1

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Services des thèses canadiennes

Ottawa, Canada
K1A 0N4

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

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

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

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

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

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R S C 1970, c C-30.

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

AVIS

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

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

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

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

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c C-30.

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

SMALL AREA ESTIMATION
AND
MEASUREMENT OF RESPONSE ERROR VARIANCE IN SURVEYS

by

N.G. NARASIMHA PRASAD, M.Sc.

A thesis submitted to the faculty of Graduate
Studies for partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Department of Mathematics
and Statistics

Carleton University

Ottawa, Ontario

August, 1985

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-29873-1

The undersigned recommend to the Faculty of Graduate Studies and
Research acceptance of the thesis "SMALL AREA ESTIMATION AND MEASUREMENT
OF RESPONSE ERROR VARIANCE IN SURVEYS" submitted by N.G. Narasimha Prasad,
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

External Examiner

Thesis Supervisor

Chairman Department of Mathematics
and Statistics

Carleton University

August, 1985

ABSTRACT

Statistical methods for variance components models are utilized to predict small area means and to measure the response error variance in surveys. Chapters 2 to 5 of this thesis deal with the small area estimation, while the measurement of response error variance is studied in Chapter 6.

Three regression models with components of variance error structure proposed in the context of small area estimation are described. Under these models, with known variance components, the Best Linear Unbiased Predictors (BLUPs) of small area means and their variances are derived in Chapter 2.

In Chapter 3, jackknife estimates of the variance of estimated variance components and the variance of non-linear functions of estimated variance components are obtained. These estimates are shown to be robust by establishing their asymptotic consistency without normality assumption.

When the variance components are unknown, BLUPs of small area means are estimated by replacing the unknown variance components by their respective asymptotic consistent estimators. For the estimated BLUP an approximate expression for the mean square error and its estimator under normality assumption, are obtained in Chapter 4, for the models considered in Chapter 2. An extension to a stratified sample design is given in Chapter 5.

In Chapter 6, a variance component model for measurement of response errors in surveys is proposed. The method of minimum norm quadratic unbiased estimation is employed to obtain the optimal estimator of the response error variance.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Professor J.N.K. Rao for his unfailing continuing guidance and encouragement which he provided throughout my graduate career at Carleton University. The financial support he supplied is also gratefully acknowledged.

As well, my sincere thanks go to Dr. J. Kleffe for his guidance and comments during my research into the MINQUE aspects of this research.

I am very grateful to Professor J.E. Graham and Dr. R. Thomas for their continuing help in preparing this thesis.

I would also like to acknowledge Gill Murray for her excellent job in translating the manuscript into typed form and for making grammatical corrections where needed.

I am thankful to the entire Department of Mathematics and Statistics for making my stay at Carleton University a pleasant one. I will always remember the stimulating discussions I had with my friends, both staff and students in the department.

Finally, I express special thanks to my wife Vijaya, my daughters Laxmi and Deepa, who patiently tolerated my long hours of study away from home.

TABLE OF CONTENTS

Abstract	i
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
 CHAPTER 1. INTRODUCTION	 1
1.1 An Overview of Small Area Methodology	1
1.2 Response Errors: an Overview	11
1.3 Thesis Outline	13
 CHAPTER 2. BEST LINEAR UNBIASED PREDICTOR	 15
2.1 Description of Small Area Models	15
2.2 Prediction under a General Mixed Linear Model (Known Variance Components)	20
2.3 BLUPs under the Small Area Models	22
 CHAPTER 3. ESTIMATION OF VARIANCE COMPONENTS	 27
3.1 The Method of Fitting of Constants	28
3.2 The Jackknife Method: Introduction	31
3.3 A Jackknife Procedure for Small Area Models	33
3.4 The Jackknife for Non-Linear Functions $g(\hat{\theta})$	44
3.5 Monte Carlo Study	50

