^2$ then

$$E_\xi(e_i^{*2}) = \sigma^2\left(1 - 2\frac{x_i}{\sum_s x_i} + \frac{nx_i^2}{(\sum_s x_i)^2}\right), \quad E_\xi(\hat{R}-R)^2 = \frac{\sigma^2 n}{(\sum_s x_i)^2},$$

and

$$E_\xi(\tilde{\theta}_{WJ}) = \theta + \frac{1}{2}\sum_{i=1}^{n}\frac{w_i\sigma^2}{(\sum_s x_i)^2}f''(R) - \frac{1}{2}(n-1)\sum_{i=1}^{n}\frac{w_i^2\sigma^2}{(\sum_s x_i)^2}f''(R)$$
$$+ \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3}f'''(R).$$

In this case,

$$\text{Bias}(\tilde{\theta}_{WJ}) = \frac{1}{2}\sum_{i=1}^{n}\frac{w_i\sigma^2}{(\sum_s x_i)^2}f''(R) - \frac{1}{2}(n-1)\sum_{i=1}^{n}\frac{w_i^2\sigma^2}{(\sum_s x_i)^2}f''(R)$$
$$+ \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3}f'''(R).$$
$$= O(n^{-2}). \qquad (9.76)$$

Corollary 9.11

If the variance function in the model of Theorem 9.4 is proportional to x_i, that is, $v_i = \sigma^2 x_i$, then

$$E_\xi(e_i^{*2}) = \sigma^2(x_i - \frac{x_i^2}{\Sigma x_i}) = \sigma^2 x_i(1-w_i), \quad E_\xi(\hat{R}-R)^2 = \frac{\sigma^2}{\sum_s x_i},$$

and

$$E_\xi(\tilde{\theta}_{WJ}) = \theta + \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3}f'''(R).$$

Thus

$$\text{Bias}(\tilde{\theta}_{WJ}) = \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3}f'''(R) = O(n^{-2}). \qquad (9.77)$$

This implies that in this case the weighted jack-knifed ratio estimator is unbiased to $O(n^{-1})$.

Corollary 9.12

If the variance function in the model of Theorem 9.4 is quadratic in x, that is, $v_i = \sigma^2 x_i^2$ then $E_\xi(\hat{R}-R)^2 = \sigma^2 \sum_s x_i^2 / (\sum_s x_i)^2$,

$$E_\xi(e_i^{*2}) = \sigma^2\left(x_i^2 - \frac{x_i^3}{\sum_s x_i} + \frac{x_i^4}{(\sum_s x_i)^4}\right), \text{ and}$$

$$E_\xi(\tilde{\theta}_{WJ}) = \theta + \frac{\sigma^2 \sum_{i=1}^{n} x_i^3}{(\sum_s x_i)^3} f''(R) - \frac{\sigma^2}{2} \sum_{i=1}^{n} \frac{\sum_s x_i^4}{(\sum_s x_i)^4} f''(R) + \frac{1}{6} \sum_{i=1}^{n} \frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R).$$

Hence

$$\text{Bias}(\tilde{\theta}_{WJ}) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^3}{(\sum_s x_i)^3} f''(R) - \frac{\sigma^2}{2} \sum_{i=1}^{n} \frac{\sum_s x_i^4}{(\sum_s x_i)^4} f''(R) + \frac{1}{6} \sum_{i=1}^{n} \frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R).$$

$$= O(n^{-2}) \qquad (9.78)$$

Thus we note from Theorem 9.4 and the Corollaries 9.10, 9.11 and 9.12 that, in general, the bias of the weighted jack-knife is of $O(n^{-2})$. In the next section we give the bias of the unweighted jack-knife under the above situations.

9.5.2 The Unweighted Jack-knifed Ratio Estimator

The corresponding developments for the ordinary jack-knifed ratio estimator under the model ξ with different variance functions are given as under:

Lemma 9.5

If the conditions of Lemma 9.3 hold, then

$$\tilde{\theta}_{UWJ} = f(R)+(\hat{R}-R)f'(R)+\frac{1}{2}(\hat{R}-R)f''(R) - \frac{n-1}{2n}\sum_{i=1}^{n}\frac{(1-w_i)^{-2}e_i^{*2}}{(\sum_s x_i)^2}f''(\hat{R})$$

$$+\frac{n-1}{6n}\sum_{i=1}^{n}\frac{(1-w_i)^{-3}e_i^{*3}}{(\sum_s x_i)^3}f'''(\hat{R}),$$

where $\tilde{\theta}_{UWJ}$ is the unweighted jack-knifed ratio estimator.

Proof:

The pseudovalues for the unweighted ratio estimator are

$$\tilde{\theta}_i = n\hat{\theta}-(n-1)\hat{\theta}_{-i} = \hat{\theta}+(n-1)(\hat{\theta}-\hat{\theta}_{-i}) \qquad (9.79)$$

Substituting for $\hat{\theta}-\hat{\theta}_{-i}$ from (9.65) in (9.79), we get

$$\tilde{\theta}_i = \hat{\theta}+(n-1)[(\hat{R}-\hat{R}_{-i})f'(\hat{R})-\frac{1}{2}(\hat{R}-\hat{R}_{-i})^2f''(\hat{R})+\frac{1}{6}(\hat{R}-\hat{R}_{-i})^3f'''(\hat{R})+\ldots],$$

$$= \hat{\theta}+(n-1)\left[\frac{e_i^*f'(\hat{R})}{(1-w_i)\sum_s x_i}-\frac{1}{2}\frac{e_i^{*2}}{(1-w_i)^2(\sum_s x_i)^2}f''(\hat{R})+\frac{1}{6}\frac{e_i^{*3}}{(1-w_i)^3(\sum_s x_i)^3}f'''(\hat{R})\right. \qquad (9.80)$$

$$\left. + \text{terms of higher order}\right].$$

The unweighted jack-kifed ratio estimator is the average of the unweighted pseudo-values $\tilde{\theta}_i$ and is given by

$$\tilde{\theta}_{UWJ} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i = \hat{\theta}+\frac{n-1}{n}\left[\sum_{i=1}^{n}\frac{e_i^*}{(1-w_i)\sum_s x_i}f'(\hat{R})-\frac{1}{2}\sum_{i=1}^{n}\frac{e_i^{*2}}{(1-w_i)^2(\sum_s x_i)^2}f''(\hat{R})\right.$$

$$\left.+\frac{1}{6}\sum_{i=1}^{n}\frac{e_i^{*3}}{(1-w_i)^3(\sum_s x_i)^3}f'''(\hat{R})\right],$$

$$= \hat{\theta}+\frac{n-1}{n}\sum_{i=1}^{n}\frac{e_i^*}{(1-w_i)\sum_s x_i}f'(\hat{R})-\frac{n-1}{2n}\sum_{i=1}^{n}\frac{e_i^{*2}}{(1-w_i)^2(\sum_s x_i)^2}f''(\hat{R})$$

$$+\frac{n-1}{6n}\sum_{i=1}^{n}\frac{e_i^{*3}}{(1-w_i)^3(\sum_s x_i)^3}f'''(\hat{R}),$$

$$= f(\hat{R})-\frac{n-1}{2n}\sum_{i=1}^{n}\frac{(1-w_i)^{-2}e_i^{*2}}{(\sum_s x_i)^2}f''(\hat{R})+\frac{n-1}{6n}\sum_{i=1}^{n}\frac{(1-w_i)^{-3}e_i^{*3}}{(\sum_s x_i)^3}f'''(\hat{R}). \qquad (9.81)$$

Now expanding $f(\hat{R})$ and substituting in (9.81), we get

$$\tilde{\theta}_{UWJ} = f(R)+(\hat{R}-R)f'(R)+\frac{1}{2}(\hat{R}-R)f''(R)-\frac{n-1}{2n}\sum_{i=1}^{n}\frac{(1-w_i)^{-2}e_i^{*2}}{(\sum_s x_i)^2}f''(\hat{R}) + \frac{n-1}{6n}\sum_{i=1}^{n}\frac{(1-w_i)^{-3}e_i^{*3}}{(\sum_s x_i)^3}f'''(\hat{R}) \cdot \qquad (9.82)$$

9.5.2a Bias of the Unweighted Jack-knifed Ratio Estimator

Now we give the bias of the unweighted jack-knifed ratio estimator under the model ξ with general variance function. This result is expressed as Theorem 9.5. The results for other variance functions are given as the corollaries of Theorem 9.5.