CHAPTER 4. PREDICTION OF SMALL AREA MEANS WHEN THE VARIANCE COMPONENTS ARE UNKNOWN	60
4.1 Approximate Mean Square Error	62
4.2 Estimation of Mean Square Error	72
4.3 A Monte Carlo Study of the Nested Error Model	76
CHAPTER 5. EXTENSION OF NESTED ERROR REGRESSION MODEL TO STRATIFICATION	91
5.1 The Model and Assumptions	92
5.2 BLUP when the Variance Components are Known	94
5.3 Approximation to MSE of the Estimated BLUP	96
CHAPTER 6. ESTIMATION OF CORRELATED RESPONSE VARIANCE (CRV)	101
6.1 Interviewer Allocation Scheme and the Model	102
6.2 Minimum Norm Quadratic Unbiased Estimation (MINQUE)	104
6.3 Optimal Estimator of σ_a^2 : Data Set A	106
6.4 Optimal Estimator of σ_a^2 : Data Set A and Data Set B Combined	109
6.5 A Monte Carlo Study	115
CHAPTER 7. SUMMARY AND CONCLUSIONS	122
APPENDIX	125
A.1 Proof of Lemma 3.2.	125
A.2 Values of n_i , x_{ij} and $\bar{X}_i$ ($i=1, \dots, t; j=1, \dots, n_i$) used in Monte Carlo Studies.	130
REFERENCES	131

LIST OF TABLES

	Page
Table 3.1a: Coverage rates of the confidence intervals for the nominal level 95% (90%) and percentage relative biases of μ , t_J , $v_N(\hat{\sigma})$ and $v_J(\hat{\sigma})$ (with $t = 10$)	54
Table 3.1b: Coverage rates of the confidence intervals for the nominal level 95% (90%) and percentage relative biases of μ , t_J , $v_N(\hat{\sigma})$ and $v_J(\hat{\sigma})$ (with $t = 20$)	57
Table 4.1: Relative Efficiency (R.E.) of the estimated BLUP with respect to (w.r. to) $t_1(1)$, $t_1(2)$ and $t_1(3)$.	81
Table 4.2a: Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: ($t = 10$, Normal distribution).	87
Table 4.2b: Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: ($t = 20$, Normal Distribution)	87
Table 4.3a: Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: ($t = 10$, Exponential distribution).	88
Table 4.3b: Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: ($t = 20$, Exponential Distribution)	88
Table 4.4a: Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: ($t = 10$, Double Exponential Distribution)	89
Table 4.4b: Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: ($t = 20$, Double Exponential Distribution).	89

Table 4.5a:	Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: (t = 20, Uniform Distribution)	90
Table 4.5b:	Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: (t = 20, Uniform Distribution)	90
Table 6.1:	Relative Efficiency of CRV estimators with respect to the optimal estimators $\hat{\sigma}_a^2$, and the percentage increase in efficiency (IE) by using Data Set B (p = 5, P = 30, n = 10, $\sigma_e^2 = 1.0$)	121
Table A:	Values of n_i , x_{ij} and $\bar{X}_1$ used in Monte Carlo Studies	130

CHAPTER 1

INTRODUCTION

Dalenius (1977a), in his article "Strain at a Gnat and Swallow a Camel, or, the Problem of Measuring Sampling and Non-sampling Errors", commented that statisticians, at many times, pay considerable attention to measuring sampling error which is often not very large. At the same time they tend to neglect non-sampling errors (e.g., response and non-response errors). However, one area where the magnitude of the sampling error may be large is in the context of the estimation of the "small area"-parameters based on a sample from a finite population. In this thesis, using variance component models, problems associated with the estimation of small area parameters are studied. The estimation of response error variance component in the context of estimating population totals from sample survey data is also investigated. Chapters 2 to 5 deal with the estimation of small area means and of their mean square errors. The estimation of response error variance is studied in Chapter 6. In Sections 1.1 and 1.2 of Chapter 1, an overview of small area methodology and a discussion of response errors respectively are presented. A thesis outline is provided in Section 1.3.

1.1 An Overview of Small Area Methodology.

Most large-scale nationwide surveys are designed to provide estimates for a variety of characteristics at national (population) level with a prespecified reliability for at least some characteristics. The publication of estimates at the population level induced a demand

for similar estimates for subpopulations which are often referred to as 'domains'. These domains usually vary in size. Purcell and Kish (1980) classified domains into; (i) 'major domains', consisting of 1/10 or more of the population units; (ii) 'minor domains', consisting of between 1/10 and 1/100 of the population units; (iii) 'mini domains', consisting of between 1/100 and 1/10000 of the population units and (iv) 'rare domains' consisting of less than 1/10000 of the population units. They further commented that "For major domains, standard estimates, basically without a bias, are generally available from probability methods of survey sampling. However, frequently for minor domains and usually for mini domains, the standard methods of survey estimation breaks down, because the sample bases are ordinarily too small for any useable reliability, and new methods are needed. For rare domains, sample surveys are usually useless: separate and distinct methods are required...". (See Purcell and Kish (1980); p.4.) Throughout this thesis minor and mini domains are considered to be "small area". Suppose that a target population consisting of N units is divided into t mutually exclusive and exhaustive small areas, S_i ($i=1, \dots, t$). Let n_i ($i=1, \dots, t$) be the number of units falling in the small area S_i , in a sample S of fixed size n drawn from this population. 'Small area estimation' entails the estimation of t small area means for a characteristic y of interest, i.e.,

$$\mu_i = \sum_{j \in S_i} y_{ij} / N_i$$

where N_i is the total number of units in the i -th small area

($i=1, \dots, t$) and y_{ij} is the value of the characteristic y for the

j-th unit in the i-th small area.

Many techniques for estimating the means μ_i have been proposed. These techniques, which differ considerably in their assumptions and use of available data, can be classified into three groups: (i)

'Symptomatic' methods (ii) traditional sample survey methods and (iii) methods utilizing both sample survey and census or administrative data. Purcell and Kish (1979), Harter (1983) and Dumaïs et al. (1985) have provided overviews of these techniques.

(1) 'Symptomatic' methods: These methods, primarily developed by demographers for the estimation of population characteristics utilize current administrative data on symptomatic variables such as births, deaths, school enrolments, etc., in combination with previous census data.

The reliability of estimates developed using administrative data is necessarily dependent on the coverage and quality of the administrative

files. This class of techniques includes census component methods I and II, developed by the U.S. Bureau of the Census in the 1940's (U.S. Bureau of the Census, 1949, 1966), the Vital Rates Method proposed by Boque (1950), the Housing Unit Method by Rives (1976) and the Ratio-

Correlation Method by Schmitt and Crosetti (1954). Detailed descriptions of these methods are given in Harter (1983) and Dumaïs et al. (1985).

(ii) Traditional sample survey based method: It is sometimes possible to obtain useable estimates for small areas from standard sample survey techniques, provided the sample falling into the small area is sufficiently large.

Under simple random sampling, three traditional estimators of the small area mean μ_1 are provided below. These estimators can be readily extended to general sampling designs.

Direct estimator: For a single characteristic, y , the estimator of μ_1 is given by

$$\hat{\mu}_1(D) = \bar{y}_1; \quad i=1, \dots, t,$$

where $\bar{y}_1$ is the mean of the sample observations falling in the small area S_1 .

Ratio estimator: When the information on an auxiliary variable x is available for all the units in the entire population, direct estimators can be improved when y is positively correlated with x by taking

$$\hat{\mu}_1(\text{Ratio}) = \bar{X}_1 \bar{y}_1 / \bar{x}_1; \quad i=1, \dots, t,$$

where $\bar{X}_1 = N_1^{-1} \sum_{j \in S_1} x_{1j}$, with x_{1j} as the value corresponding to

the j -th unit in the small area S_1 , and $\bar{x}_1$ is the mean of x for the sample observations falling in the small area S_1 . This estimator

$\hat{\mu}_1(\text{Ratio})$ is biased under the repeated sampling framework i.e., design-biased.

Regression estimator: Like the ratio estimator defined above, a regression estimator can be used to improve the direct estimator when y is correlated with the auxiliary variable x :

$$\hat{\mu}_1(\text{Reg}) = \bar{y}_1 + \hat{\beta}_1(\bar{X}_1 - \bar{x}_1),$$

where $\hat{\beta}_1$ is the regression estimator of y on x based on the sample observations falling in the small area S_1 .