Theorem 9.5

If the conditions of Lemma 9.3 hold, then under the model $y_i = \beta x_i + e_i$ with $E(e_i) = 0$, $Var(e_i) = v_i$ and $Cov(e_i, e_j) = 0$ for $i \neq j$, the bias of the unweighted jack-knifed ratio estimator is of $O(n^{-2})$.

Proof

From (9.82), we have

$$\tilde{\theta}_{UWJ} = f(R)+(\hat{R}-R)f'(R)+\frac{1}{2}(\hat{R}-R)^2f''(R)-\frac{n-1}{2n}\sum_{i=1}^{n}\frac{(1-w_i)^{-2}e_i^{*2}}{(\sum_s x_i)^2}f''(\hat{R}) + \frac{n-1}{6n}\sum_{i=1}^{n}\frac{(1-w_i)^{-3}e_i^{*3}}{(\sum_s x_i)^3}f'''(\hat{R}) \cdot$$

Under the above model and using (9.72) and (9.73), the expected value of the unweighted jack-knifed estimator is

$$E_\xi(\tilde{\theta}_{UWJ}) = f(R) + \frac{1}{2}\frac{\sum_s v_i}{(\sum_s x_i)^2} f''(R) - \frac{n-1}{2n}\sum_{i=1}^{n} \frac{(1-w_i)^{-2}\left(v_i - 2\frac{v_i x_i}{\sum_s x_i} + \frac{x_i^2 \sum_s v_i}{(\sum_s x_i)^2}\right)}{(\sum_s x_i)^2} f''(R)$$

$$+ \frac{n-1}{6n}\sum_{i=1}^{n} \frac{(1-w_i)^{-3} E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R) ,$$

$$= \theta + \frac{1}{2}\frac{\sum_s v_i}{(\sum_s x_i)^2} f''(R) - \frac{n-1}{2n}\sum_{i=1}^{n}\frac{v_i}{(\sum_s x_i)^2} f''(R) + \frac{n-1}{n}\sum_{i=1}^{n}\frac{v_i x_i}{(\sum_s x_i)^3} f''(R)$$

$$- \frac{n-1}{2n}\sum_{i=1}^{n}\frac{x_i^2 \sum_s v_i}{(\sum_s x_i)^4} f''(R) + \frac{n-1}{6n}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R) .$$

Now bias is given by

$$\text{Bias}(\tilde{\theta}_{UWJ}) = E_\xi(\tilde{\theta}_{UWJ}) - \theta = \frac{1}{2}\frac{\sum_s v_i}{(\sum_s x_i)^2} f''(R) - \frac{n-1}{2n}\sum_{i=1}^{n}\frac{v_i}{(\sum_s x_i)^2} f''(R)$$

$$+ \frac{n-1}{n}\sum_{i=1}^{n}\frac{v_i x_i}{(\sum_s x_i)^3} f''(R) - \frac{n-1}{2n}\sum_{i=1}^{n}\frac{x_i^2 \sum_s v_i}{(\sum_s x_i)^4} f''(R)$$

$$+ \frac{n-1}{6n}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R) . \qquad (9.83)$$

$$= O(n^{-2}) . \qquad (9.84)$$

Special Case I

If n is large, then (9.83) reduces to

$$\text{Bias}(\tilde{\theta}_{UWJ}) = \sum_{i=1}^{n}\frac{v_i x_i}{(\sum_s x_i)^3} f''(R) - \frac{1}{2}\sum_{i=1}^{n}\frac{x_i^2 \sum_s v_i}{(\sum_s x_i)^4} f''(R) + \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R) ,$$

$$= O(n^{-2}) . \qquad (9.85)$$

Thus from expressions (9.75) and (9.85) we gather that when n is large the bias of the unweighted jack-knife estimator is equivalent to that of the weighted jack-knifed estimator.

Corollary 9.13

If the variance function in the model of Theorem 9.5 is constant, that is, $v_i = \sigma^2$ then

$$\text{Bias}(\tilde{\theta}_{UWJ}) = \frac{1}{2}\frac{n\sigma^2}{(\sum_s x_i)^2} f''(R) - \frac{n-1}{2n}\sum_{i=1}^{n}\frac{n\sigma^2}{(\sum_s x_i)^2} f''(R) - \frac{(n-1)^2}{2n}\sum_{i=1}^{n}\frac{w_i^2\sigma^2}{(\sum_s x_i)^2} f''(R)$$

$$+ \frac{n-1}{6n}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R) \quad . \tag{9.86}$$

$$= O(n^{-2}) \tag{9.87}$$

Special Case II

If n is large, then (9.86) reduces to

$$\text{Bias}(\tilde{\theta}_{UWJ}) = -\frac{n}{2}\sum_{i=1}^{n}\frac{w_i^2\sigma^2}{(\sum_s x_i)^2} f''(R) + \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R) \quad .$$

$$= O(n^{-2}) \tag{9.88}$$

Corollary 9.14

If the variance function in the model of Theorem 9.5 is proportional to x, that is, $v_i = \sigma^2 x_i$ then

$$\text{Bias}(\tilde{\theta}_{UWJ}) = -\frac{(n-1)}{2n}\sum_{i=1}^{n}\frac{w_i x_i}{(\sum_s x_i)^2}\sigma^2 f''(R) + \frac{n-1}{6n}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R). \tag{9.89}$$

$$= O(n^{-2}) \quad . \tag{9.90}$$

Special Case III

If n is large, then (9.89) reduces to

$$\text{Bias}(\tilde{\theta}_{UWJ}) = -\frac{1}{2}\sum_{i=1}^{n}\frac{w_i x_i}{(\Sigma x_i)^2}\sigma^2 f''(R) + \frac{1}{6}\sum_{i=1}^{n}\frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R) \quad ,$$

$$= O(n^{-2}). \tag{9.91}$$

Corollary 9.15

If the variance function in the model of Theorem 9.5 is quadratic in x, that is $v_i = \sigma^2 x_i^2$ then

$$\text{Bias}(\tilde{\theta}_{UWJ}) = \frac{\sigma^2}{2} \frac{\sum_s x_i^2}{(\sum_s x_i)^2} f''(R) - \frac{(n-1)\sigma^2}{2n} \sum_{i=1}^{n} \frac{x_i^2}{(\sum_s x_i)^2} f''(R) + \frac{n-1}{n} \sigma^2 \frac{\sum_s x_i^3}{(\sum_s x_i)^3} f''(R)$$

$$- \frac{(n-1)\sigma^2}{2n} \frac{\sum_s x_i^4}{(\sum_s x_i)^4} f''(R) + \frac{n-1}{6n} \sum_{i=1}^{n} \frac{E_\xi e_i^{*3}}{(\sum_s x_i)^3} f'''(R) \quad . \quad (9.92)$$

$$= O(n^{-}) \quad . \quad (9.93)$$

Special Case IV

For large n, the expression (9.92) reduces to

$$\text{Bias}(\tilde{\theta}_{UWJ}) = \sigma^2 \frac{\sum_s x_i^3}{(\sum_s x_i)^3} f''(R) - \frac{2}{2} \sum_{i=1}^{n} \frac{\sum_s x_i^4}{(\sum_s x_i)^4} f''(R) + \frac{1}{6} \sum_{i=1}^{n} \frac{E_\xi(e_i^{*3})}{(\sum_s x_i)^3} f'''(R) \quad . \quad (9.94)$$

$$= O(n^{-2}) \quad . \quad (9.95)$$

Hinkley (1977b), while discussing the need for an appropriate jack-knife procedure for unbalanced statistics, remarks that the principle question with a given jack-knife procedure is that whether it removes a first order bias. We note from Theorem 9.4 and 9.5 that, in general, under the model ξ the bias of weighted jack-knife is $O(n^{-2})$ and that of the unweighted jack-knife is also $O(n^{-2})$. However, when the variance function is proportional to x the bias of weighted jack-knife is of $O(n^{-2})$ and that of the unweighted jack-knife of $O(n^{-2})$. If we assume that n is large, then the unweighted jack-knife also removes the first order bias term. These results indicate that the performance of the weighted jack-knife, for ratio estimation, is better than the unweighted jack-knife method. We also note that the unweighted jack-knife is sample

size dependent whereas the weighted jack-knife does not depend on the sample size. Thus we gather from these results that small sample performance of the weighted jack-knife method for ratio estimation may be better than the unweighted jack-knife. This,however, requires numerical comparison.