These sample-based estimators, generally, are not reliable as they are based only on the sample falling in the small area under consideration. This drawback prompted the study of a new class of estimators that utilize reliable estimates of the large area characteristics. These are described below:

(iii) Methods utilizing both sample survey and census or administrative data: These methods rely upon the modelled relationships between the survey data and the auxiliary data (census data or administrative records):

Synthetic estimation: Synthetic estimators are obtained by combining survey estimators of large area characteristics with census or other data services such as administrative records pertaining to the small area under consideration. Thus, suppose $\bar{X}_i$ denote the mean of the auxiliary variable which is available for the entire i -th small area and let $\bar{Y}$ be a sample survey estimator of the mean variable of interest for the entire population (e.g., $\bar{Y} = \bar{y}_{..} = n^{-1} \sum_{i,j} y_{ij}$ under simple random sampling). Then a synthetic estimator for the i -th small area mean is given by

$$\hat{\mu}_i(S) = \bar{X}_i \cdot \bar{Y} / \bar{X},$$

where $\bar{X}$ is the sample survey estimator for the population mean of the auxiliary variable x . The estimator $\hat{\mu}_i(S)$ assumes that the small area populations μ_i ; $i=1, \dots, t$, for a given characteristic, obtained across areas, are approximately equal to the overall mean. The magnitude of the bias of this estimator evidently depends on the degree of

departure from this assumption.

— Another synthetic estimator, under simple random sampling design, is found by assuming the model

$$y_{ij} = x_{ij}\beta + e_{ij}; \quad i=1, \dots, t; \quad j=1, \dots, n_i,$$

with expectation $E(e_{ij}) = 0$ and variance $v(e_{ij}) = \sigma_e^2$, and obtaining the least square estimator of $\bar{X}_{i.}\beta$, leading to

$$\hat{u}_1(S) = \bar{X}_{i.}\hat{\beta},$$

where $\hat{\beta}$ is the least square estimator of β .

A drawback of synthetic estimators is that they do not directly utilize the small area data on y . Further discussions and evaluations of synthetic estimators are given in Gonzalez (1973), Laake (1979) and Levy (1971).

Categorical data analysis approach: At times small area estimates are desired for categorical variables such as age, sex and marital status. In such cases, the problem of small area estimation can be structured within a categorical data analysis framework. For example, assume the data are categorized by two attributes, age (Rows) and sex (Columns). The objective is to estimate the number of males and females in each age group in the small area. It is assumed that survey estimates for age group and sex totals are available but not reliable at the small area level. It is also assumed that auxiliary information is available at the small area level from an alternative source (e.g., census or

administrative records). Small area estimates are obtained using the Deming-Stephan's iterative proportional fitting (IPF) procedure (Deming and Stephan (1940)). Each row of the auxiliary data table is multiplied by a constant so that the entries in a row aggregate to the survey estimated row total and this process, known as 'raking', is repeated also on the columns. The rows and columns are raked alternatively until the small area estimates in the age-by-sex table converge. The entries in the final age-by-sex table constitutes the small area estimates. Purcell and Kish (1980) call such estimators as 'members of the class of structure preserving estimators (SPREES)'. A synthetic estimator is also a member of the SPREE class. In general, SPREES tend to be less biased than a customary synthetic estimator, but relatively more complex to apply. However, generalized computer programmes are available to carry out the computational procedures. A detailed description of SPREE is given in Purcell and Kish (1980).

Composite method of estimation: Composite estimators are obtained by combining two or more different estimators. One such composite estimator is derived by combining a survey based estimator e.g., $\hat{\mu}_1(D)$, with a synthetic estimator e.g., $\hat{\mu}_1(S)$. Although it is possible to obtain the optimal combination of these two estimators which minimizes the overall variance, the weights generally involve a knowledge of unknown parameters. Drew, Singh and Choudhry (1982), and Srinath and Hidioglou (1985) have proposed composite estimators with sample-dependent weights. These weights have the characteristic that when no sampled unit falls

within the small area, the composite estimator is identical to the synthetic estimator.