9.6 Summary and Discussion of Results

In this chapter we have extended the application of the weighted jack-knife technique using the empirical influence function of R at the point (x_i, y_i). This procedure, which is based on the weighted pseudovalues, produces the weighted jack-knifed ratio estimator $\tilde{R}_{WJ}$ and the weighted jack-knifed variance estimator $\tilde{V}_{WJ}$. Both these estimators possess nice properties and reproduce the original ratio estimator $\hat{R}$ and the variance estimator $\hat{V}_{2c}$ (Cochran, 1977, p.155). Under the model ξ the weighted jack-knifed variance estimator is approximately unbiased for any variance function provided n is large and f is small. This estimator is robust against the non-homogneity of the error variances. However, when the regression function $E(y_i) = \beta x_i$ in the model ξ changes the weighted jack-knifed variance estimator has a positive bias. This implies that the failure of simple regression model have a conservative effect on the weighted jack-knifed variance estimator. The comparison of the weighted and the unweighted jack-knife procedures for ratio estimation, in terms of their bias reduction properties, indicates that in general under the model ξ the performance of the weighted jack-knife is better than the unweighted jack-knife. It is also observed from Theorems 9.4 and 9.5 that the weighted jack-knife does not depend on the sample size whereas the unweighted jack-knife does. This may result in a better performance for the weighted jack-knife

in the case of small samples. This, however, needs further investigation. On the basis of the results in this chapter and the available relevant literature, we conclude that the weighted jack-knifed variance estimator may be a good competitor with the estimators currently in practice.

9.7 Suggestions for Further Research

In the course of this investigation, not only have we settled many questions about the jack-knife technique, but as is the usual case in all research we have opened the door to different areas for future exploration.

(i) The asymptotic normality of the ratio estimator for the finite population total for simple random sampling is widely used. Scott and Wu (1981) discuss the asymptotic normality of the ratio estimator under simple random sampling. The model based approach has produced results most comparable to the randomisation based approach. It may be worthwhile to examine the asymptotic distribution of the weighted jack-knifed ratio estimator for the population total under the model based approach.

(ii) Royall and Cumberland (1978a, 1981a) have produced a series of empirical results on the performance of the randomisation based variance estimator $\hat{V}_{1C}$, the unweighted jack-knifed variance estimator $\hat{V}_J$ and the model based variance estimators $\hat{V}_D$ and $\hat{V}_H$. The performance of the weighted jack-knifed estimator $\hat{V}_{WJ}$ and other competing variance estimators such as $\hat{V}_J$, $\hat{V}_{1C}$ and $\hat{V}_H$ should be investigated.

(iii) Another interesting area for further research is to compare the performance of the weighted jack-knifed variance estimater $\hat{V}_{WJ}$, the unweighted jack-knifed variance estimator $\hat{V}_J$, the randomisation based variance estimator $\hat{V}_{IC}$ and the model based variance estimator $\hat{V}_H$ in giving the approximate confidence intervals.

(iv) Simple random sampling provides balance in expectation, with the probability of extreme imbalance decreasing as the sample size increases. This has lead to the common belief that with simple random sampling one needs to worry about imbalance only in very small samples and that reasonably large samples can provide protection against bias and give robust inference, See, for example, Brewer (1979), Särndal (1980), Godambe (1982), Hansen et al. (1983) and Kalton (1983). However, Royall and Herson (1973a) and Royall and Cumberland (1978b) have reported some theoretical results which negate this belief and stress the need of balance for robust inference. Further investigations to study the implications of increasing sample size on the bias, on the performance of the variance estimates $\hat{V}_{WJ}$, $\hat{V}_{J}$, $\hat{V}_{IC}$ and $\hat{V}_{H}$ and on the behaviour of the corresponding standardised errors may be useful.

APPENDIX I

Relative Biases for Simple, Partial and Multiple Correlations and Regression Coefficients

Estimate	Actual Value	Relative Biases		
		6 Strata	12 Strata	30 Strata
Corr. 1,2	0.9152	-0.0040	0.0024	-0.0004
Corr. 1,3	0.1724	-0.0524	0.0102	-0.0011
Corr. Y,1	0.1423	0.1543	0.1214	0.0256
Corr. 2,3	0.3251	-0.0236	-0.0076	0.0053
Corr. Y,2	0.1941	0.1401	0.1159	0.0350
Corr. Y,3	0.2655	0.1159	0.0922	0.0439
Corr. 4,5	-0.3165	-0.0110	-0.0086	0.0068
Corr. 4,6	0.1398	0.0282	0.0249	-0.0050
Corr. Y',4	0.3834	0.1080	0.0983	0.0312
Corr. 5,6	-0.3557	0.0016	0.0076	-0.0132
Corr. Y',5	-0.1468	0.1171	0.1140	0.0180
Corr. Y',6	0.3304	0.0804	0.0448	0.0293
All Correlations	-	0.06972	0.05399	0.01748
Part. RY1.23	-0.0193	-0.2541	0.0955	0.0937
Part. RY2.13	0.0637	0.0998	0.1549	0.0517
Part. RY3.12	0.2007	0.1329	0.0945	0.0446
Part. RY'4.56	0.3667	0.1161	0.1051	0.0396
Part. RY'5.46	0.0793	0.0418	0.0018	0.0789
Part. RY'6.45	0.3107	0.0953	0.0501	0.0433
All Partial R.S	-	0.12333	0.08365	0.05863
Multi Corr. Y(123)	0.2895	0.2028	0.1372	0.0559
Multi Corr. Y'(456)	0.4796	0.1212	0.0851	0.0375

Estimate	Actual Value	Relative Biases		
		6 Strata	12 Strata	30 Strata
All Multiple R.S	-	0.16200	0.11115	0.04670
Beta 0,Const.	2168.140	0.0207	0.0082	0.0111
Beta Y1.23	-187.956	-0.2463	0.0215	0.0984
Beta Y2.13	548.160	-0.0314	0.0066	0.0251
Beta Y3.12	3361.730	-0.0337	0.0275	0.0024
Beta 0 Const.	-4068.830	0.0232	0.0914	0.0278
Beta Y'4.56	3163.600	0.0227	0.0000	-0.0041
Beta Y'5.46	34.478	0.0092	0.0645	0.0443
Beta Y'6.45	610.484	0.0110	0.0459	0.0089
All Betas		0.04978	0.03320	0.02776

Source: Kish and Frankel (1974) Table 5.1

REFERENCES

ADAMS, J.E., GRAY, H.L. and WATKINS, T.A. (1971). An Asymptotic Characterisation of Bias Reduction by Jack-knifing. Ann. Math. Statist., 42, 1606-1612.

ANDREWS, D.F., BICKEL, P., HAMPEL, F., HUBER, P., ROGERS. W.H. and TUKEY, J.W. (1972). Robust Estimates of Location. Princeton, N.J. : Princeton.

ARVESEN, J.N. (1969). Jackknifing U. Statistic. Ann. Math Statist., 40, 2076-2100.

ARVESEN, J.N. and LAYARD, MAXWELL W.J. (1975). Asymptotically Robust Tests in Unbalanced Variance Component Models. Annals of Statistics, Vol.3, No.5, 1122-1134.

BAILAR, B.A. and DALENIUS, T. (1969). Estimating Response Variance Components of the U.S. Bureau of the Census Survey Model. Sankyā, B31, 341-360.

BASU, D. (1969). Role of Sufficiency and Likelihood in Sample Survey Theory. Sankyā, 31, 441-454.

BASU, D. (1971). An Essay on the Logical Foundarions of Survey Sampling, in "Foundations of Statistical Inference" (V.P. Godambe and D.A. Sprotts, eds.), Holt, Rinehart and Winston. Toronto: 203-242.

BEALE, E.M.L. (1962). Some Uses of Computers in Operational Research. Industrial ORGANIS, V-31, 27-28.

BEAN, J.A. (1975). Distribution and Properties of Variance Estimators for Complex Multistage Probabiltiy Samples - An Empirical Multistage Probability Samples - an Empirical Distribution. Department of H.E.W.. Publication No. (HRA) 75-1339, Rockville, MD.

BELSLEY, D.A., KUH, E., and WELSCH, R.E. (1980). Regression Diagnostics. New York : Wiley.

BINGHAM, C. (1977). Some Identities Useful in the Analysis of Residuals from Linear Regression. University of Minnesota, School of Statistics, Technical Report, No. 300.

BISSEL, A.F. and FERGUSON, R.A. (1975). The Jackknife Toy, Tool or Two-Edged Weapon? Statistician, 24, 79-100.

BOX, G.E.P. and DRAPER, N.R. (1975). Robust Designs. Biometrika, 62, 347-52.

BREWER, K.R.W. (1963). Ratio Estimation and Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process. Aust. Jour. Statist., 5. No.:3, 93-105.

BREWER, K.W.R., and HANIF, M. (1969). Sampling Without Replacement and Probability of Inclusion Proportional to Size. I Methods Using the Horvitz and Thompson Estimator. II Methods Using Special Estimators. Unpublished Manuscript.

BREWER, K.R.W. and MELLOR, R.W. (1973). The Effect of Sample Structure on Analytical Surveys. Aust. J. Statist. Vol.15, 145-152.

BRILLINGER, D. (1964). The Asymptotic Behaviour of Tukey's General Method of Setting Approximate Confidence Limits (The Jackknife) When Applied to Maximum Likelihood Estimates. Rev. Int. Statist. Inst., 32, 202-206.

BRILLINGER, D.R. (1966). The Application of the Jackknife to the Analysis of Sample Surveys. Commentary, 8, 74-80.

BRILLINGER, D.R. (1977). Approximate Estimation of the Standard Errors of Complex Statistics Based on Sample Surveys. New Zealand Statistician, Vol. 11, No.2,35-40.

CAMPBELL, C. and MEYER, M. (1978). Some Properties of T Confidence Intervals for Survey Data. Poc. Sec. Survey Res. Methods. Amer. Statist Assoc. 437-442.

CASADY, R.J. (1975). The Estimation of Variance Components Using Balance Repeated Replications. Pro. Soc. Stat. Section Amer. Stat. Assoc. 352-357.

CASSEL, C.M., SÄRNDAL, C.E., and WRETMAN, J.H. (1977). Foundations of Inference in Survey Sampling. Wiley, New York.

CHAKRABARTY, R.P. and RAO, J.N.K. (1967). The Bias and Stability of Jackknife Variance Estimator in Ratio Estimation. Proceedings Amer. Stat. Association (Social Stat. Sect.) 326-332.

CHAKRABARTY, R.P. and RAO, J.N.K. (1968). The Bias and Stability of the Jackknife Variance Estimator in Ratio Estimation. Jour. Amer. Statist. Assoc. 63, 748-49.

CHIBISOV, D.M. (1973). An Asymptotic Expansion for a Class of Estimators Containing Maximum Likelihood Estimators. Theor. Probability Appl. 18. 295-303.

COCHRAN, W.G. (1942). Sampling Theory When the Sampling Units are of Unequal Sizes. JASA, 37, 199-212.

COCHRAN, W.G. (1963). Sampling Techniques. John Wiley and Sons, New York.

COCHRAN, W.G. (1965). The Planning of Observational Studies of Human Populations. Jour. Roy. Statist. Soc. A128 234-260.

COCHRAN, W.G. (1976). Dicussion of Paper by Smith, T.M.F. Jour. Roy. Statist. Soc. A139 201.

COCHRAN, W.G. (1977). Sampling Techniques, 3rd Ed. Wiley, New York.

COOK, R.D. (1977). Detection of Influential Observations in Linear Regression. Technometrics, 19, 15-18.

COOK, R.D. (1979). Influential Observations in Linear Regression. Jour Amer. Statist. Assoc., 74, 169-174.

COOK, R.D. and WEISBERG, S. (1980). Characterisation of an Empirical Influence Function for Detecting Influential Cases in Regression. Technometrics, 22, 495-508.

COOK, R.D. and WEISBERG, S. (1982). Residuals and Influence in Regression. Chapman and Hall, New York.

COX, D.R. (1958). Some Problems Connected with Statistical Inference. Ann. Math. Stat., 29, 357-372.

COX, D.R. and HINKLEY, D.V. (1974). Theoretical Statistics. Chapman and Hall, London.

CRAMER, H. (1946). Mathematical Methods of Statistics. Princeton, N.J., Princeton University Press.

CUMBERLAND, W.G. and ROYALL, R.M. (1983). Does Simple Random Sampling Provide Adequate Balance? Unpublished Manuscript.

DAVID, I.P. and SUKHATME, B.V. (1974). On the Bias and Mean Square Error of the Ratio Estimator. Jour. Amer. Stat. Assoc. Vol. 69, 464-66.

DAVIES, R.B. and HUTTON, B. (1975). The Effects of Errors in the Independent Variables in Linear Regression. Biometrika, 62, 383-391.

DEMING, W.E. (1950). Some Theory of Sampling. New York, John Wiley and Sons., Inc.

DEMING, W.E. (1960). Sample Design in Business Research. John Wiley and Sons, New York.

DEMING, W.E. (1963). On the Correction of Mathematical Bias by Use of Replicated Designs. Metrika, 6, 37-42.

DRAPER, N.R. and SMITH, H. (1981). Applied Regression Analysis, 2nd. Edn. New York: Wiley.

DUNCAN, G.T. and LAYARD, M.W.J. (1973). A Monte Carlo Study of Asymptotically Robust Tests of Correlation Coefficients. Biometrika, 60, 3, 551-559.

DURBIN, J. (1959). A Note on the Application of Quenouille's Method of Bias Reducation to the Estimation of Ratios. Biometrika, 46, 477-80.

EFRON, B. (1979a). Bootstrap Methods: Another Look at the Jack-knife. Ann. Statist., 7, 1-26.

EFRON, B. (1979b). Computers and the Theory of Statistics: Thinking the Unthinkable. Society for Industrial and Applied Mathematics (SIAM) Review, Vol: 21, No.4, p.460-480.

EFRON, B. (1981a). Censored Data and the Bootstrap. JASA, Vol. 76, No.374, p.312-319.

EFRON, B. (1981b). Non-parametric Estimates of Standard Error. The Jack-knife, the Bootstrap and Other Methods. Biometrika, 68, 3, pp.589-99.

EFRON, B. (1981c). Non-Parametric Standard Errors and Confidence Intervals. Canadian Jour. of Stat., 9, 139-172.

EFRON, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, Monograph No.38, CBMS-NSF.

EFRON, B. and GONG, G. (1983). A Leisurely Look at the Bootstrap, the Jackknife and Cross-Validation. The American Statistician, Vol. 37, No.1, 36-48.

EFRON, B. and STEIN, C. (1981). The Jackknife Estimate of Variance. The Annals of Statist., Vol.9, No.3, 586-596.

ERICSON, W.A. (1969). Subjective Bayesian Models in Sampling Finite Populations, J.R. Statist. Soc. B. 31, 195-234.

FAREWELL, V.T. (1978). Jackknife Estimation With Structured Data. Biometrika, 65, 444-447.

FENECH, P.A. (1983). A Pseudo-Replicate Estimator of the Variance of Function of Sample Means. Commun. Statist. Theory and Methods., 12(7), 777-798.

FISHER, R.A. (1935). The Logic of Inductive Inference (With Discussion). J.R. Statist. Soc., 98, 39-82.

FISHER, R.A. (1955). Statistical Methods and Statistical Inference. J.R. Statist. Soc., Ser.B, Vol.17, 69-78.

FOLSOM, R.E. Jr. (1974). National Assessment Approach to Sampling Error Estimation, Sampling Error Monograph. Research Triangle Park, North Carolina: Research Triangle Institute. Project No. 25U-796-5, Prepared for National Assessment of Educational Progress.

FOLSOM, R.E. BAYLESS, D.L. and SHAH B.V.(1971). Jackknifing for Variance Components in Complex Sample Survey Designs. Proceedings Americal Statistical Association (Social Statistics) Section). 36-39.

FRANKEL, M.R. (1971). Inference From Clustered Sample. Proceedings American Statistical Association (Social Statistics Section). 20-27.