Holt, Smith and Tomberlin (1979) assumed the model

$$y_{ij} = \mu + e_{ij}, \quad i=1, \dots, t; \quad j=1, \dots, n_i$$

with $E(e_{ij}) = 0$ and $\text{Var}(e_{ij}) = \sigma_e^2$ and arrived at the composite estimator

$$\hat{\mu}_1(\text{HST}) = n_1 N_1^{-1} \bar{y}_i + (1 - n_1 N_1^{-1}) \bar{Y}..$$

using a predictive approach (Royall (1970)).

Generalized Regression Method: Särndal (1984) proposed an approximate design unbiased estimator that incorporates auxiliary information through the generalized regression technique. For a sample design with arbitrary inclusion probability $\pi_j(i)$ for the j -th unit in the i -th small area, his estimator, motivated by the general linear model:

$$y_{ij} = x_{ij} \beta + e_{ij}, \quad i=1, \dots, t; \quad j=1, \dots, n_i$$

with $E(e_{ij}) = 0$ and $\text{Var}(e_{ij}) = \sigma_e^2 x_{ij}$, is given by

$$\hat{\mu}_1(\text{Sar}) = N_1^{-1} \sum_{j=1}^{n_1} y_{ij} / \pi_j(i) + \hat{\beta} (\bar{X}_{1.} - N_1^{-1} \sum_{j=1}^{n_1} x_{ij} / \pi_j(i)),$$

where

$$\hat{\beta} = [\sum_{i,j} y_{ij} / \pi_j(i)] / [\sum_{i,j} x_{ij} / \pi_j(i)]$$

Under simple random sampling design, i.e., $\pi_j(i) = n N^{-1}$, the above estimator reduces to

$$\hat{\mu}_1(\text{Sar}) = (N n_1 / N_1 n) \bar{y}_i + \frac{\bar{Y}..}{\bar{x}..} \{ \bar{X}_{1.} - (N n_1 / N_1 n) \bar{x}_i \}.$$

James-Stein and Empirical Bayes methods: When the means of several small areas are being estimated simultaneously, James-Stein theory (James and Stein (1961)) can be used to obtain 'shrinkage estimators'. The small area estimates are shrunk towards the overall mean. As an example, suppose that for given θ_i , Y_i is normally distributed with mean θ_i and known variance D_i ($i=1, \dots, t$)

$$\text{i.e. } Y_i | \theta_i \sim N(\theta_i, D_i) \quad i = 1, \dots, t.$$

Then James and Stein (1961) showed that when $D_i \equiv D$ an estimator of the form

$$\hat{\theta}_i(\text{JS}) = \tilde{\theta}_i + (1 - (t-2)D/S)(Y_i - \tilde{\theta}_i),$$

where $\tilde{\theta}_i$ is the initial guess of θ_i and $S = \sum_{i=1}^t (Y_i - \tilde{\theta}_i)^2$, exists such that

$$E \sum_{i=1}^t (\hat{\theta}_i(\text{JS}) - \theta_i)^2 < E \sum_{i=1}^t (Y_i - \theta_i)^2,$$

i.e., the total mean square error (MSE) of James-Stein estimator over small areas is uniformly smaller than the total MSE of the unbiased

estimators. Efron and Morris (1973) established that $\hat{\theta}_i(\text{JS})$ is in fact an empirical Bayes estimator. Their proof involves the following steps:

Step 1: By using the prior distribution $\theta_i \sim N(\tilde{\theta}_i, A)$ with known A , they obtained the Bayes estimator

$$\hat{\mu}_i(B) = \tilde{\theta}_i + [1 - \frac{D}{D+A}](Y_i - \tilde{\theta}_i).$$

Step 2: From the marginal distribution of Y_1 which is given by

$$Y_1 \sim N(\mu_1, D+A),$$

it follows that

$$E[(t-2)/S] = 1/(D+A).$$

Step 3: Thus when A is unknown, by replacing $(D+A)^{-1}$ by its estimator $(t-2)/S$ in the Bayes estimator $\mu_1(B)$, yields an empirical Bayes estimator which is identical to $\mu_1(JS)$.