FREEMAN, D.H., FREEMAN, J.L., BROCK, D.B., and KOCH, G.G. (1976). Strategies in the Multivariate Analysis of Data from Complex Surveys II: An Application to the United States National Health Interview Survey. Inst. Stat. Rev., 44, 317-30.

FRYER, J. (1970). Jack-knifing Maximum Likelihood Estimates in the Multiparameter Case (Preliminary Report, Abstract). Annal. Math. Statist., 41, 1392.

FULLER, W.A. (1973). Regression Analysis for Sample Surveys. Int. Assoc. of Survey Statisticians. 1st Meeting, Vienna, August, 1973.

FULLER, W.A. (1974). Regression Analysis for Sample Surveys. Report Prepared for the US Bureau of the Census on Work Conducted Under the Joint Statistical Agreement, Iowa State Univ., Ames., Iowa.

FULLER, W. A. (1975). Regression Analysis for Sample Survey. Sankhyā. Series, C, Pt.3, 37, 117-132.

GHANGURDE, P.D. (1981). Models for Estimation of Sampling Errors. Proc. Amer. Statist. Assoc. (Sur. Res. Meth. Sect.). 209.

GODAMBE, V.P. (1955). A Unified Theory of Sampling from Finite Populations. J.R. Statist Soc. B, 17, 269-278.

GODAMBE, V.P. (1966). A New Approach to Sampling from Finite Populations. J.R. Statist. Soc. B., 28, 310-328.

GODAMBE, V.P. (1969). A Fiducial Argument With Application to Survey Sampling. J.R. Statist. Soc., B31, 246-260.

GODAMBE, V.P. (1982). Estimation in Survey Sampling. Robustness and Optimality. Journal of the American Statistical Association, 77, 393-403.

GODAMBE, V.P. and JOSHI, V.M. (1965). Admissibility and Bayes Estimation in Sampling Finite Populations 1. Ann. Math. Statist., 36, 1707-1722.

GODAMBE, V.P. and THOMPSON, M.E. (1976). Philosophy of Survey Sampling Practice, in Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science (Harper and Hooker, eds). Dordecht - Holland: D. Reidel Publishing Co.

GOODMAN, L.A. and HARTLEY, H.O. (1958). The Precision of Unbiased Ratio Type Estimators. J. Amer. Stat. Assoc., 53, 491-508. Corrigenda: JASA, 58, 1162.

GRAY, H.L. and SCHUCANY, W.R. (1972). The Generalised Jackknife Statistic, New York: Marcel Dekker.

GURNEY, M. and JEWETT, R.S. (1975). Constructing Orthogonal Replications for Variance Estimation. JASA, V.70, N.352, 819-821.

HÁJEK, J. (1958). On the Theory of Ratio Estimates. Aplik. Mat. 3, 384-398. International Journal of Abstracts (Theory and Methods), I, 4.3 (8.1), 50.

HÁJEK, J. (1960). On the Theory of Ratio Estimate. Bull Int. Stat. Inst., 37(2), 219-226.

HAMPEL, F. (1974). The Influence Curve and its Role in Robust Estimation. Amer. Statist. Assoc., 69, 383-393.

HANSEN, M.H. and HURWITZ, W.N. (1943). On the Theory of Sampling From Finite Populations. Ann. Math. Statist., 14, 333-362.

HANSEN, M.H. and HURWITZ, W.N. and MADOW, W.G. (1953 a,b). Sampling Survey Methods and Theory, Vols 1,2. Wiley, New York.

HANSEN, M.H., MADOW, W.G. and TEPPING, B.J. (1983). An Evaluation of Model-dependent and Probability Sampling Inferences in Sample Surveys. Jour. Amer. Statist. Assoc., 78, 776-793.

HARTLEY, H.O. and RAO, J.N.K. (1968). A New Estimation Theory of Sample Surveys. Biometrika, 55, 547-557.

HARTLEY, H.O. and ROSS, A. (1954). Unbiased Ratio Estimate. Nature, 174, 270-271.

HARTLEY, H.O. and SIELKEN, R.L. (1975). A 'Superpopulation Viewpoint' for Finite Population Sampling. Biometrika, 31, 411-422.

HENDERSON, H.V. and SEARLE, S.R. (1981). On Deriving the Inverse of a Sum of Matrices. SIAM Review, 23, 53-60.

HERSON, J. (1976). An Investigation of Relative Efficiency of Least Squares Prediction to Conventional Probability Sampling Plans. JASA, 71, 701-703.

HIDIROGLOU, M.A., FULLER, W.A. and HICKMAN, R.D. (1979). SUPER CARP, 5th ed., Survey Section, Statistical Laboratory, IOWA State University.

HINKLEY, D.V. (1977a). Jack-knife Confidence Limits Using Student-t Approximations. Biometrika, 64, 21-28 and Correction, 65, 677.

HINKLEY, D.V. (1977b). Jack-knifing in Unbalanced Situations. Technometrics, 19, 285-292.

HINKLEY, D.V. (1978). Improving the Jack-knife With Special Reference to Correlation Estimation. Biometrika, 65, 13-21.

HOAGLIN, D.C. and WELSCH, R. (1978). The Hat Matrix in Regression and ANOVA. Amer. Statistician, 32, 17-22.

HOBBS, C., CHILKO, D., BILLINGS, A., KING, R., and HSHIEH, P. (1980). Jackknife Techniques in Regression Analysis. Annual Conf. SAS. Users Group. Internat. SAN. ANTONIO, 203-207.

HOLT, M.M. (1977). SURRGER: Standard Errors of Regression Coefficients from Sample Survey Data (Manual), Research Triangle Institute.

HOLT, D., SMITH, T.M.F. and WINTER, P.D., (1980). Regression Analysis of Data from Complex Surveys. J.R. Statist. Soc, A; 143, 474-487.

HORVITZ, D.G., and THOMPSON, D.J. (1952). A Generalization of Sampling Without Replacement from a Finite Universe. Jour. Amer. Statist. Assoc., 47, 663-685.

HUBER, P.J. (1972). Robust Statistics - A Review. Annals. Math. Stat. Vol.43, 1011-1067.

HUBER, P. (1975). Robustness and Designs, Srivastava, J.N. (ed). A Survey of Statitical Design and Linear Models. Amsterdam, North Holland.

HUBER, P. (1981). Rubust-Statistics, New York, Wiley.

JOHNSON, N.J. and KOTZ, S. (1970). Continuous Univariate Distributions, Vol.2, Houghton Millin, Boston.

JONES, H.L. (1962). Some Basic Problems in Replicated Sampling, New Frontiers in Administrative and Engineering Quality Control: Proceedings of the 1962 Middle Atlantic Conference of the American Society for Quality Control, American Society for Quality Control, Milwaukee, 17-81.

JONES, H.L. (1963). The Jack-knife Method. Proc. IBM Scientific Computing Symposium on Statistics, IBM. 185-201.

JONES, H.L. (1974). Jackknife Estimation of Function of Stratum Means. Biometrika, 61, 343-348.

JONRÜP, H. and RENNERMALM, B. (1976). Regression Analysis in Samples From Finite Populations. Scand. J. Statist. 3, 33-37.

KALBFLEISCH, J.D. and SPROTT, D.A. (1969). Application of Likelihood And Fiducial Probability to Sampling Finite Populations. In "New Developments in Survey Sampling". (Johnson, N.L. and Smith, H. ed.) Wiley, New York, 358-389.

KALTON, G. (1974). Discussion of Kish and Frankel (1974) Paper.

KALTON, G. (1977). Practical Methods for Estimating Survey Sampling Errors. Bull. Int. Statist. Inst. 47(3), 495-514.

KALTON, G. (1983). Models in Practice of Survey Sampling. Inst. Statist. Rev., 51, 175-188.

KENDALL, M.G. and STUART, A. (1973). The Advanced Theory of Statistics, Vol.2, Griffin, London.

KEYFITZ, N (1957). Estimates of Sampling Variance When Two Units are Selected From Each Stratum. Jour. Amer. Stat. Assoc., 52, 503-510.

KISH, L. (1965). Survey Sampling. John Wiley and Sons, New York.