In the context of small area estimation with no auxiliary information, Carter and Rolph (1974) adapted the James-Stein estimator within the empirical Bayes framework. Their estimator, when D_i 's are known, is

$$\hat{\mu}_1(EB) = \frac{D_i}{\hat{A} + D_i} \bar{y}_{..} + \frac{\hat{A}}{\hat{A} + D_i} \bar{y}_{i..}$$

where $\hat{A}$ is the estimator obtained from the marginal distribution of $\bar{y}_{i..}$. For estimating per capita income for small areas (populations less than 1000) in U.S.A., Fay and Herriot (1979) adapted an empirical Bayes procedure. Their procedure assumes the knowledge of auxiliary variables related to the small area means of the study variable... Section 2.1 provides a description of their work.

Battese and Fuller (1982) obtained a James-Stein type estimator in the context of utilizing Landsat satellite data for predicting mean hectares under corn and soyabean for 12 counties (small areas) in Iowa. For their procedure, the values of auxiliary variables for

the sampled units in each small area and the population mean of auxiliary variables for each small area are needed. A description of their work is provided in Chapters 2 and 4 of this thesis.

All these James-Stein type estimators are composite estimators with sample-dependent weights, which are assigned in a natural way. Raq (1984) provides a detailed discussion on empirical Bayes methods in the context of small area estimation.

Stroud (1984) and Dempster and Raghunathan (1985) adapt a purely Bayesian approach to small area estimation by assuming a prior distribution on the parameter, e.g., a prior on the parameter A in Carter and Rolph model (1974).

1.2 Response Errors: an Overview

Standard textbook formulae for the variance of an estimator under probability sampling are based on the assumption that the observation y_i on the i -th unit is the correct value, say y_i^* , for that unit. This assumption holds reasonably well only in the simpler types of surveys in which the measuring devices are accurate and quality of work is high. When the errors of measurement $y_i^* - y_i$; $i=1, \dots, N$, known as 'response errors' are not 'small', then textbook formulae can lead to serious under-estimates of standard errors. Thus, over the years, the importance of studying measurement errors has been emphasized; Dalenius (1977 b) provides a comprehensive bibliography. There are various causes for measurement errors. Some of the most common ones are:

- (i) a lack of trained or experienced interviewers;
- (ii) a respondent may not possess accurate information;
- (iii) a misinterpretation of questions;
- (iv) errors in recording, editing and tabulating the survey results.

Apart from measurement errors, the presence of other non-sampling errors viz., non-response errors, are also common in complex surveys. Taking into account these errors, Hansen, Hurwitz and Bershad (1961) decomposed the total mean square error of the sample estimate into sampling error and non-sampling error components.

A mathematical model for non-sampling errors was first proposed by Sukhatme and Seth (1952). Cochran (1968), Battese, Fuller and Hickman (1976), Dodds and Smith (1973) and Hartley and Rao (1978) have proposed similar models. Under these models, it is assumed that a survey response differs from its 'true value' by additive error contributions associated with various source of error, such as interviewers, coders and so on. Also, these models indicate that the errors made by a specified error source (e.g., interviewer, coders) are correlated. The variance of such errors is called the 'correlated response variance' (CRV) (see Cochran (1977)). Most of the studies (e.g., Hansen et al. (1959), Fellegi (1974)) dealing with the estimation of CRV use the 'interpenetration' of interviewers (or coders) technique to meet the condition for estimability of the response variance components. A necessary and sufficient condition for estimability is given in Hartley, Rao and Lamotte (1978) and Biemer (1978).

The technique of 'interpenetration' involves randomly dividing the sample under consideration into k subsamples of equal sizes, where k is the number of interviewers (coders). These subsamples are then randomly assigned to the k interviewers. Since interpenetration is expensive, only a fraction of the total sample will be subjected to interpenetration in practice. Fellegi (1974) obtained two estimators of CRV: the first estimator based only on the interpenetrated data set and the second estimator based on the entire sample. As an improved method of estimation, he suggested to use the arithmetic mean of the above two estimators. His estimators are given in Chapter 6 of this thesis.