KISH, L. and FRANKEL, M. (1968). Balanced Repeated Replication for Analytical Statistics. Proc. Amer. Stat. Assoc. (Soc. Stat. Scet). 2-10.

KISH and FRANKEL (1970). Balanced Repeated Replications for Standard Errors. JASA, Vol.65, No.331 (Application Section). 1071-1093.

KISH, L., FRANKEL, M., and Van ECK, N. (1972). SEPP - Sampling Error Programme Package. (PSALMS, BRRP, ABSERD). Institute for Social Research, University of Michigan, Ann Arbor, Michigan.

KISH, L. and FRANKEL, M.R. (1974). Inference from Complex Samples. (With Discussion). J.R. Statist. Soc., B., 36, 1-37.

KISH, L. and HESS, I. (1959b). On Variances of Ratios and Their Differences in Multistage Samples. Jour Amer. Stat. Assoc., 54, 416-46. Corrigenda, Jour. Am. Stat. Assoc. 58, (1963), 1162. (Sastry, K.V.R.).

KNOTT, M. and FRANGOS, C.C. (1983). Variance Estimation for the Jackknife Using Von Mises Expansions. Biometrika, 70, 2, 501-504.

KOCH, G.G. and LEMESHOW, S. (1972). An Application of Multivariate Analysis to Complex Sample Survey Data. JASA, 67, 780-2.

KOKAN, A.R. (1963). A Note on the Stability of the Estimates of Standard Errors of the Ordinary Mean Estimate and the Ratio Estimate. Calcutta. Statist. Assoc. Bull. 12, 149-58.

KONIJN, H.S. (1973). Statistical Theory of Sample Survey Design and Analysis. North-Holland Publishing Co., Amsterdam.

KONIJN, H.S. (1974). Discussion of Paper by Kish and Frankel, "Inference From Complex Samples". J.R. Statist. Soc., B36, 1-37

KOOP, J.C. (1960). On Theoretical Questions Underlying the Techniques of Replicated or Interpenetrating Sampling. Proc. Social Stat. Section. ASA, 196-205.

KOOP, J.C. (1962). On Upper Limits to the Difference in Bias Between Two Ratio Estimators. Metrika, 5, 145-149.

KOOP, J.C. (1967). Replicated (or interpreting) Samples of Unequal Sizes. Ann. Math. Statist., 38, 1142-1147.

KOOP, J.C. (1970). Estimation of Correlation for a Finite Universe. Metrika, 15, pp.105-109.

KREWSKI, D., (1978a), On the Stability of Some Replication Variance Estimators in the Linear Case. Jour. Statist. Planning and Inference 2, 45-51.

KREWSKI, D. (1978b). Jackknifing U Statistics In Finite Population. Commun. Statist., A7, 1-12.

KREWSKI, D., and CHAKARBARTY, R.P. (1981). On the Stability of the Jackknife Variance Estimator in Ratio Estimation. Jour. Stat. Planning and Inference., 5, 71-78, North Holland Publishing Co.

KREWSKI, D. and RAO, J.N.K. (1978). Inference from Stratified Samples - I Large Sample Properties of Linearization, Jack-knife and Balanced Repeated Replication Methods. Carleton Mathematical Series No. 155, Department of Mathematics, Carleton Unversity, Ottawa, Canada. 1-32.

KREWSKI, D. and RAO, J.N.K. (1981). Inference from Stratified Samples: Properties of the Linearisation, Jack-knife and Balanced Repeated Replication Methods. The Annals of Statistics, Vol.9, 1010-1019.

LAHIRI, D.B. (1951). A Method of Sample Selection Providing Unbiased Ratio Estimates. Bull. Int. Statist. Inst. 33:11, 133-140.

LAHIRI, D.B. (1958). Observations on the Use of Interpreting Samples in India. Bull. Int. Stat. Inst., Vol.36, 144-152.

LAHIRI, D.B. (1968). On the Unique Sample, The Surveyed One. Unpublished Paper Presented at the Symposium on the Foundations of Survey Sampling, Chapel Hill, North Carolina.

LANKE, J. (1974). On Non Negative Variance Estimators in Survey Sampling. Sankhya C.36, 33-42.

LANKE, J. (1973). On UMV - Estimators in Survey Sampling. Metrika, 20, 196-202.

LANKE, J. (1975). Some Contributions to the Theory of Survey Sampling. Lund: AV-Centralen.

LAUH, E. and WILLIAMS, W.H. (1963). Some Small Sample Results for the Variance of a Ratio. Proc. Amer. Stat. Assoc. (Social Statistics Section), 273-83.

LEE, K.H., (1972). Partially Balanced Designs for Half Sample Replication Method of Variance Estimation. JASA, V.67, No.338, Applications Section, 324-334.

LEE, K.H., (1973). Variance Estimation in Stratified Sampling. JASA, V.68, 342. Application Section, 336-342.

LEMESHOW, S. and EPP, R. (1977). Properties of the Balanced Half-sample and Jack-knife Variance Estimation Techniques in the Linear Case. Commun. Statist., A6, 1259-1274.

LEMESHOW, S., and LEVY, P.S. (1979). Estimating the Variances of Ratio Estimate in Complex Sample Surveys with Two Primary Sampling Units Per Stratum - a Comparison of Balanced Replication and Jack-knife Techniques. J. Statist. Comput. Simul., 8, 191-205.

LITTLE, R.J.A., (1982). Models for Non-Response in Sample Surveys. Jour. Amer. Stat. Assoc, 77, 237-250.

LITTLE, R.J.A. (1983). Discussion of Hansen et al., Paper "An Evaluation of Model Dependent and Probability Sampling Inferences in Sample Surveys". JASA, 78, 797-799.

MADOW, W.G. and MADOW, L.H., (1944). On the Theory of Systematic Sampling. Ann. Math. Stat., 15, 1-24.

MAHALANOBIS, P.C., (1938-41). A Statistical Report on the Experimental Crop Census of 1937; Subsequently Printed in the Series, On the Sample Survey of Jute in Bengal, Indian Central Jute Committee, Calcutta, 1938-41.

MAHALANOBIS, P.C. (1946). Recent Experiments in Statistical Sampling in India. JRSS (A). A109, 325-378.

MAJUMDAR, H., and SEN, P.K., (1978). Invariance Principles for Jackknifing U-Statistics for Finite Population Sampling and Some Applications. Commun. Statist. Theor. Meth., A7, 1000-1025.

MALLOWS, C.L. (1975). On Some Topics in Robustness. Unpublished Bell Telephone Laboratories Report, Murray Hill, N.J.

MANN, H.B., and WALD, A., (1943). On Stochastic Limit and Order Relationship. Ann. Math. Statist., 14, 217-226.

MANTEL, N., (1967). Assumption Free Estimators Using U-Statistic and a Relationship to the Jackknife Method. Biometrics, 23, 567-571.

MAURER, K., JONES, G. and BRYANT, E., (1978). Relative Computational Efficiency of the Linearised and Balanced Repeated Replication Procedures for Computing Sampling Variances. Proc. Amer. Statist. Assoc., (Sur. Res. Meth) 388-389.

McCARTHY, P.J. (1966). Replication - An Approach to the Analysis of Data from Complex Surveys. Public Health Service Publication No.1000 - Series 2, N.14. U.S. Department of Health Education and Welfare. Washington. 1-37.

McCARTHY, P.J. (1969a). Pseudo-replications: Half Sampel, Rev. Inst. Statist. Int., 37, 239-264.

McCARTHY, P.J., (1969b). Pseudo-Replication. Further Evaluation and Application of the Balanced Half-Sample Technique. Vital and Health Statistics, Series 2, No.31 U.S. Department of Health, Education and Welfare. Washington U.S. Government Printing Office.

MIDZUNO, H. (1950). An Outline of the Theory of Sampling Systems; AISM, 1, 149-156.

MILLER, R.G., Jr., (1964). A Trustworthy Jackknife. Ann. Math. Statist., 35, 1594-1605.

MILLER, R.G., Jr., (1968). Jackknifing Variances. Ann. Math. Statist., 39, 567-82.

MILLER, R.G., Jr., (1974a). An Unbalanced Jack-knife. Ann. Statist., 2, 880-891.

MILLER, R.G., Jr., (1974b). The Jack-knife, a Review. Biometrika, 61, 1-15.