1.3 Thesis Outline

1. The regression models with components of variance error structure, proposed by Battese and Fuller (1982), Dempster, Rubin and Tsutakawa (1981) and Fay and Herriot (1979), have been used to obtain small area 'predictors'. In Section 1 of Chapter 2, descriptions of these models are provided. The small area means, under these models, are expressed as linear combinations of fixed and random effects. In Section 2.2, under each of the above models, when the variance components are known, the 'Best Linear Unbiased Predictors' (BLUPs) of the small area means are obtained using Henderson's (1975) result on BLUP for general linear models.

2. When the variance components of error structure are unknown,

the BLUPs need to be estimated by replacing the unknown variance components by their asymptotically consistent estimators. The method of fitting of constants for obtaining such estimators is discussed in Section 1 of Chapter 3. A jackknife procedure for estimating the variances and covariances of the variance component estimators is presented in Section 3.3. The jackknife procedure is extended in Section 3.4 to deal with the estimation of a non-linear function of variance components and its variance.

3. An exact expression for the mean square error (MSE) of the estimated BLUP is not tractable. In Section 1 of Chapter 4 an approximate expression for the MSE of an estimated BLUP is derived under normality assumption. Three mean square error estimators are suggested in Section 4.2. Using a Monte Carlo study, the finite sample behaviour of the approximate MSE and the MSE estimators are examined in Section 4.3. The effect of normality is also studied.

In Chapter 5, the Battese-Fuller model is extended to the stratified sampling case along the lines of Walker and Sigman (1984). For this model, the BLUP and an approximation to MSE of the estimated BLUP are provided.

4. A variance component model for correlated response error variance is suggested in Section 1 of Chapter 6. Optimal estimators of the CRV based on (a) the interpenetrated data set and (b) both the interpenetrated and the non-interpenetrated data sets are obtained in Sections 6.3 and 6.4. The method of minimum norm quadratic unbiased estimation (MINQUE) is employed to obtain these estimators.

BEST LINEAR UNBIASED PREDICTOR

In this chapter general linear model methods are applied to obtain small area predictors. For this purpose, the models proposed by Battese and Fuller (1982), Dempster, Rubin and Tsutakawa (1981) and Fay and Herriot (1979) are considered. A detailed description of these models is given in Section 2.1. For the Battese-Fuller, and Dempster models, the values of auxiliary variables for the sampled units in each small area and the population means of auxiliary variables for each small area are required. On the other hand, the Fay-Herriot model only requires the knowledge of auxiliary variables related to small area means of the study variable.

The works of Henderson (1975) and Harville (1976) on best linear unbiased predictors (BLUP) in a general linear model is reviewed in Section 2.2. The BLUP's of small area means and their mean square errors (MSE) under the three models are spelled out in Section 2.3.

2.1. Description of Small Area Models

2.1(a). Nested Error Regression Model (model 1).

In the context of utilizing Landsat Satellite data for predicting mean hectares under corn for 12 counties (small areas) in North-Central Iowa, Battese and Fuller (1982) proposed a nested error regression model. The scatter plot of y (hectares of corn in a segment) against x (number of pixels of corn) indicated that the segments for a particular county fall roughly around a regression line, but they tend to cluster

a bit above or below the line. One way of representing this type of phenomenon is to use a nested error regression model given by

$$y_{ij} = x_{ij}\beta + u_{ij}, \quad (2.1)$$

and

$$u_{ij} = v_i + e_{ij}, \quad i=1, \dots, t; \quad j=1, \dots, n_i, \quad (2.2)$$

where y_{ij} is the reported value of the study characteristic y (e.g. number of hectares of corn) on the j^{th} sampled unit in the i^{th} small area; $x_{ij} = (x_{ij1}, \dots, x_{ijk})$ is a $1 \times k$ vector of corresponding values of auxiliary variables (e.g. number of pixels of corn), $\beta = (\beta_1, \dots, \beta_k)$ is a $k \times 1$ vector of unknown parameters and n_i is the number of sampled units (segments) observed in the i^{th} small area. The random errors v_i , $i=1, \dots, t$ are assumed to be independent of the unit errors e_{ij} , $j=1, \dots, n_i$, $i=1, \dots, t$, and both are assumed to be independent and identically distributed (i.i.d) with zero mean and variances σ_v^2 and σ_e^2 respectively.