MICKEY, M.R. (1959). Some Finite Population Unbiased Ratio and Regression Estimators. J. Amer. Stat. Assoc., 54, 594-612.

MOKASHI, V.K. (1950). A Note on Interpreting Samples. J. Ind. Soc. Agri. Stat., 2, 189.

MURTHY, M.N., (1962). Almost Unbiased Estimators Based on Interpenetrating Sub-samples. Sankhya, A, 303-314.

MURTHY, M.N. (1967). Sampling Theory and Methods. Statistical Publishing Co., India.

NANDI, H.K., and SEN, P.K., (1963). On the Properties of U-Statistics When the Observations are Not Independent. Pt.II. Unbiased Estimation of the Parameters of a Finite Population. Cal. Statist. Assoc. Bull., 12, 125-143.

NATHAN, G. (1973). Utilisation of Information on Sampling and Non-Sampling Errors for Survey Design - Experiences in Israel. Bull. Int. Statist., 45(3), 393-409.

NATHAN, G., (1975). Tests of Independence in Contingency Tables from Stratified Proportional Samples. Sankya, C.37, 77-87.

NEYMAN, J., (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Sampling. J. Royal. Soc., 97, 558-625.

NEYMAN, J., and PEARSON, E.S., (1937). Note on Some Points in "Student", Paper on "Comparison Between Balanced and Random Arrangement of Field Plots. Biometrika, 29, 380-388.

O'MUIRCHEARTAIGH, C., VERMA, V., and SCOTT, C., (1980). Sample Designs and Sampling Errors for World Fertility Survey. Jour. Roy. Statist., Soc. A143, Part 4, 431-473.

O'MUIRCHEARTAIGH, C., and WONG, S.T., (1981). The Impact of Sampling Theory on Survey Sampling Practice: A Review. Bull. Int. Statist., Inst., 49, 465-493.

PARR, W.C. (1983). A Note on the Jackknife, the Bootstrap and the Delta Method Estimators of Bias and Variance. Biometrika, 70, 719-22.

PASCUAL, J.N., (1961). Unbiased Ratio Estimators in Stratified Sampling. J. Amer. Stat. Assoc., 56, 70-87.

PLACKET, R.L. and BURMAN, J.P. (1946) The Design of Optimum Multifactorial Experiments. Biometrika, 33, 305-325.

QUENOUILLE, M., (1949). Approximate Tests of Correlation in Time Series. J.R. Statist. Soc., Ser B; 11, 68-84.

QUENOUILLE, M., (1956). Notes on Bias in Estimation. Biometrika, 43, 353-360.

RAJ, D., (1956). Some Estimators in Sampling with Varying Probabilities Without Replacement. Jour. Amer. Stat. Assoc., 51, 269-284.

RAJ, D., (1968). Sampling Theory, McGraw-Hill, New York.

RAO, C.R. (1965). Linear Statistical Inference and its Applications, 1st Ed., Wiley, New York.

RAO, C.R., (1973). Linear Statistical Inference and its Applications, 2nd Edn., New York: Wiley.

RAO, J.N.K., (1965). A Note on the Estimation of Ratios by Quenouille's Method. Biometrika, 52, 647-649.

RAO, J.N.K. (1967). The Precision of Mickey's Unbiased Rato Estimator. Biometrika, 54, 321-324.

RAO, J.N.K., (1969a). Ratio and Regression Estimators. In New Development in Survey Sampling, Eds., (N.L. Johnson and Harry Smith, Jr.), 213-34, New York, Wiley.

RAO, J.N.K., (1969b). Small Sample Results in Ratio and Regression Estimation. Tech. Report. No. Math. Stat/15/69. Indian Statistical Institute, Calcutta.

RAO, J.N.K., (1975). Analytical Studies of Sample Survey Data. Survey Methodology, 1, Supplementary Issue 1-76.

RAO, J.N.K., and BEEGLE, L.D., (1966). A Monte Carlo Study of Some Ratio Estimators. Pro. Amer. Stat. Assoc. (Social Stat. Section), 443-450. Corrigendum: Sankhya, B29, 190.

RAO, J.N.K., and KUZIK, R.A., (1974). Sampling Errors in Ratio Estimation. Sankhya, Series C, 36, 43-58.

RAO, J.N.K., and SINGH, M.P., (1973). On the Choice of Estimator in Survey Sampling. Australian Jour. Stat., 15, 95-104.

RAO, J.N.K., and WEBSTER, J.T., (1966). On Two Methods of Bias Reduction in the Estimation of Ratios. Biometrika, 53, 571-577.

RAO, P.S.R.S., (1969). Comparison of Four Ratio-Type Estimates Under a Model. Jour. Amer. Stat. Assoc, 64, 574-580.

RAO, P.S.R.S. (1979). On Applying the Jackknife Procedure to the Ratio Estimator. Sankhya: The Indian Journal of Statistics, Volume 41, Series C, Pt.2, 115-126.

RAO, P.S.R.S. (1981). Estimation of the Mean Square Error of the Ratio Estimator, In Current Topics in Survey Sampling (Krewki, Placket and Rao; Ed.). Academic Press. 305-315.

RAO, PODURI, S.R.S. and RAO J.N.K. (1971a). Small Sample Results for Ratio Estimators. Biometrika, 58, 625-30.

RAO, PODURI, S.R.S. and RAO, J.N.K., (1971b). On the Jack-knife Estimator in Finite Populations. Bull. Int. Stat. Inst., 44(2), 325-328.

RAO, T.J., (1977). Estimating the Variance of the Ratio Estimator for the Midzuno-Sen Sampling Scheme. Metrika, 24, 203-208.

REEDS, J.A. (1976). On the Definition of von Mises Functionals. Unpublished Ph.D. Dissertation, Harvard University, Department of Statistics.

REEDS, J.A., (1978). Jackknifing Minimum Likelihood Estimates. Ann. Statistics, 6, 727-739.

RICHARD, V., and FREEMAN, D.H., (1979). A Comparison of Balanced Repeated Replication to Direct Replication for Variance Etimation. Proc. Amer. Stat. Assoc., (Surv. Res. Meth.), 106-108.

ROBSON, D.S., (1957). Application of Multivariate Polykays to the Theory of Unbiased Ratio-Type Estimation. Jour. Amer. Stat. Assoc., 52, 511-522.

ROCKE, D.M., and DOWNS, G.W., (1981). Estimating the Variance of Robust Estimators of Location: Influence Cause, Jackknife and Bootstrap. Commun. Statist. Simula. Computa., Bio, (3), 221-248.

ROYALL, R.M., (1968). An Old Approach to Finite Population Sampling Theory. Jour. Amer. Stat. Assoc., 63, 1269-1279.

ROYALL, R.M. (1970). On Finite Population Sampling Theory Under Certain Linear Regression Models. Biometrika, 57, 377-388.

ROYALL, R.M. (1971). Linear Regression Models in Finite Population Sampling Theory. In Foundations of Statistical Inference. (V.p. Godambe and D.A. Sprott, eds.). 259-278. Holt, Reinhart and Winston of Canada Ltd., Toronto, Montreal.

ROYALL, R.M., (1976). Current Advances in Sampling Theory: Implications for Human Observational Studies. Amer. Jour. Epidemiology. Vol.4, No.4, 463-474.

ROYALL, R.M., (1982). Robust Variance Estimation in Two-Stage Cluster Sampling Using Prediction Models. Unpublished Manuscript.

ROYALL, R.M. and CUMBERLAND, W.G. (1978a). An Empirical Study of Prediction Theory in Finite Population Sampling. Simple Random Sampling and the Ratio Estimator, in Survey Sampling and Measurement, ed. N.K. Manboodiri, New York. Academic Press, 293-399.

ROYALL, R.M. and CUMBERLAND, W.G. (1978b). Variance Estimation in Finite Population Sampling. J. Amer. Statist., Assoc., 73, 351-358.

ROYALL, R.M. and CUMBERLAND, W.G. (1981a). An Empirical Study of the Ratio Estimator and Estimators of its Variance. JASA, Vol.76. No.373. p.66-77.

ROYALL, R.M., and CUMBERLAND, W.G., (1981b). The Finite-Population Linear Regression Estimator and Estimators of its Variance - An Empirical Study. Jour. Amer. Stat. Assoc., 76, 924-930.

ROYALL, R.M., and CUMBERLAND, W.G., (1982). Conditional Coverage Properties of Finite Population Confidence Intervals. Paper No.535, Department of Statistics, John Hopkins University.

ROYALL, R.M. and EBERHARDT, K.R. (1975). Variance Estimates for the Ratio Estimator. Sankhya, Series C, Pt. 1, 37, 43-52.

ROYALL, R.M. and HERSON, J. (1973a). Robust Estimation in Finite Populations I. JASA, 68, No.344, 880-889.

ROYALL, R.M. and HERSON, J. (1973b). Robust Estimation in Finite Populations I. JASA, 68, No.344, 890-893.

RUBIN, D.B., (1976). Inference and Missing Data. Biometrika, 63, 581-592.

SAMPFORD, M.R., (1974). Discussion of the Paper by Kish and Frankel. J.R. Statist. Soc., Ser.B, 36, 1-37.

SARNDAL, C.E. (1978). Design-based and Model Based Inference in Survey Sampling. Scand Jour. Statistics, 5, 27-52.

SARNDAL, C.E., (1980). Two Model Based Inference Arguments in Survey Sampling. Austral. J. Statist., 22(3), 341-348.

SARNDAL, C.E., (1981). Frameworkf for Inference in Survey Sampling With Applications to Small Area Estimation and Adjustment for Non-response. Bull. of Int. Statist. Inst., XLIX, Book 1, 494-513.

SCHUCANY, W.R., GRAY, H.L., and OWEN, D.B., (1971). On Bias Reduction in Estimation. J. Amer. Statist. Assoc., 66, 524-533.

SCOTT, A.J., and SMITH, T.M.F., (1969). Estimation in Multistage Surveys. Jour. Amer. Stat. Assoc., 64, 830-840.

SEN, A.R. (1953). On the Estimate of Variance in Sampling with Varying Probabilities. Jour. Ind. Soc. Agri. Stat., 5, 119-127.

SHAH, B.V., HOLT, M.M., and FOLSOM, R.E., (1977). Inference About Regression Models From Sample Survey Data. Bull. Int. Statist. Inst. Book 3, 43-57.

SHAROT, T., (1976a). Sharpening the Jackknife. Biometrika, 63, 315-321.

SHAROT, T., (1976b). The Generalised Jackknife: Finite Sample and Sub-Sample Sizes. J. Amer. Statist. Assoc., 71, 451-454.

SIMMONS, W.R., and BAIRD, J.T., Jr., (1968). Pseudo-Replication in the NCHS Health Examination Survey. Proc. Amer. Stat. Assoc. (Soc. Stat. Sect.), 19-32.

SMITH, T.M.F., (1974). Discussion of Kish and Frankel's (1974) Paper. "Inference From Complex Samples". Jour Roy. Soc. Series B, 36, 1-37.

SMITH, T.M.F., (1976). The Foundations of Survey Sampling. J.R. Statist. Soc. A 139, 183-195.

SMTIH, T.M.F., (1979). Statistical Sampling in Auditing: A Statistician's Viewpoint. The Statistician, 28, 4, 267-280.

SMITH, T.M.F., (1981a). Regression Analysis for Complex Surveys, In "Current Topics in Survey Sampling". (Krewski, Platek and Rao, Ed.), Academic Press, London. 267-292.

SMITH, T.M.F., (1981b). "Comments" on Royall and Cumberland., (1981a) Paper. Jour. Amer. Stat. Assoc. 76, 83.

SMITH, T.M.F., (1983a). On the Validity of Inferences From Non-Random Samples. J.R. Statist. Soc., A 146, 394-403.

SMITH, T.M.F., (1983b). Disucssion of Hansen et al's Paper "An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Survey". J. Amer. Statist. Assoc., Vol. 78, 776-793.

SMITH, T.M.F., (1984). Present Position and Potential Developments: Some Personal Views (Sample Survey). Jour. Roy. Statist. Soc, A 147, 8.1-8.12.

SRIKANTAN, K.S. (1961). Testing for a Single Outlier in a Regression Model. Sankhya, A, 23, 251-260.

SRIKANTAN, K.S., (1963). A Note on Interpenetrating Sub-Samples of Unequal Sizes. Sankhya, 25 (B), 345-350.

STANEK III, E.J., and LEMESHOW, S., (1977). The Behaviour of Balanced Half Sample Variance Estimates for Linear and Combined Ratio Estimates When Strata are Paired to Form Pseudo-Strata. Proc. Amer. Stat. Assoc., (Soc Stat. Sect), 837-840.

STUART, A., (1976). Basic Ideas of Scientific Sampling. Second Edition. Charles Griffin and Company Ltd., London.

SUGDEN, R. (1979). Inference on Symmetric Functions of Exchangeable Populations. J.R. Statist., Ser. B, 41, 269-273.

SUKHATME, P.V., and SUKHATME, B.V., Sampling Theory of Surveys With Applications. Iowa State University Press, USA.

SUNDBERG, R., (1983). The Predictive Approach and Randomised Population Models for Finite Populations Inference From Two Stage Samples. Scand. J. Statist., 10, 223-238.

TEPPING, B.J. (1968). Variance Estimation in Complex Surveys. Proc. Amer. Stat. Assoc. (Soc. Stat. Sect.), 11-18.

THORBURN, D., (1976). Some Asymptotic Properties of Jack-knife Statistics. Biometrika, 63, 305-313.

THORBURN, D., (1977). On the Asymptotic Normality of the Jackknife. Scand. J. Statist., 4, 113-118.

TIN, M., (1965). Comparison of Some Ratio Estimators. J. Amer. Stat. Assoc., 60, 294-307. Corrigenda. Jour. Amer. Stat. Assoc., 65, 1012.

TUKEY, J.W. (1949). Memorandum Statistics in the Federal Government, American Statistician, Vol.3, 6-17.

TUKEY, J.W., (1954). Unsolved Problems of Experimental Statistics, JASA, 49, 706-731.

TUKEY, J.W. (1958). Bias and Confidence in Not Quite Large Samples. Ann. Math. Stat. (Abstract), 29, 614.

VIJAYAN, K., (1975). On Estimating the Variance in Unequal Probability Sampling. J. Amer. Statist. Assoc., 70, 713-716.

von MISES, R. (1947). On the Asymptotic Distribution of Differential Statistical Functions. Ann. Math. Statist., 18, 309-48.

von MISES, R. (1964). Mathematical Theory of Probability and Statistics. New York: Academic Press.

WAINER, H., and THISSEN, D., (1975). When the Jackknife Fails (Or Does it?). Psychometrika., 40, 113-114.

WAKIMOTO, K., (1971a). Stratified Random Sampling (I), Estimation of the Population Variance. Ann. Inst. Statist. Math., 23, 233-252.

WAKIMOTO, K., (1971b). Stratified Random Sampling (II). Estimation of the Population Covariance. Ann. Inst. Statist. Math., 23, 327-337.

WAKIMOTO, K., (1971c). Stratified Random Sampling (III) Estimation of the Population Correlation Coefficient. Ann. Inst. Statist. Math., 23, 339-353.

WARREN, W.G., (1971). Correlation or Regression, Bias or Precision Applied Stat., 20, 148-164.

WEISBERG, S. (1980). Applied Linear Regression. New York: Wiley.

WELSCH, R. E., and KUH, E. (1977). Linear Regression Diagnostics, Working Paper No. 173, National Bureau of Economic Research.

WILKS, S. (1963). Multivariate Statistical Outliers. Sankhyā A, 25, 507-526.

WOODRUFF, R.S., (1971). A Simple Method for Approximating the Variance of a Complicated Estimate. Jour. Amer. Stat. Assoc., 71, 315-321.

WOODRUFF, R.S. and CAUSEY, B.D. (1976). A Computerised Method for Approximating the Variance of a Complicated Estimate. J. Amer. Statist. Assoc., 66, 411-414.

WU, C.F., (1982). Estimation of Variance of the Ratio Estimator. Biometrika, 69, 183-199.

YATES, F., (1949). Sampling Methods for Censuses and Surveys. Charles Griffin and Co., London.

YATES, F. and GRUNDY, P.M. (1953). Selection Without Replacement From Within Strata with Probability Proportional to Size. Jour. Roy. Stat. Soc., B15, 253-261.