Remark 2.1. The model (2.1) - (2.2) allows an intercept term if we let $x_{ij1} = 1$ and $\beta_1 = \alpha$.

Under the above assumptions it follows that the covariance structure of the errors u_{ij} is given by

$$E(u_{ij}u_{i'j'}) = \begin{cases} \sigma_v^2 + \sigma_e^2 & \text{if } i=i' \text{ and } j=j' \\ \sigma_v^2 & \text{if } i=i' \text{ and } j \neq j' \\ 0 & \text{if } i \neq i', \end{cases} \quad (2.3)$$

where E denotes the expectation operator. Thus the covariance matrix of

$$u = \begin{matrix} \text{Col.} & \text{Col.} \\ \text{list} & \text{list} \end{matrix} (u_{ij}) = (u_{11}, \dots, u_{1n_1}, \dots, u_{t1}, \dots, u_{tn_t})$$

is given by

$$\begin{aligned} V &= \text{diag} (\sigma_e^2 I_{n_i} + \sigma_v^2 1_{n_i} 1_{n_i}') \\ &= \text{diag} (V_i) \end{aligned} \quad (2.4)$$

where I_{n_i} is an identity matrix of order n_i and $1_{n_i} = \text{Col.}(1)$.

The problem is to predict (or estimate) the small area means

$$\mu_i = \bar{X}_i \beta + v_i + \bar{E}_i \quad (2.5)$$

where $\bar{X}_i = N_i^{-1} \sum_{j=1}^{N_i} x_{ij}$ is the mean of the x_{ij} for the N_i population units in the i th small area, and $\bar{E}_i = N_i^{-1} \sum_{j=1}^{N_i} e_{ij}$. We assume $\bar{X}_i$ is

known (e.g. mean number of pixels from the Satellite data), and also

assume that N_i is large so that μ_i may be taken as (by noting that

$E_i = 0$ since $E(e_{ij}) = 0$)

$$\mu_i = \bar{X}_i \beta + v_i \quad (2.6)$$

Remark 2.2. The nested error regression model (2.1) - (2.2) can also be viewed as a random intercept model by writing (see Remark 2.1)

$$y_{ij} = \alpha_i + z_{ij} \gamma + e_{ij}$$

where $\alpha_i = \alpha + v_i$ are the random intercepts and $\gamma = (\beta_2, \dots, \beta_k)'$,

$z_{ij} = (x_{ij2}, \dots, x_{ijk})$. In the case of no intercept, we can also interpret the model as above but with known $\alpha = 0$.

A more general model with random slopes, proposed by Dempster, Rubin and Tsutakawa (1981), is described below.

2.1(b). Random Regression Model (model 2).

We are concerned here with the following model:

$$y_{ij} = x_{ij}\beta_i + e_{ij}, \quad (2.7)$$

$$\beta_i = \beta + v_i, \quad j=1, \dots, n_i; i=1, \dots, t,$$

where y_{ij} , x_{ij} , β , e_{ij} are as defined in the previous model 1, and $v_i = (v_{i1}, \dots, v_{ik})'$ are identically distributed with mean 0 and a $k \times k$ covariance matrix ψ . Furthermore, v_i and the e_{ij} are stochastically independent. The covariance matrix of $Y =$

Col. (y_{ij}) is given by
 $1 \leq i \leq t \quad 1 \leq j \leq n_i$

$$V = \sigma_e^2 I_n + \text{diag} (Z_i' \psi Z_i) \quad (2.8)$$

$$1 \leq i \leq t$$

where $Z_i = \text{Col} (x_{ij})$. Under this model we are interested in predicting the small area means

$$\mu_i = \bar{X}_i \beta_i = \bar{X}_i \beta + \bar{X}_i v_i$$

2.1(c). Pay-Herriot Model (model 3).

In the context of estimating per capita income for small areas (populations less than 1000) in the U.S.A., Fay and Herriot (1979) adapted an empirical Bayes procedure. Their work is summarized below.

Suppose $\bar{y} = (\bar{y}_1, \dots, \bar{y}_t)'$ denotes a t -vector of independent unbiased

domain estimators of the small area mean vector $\bar{Y} = (\bar{Y}_1, \dots, \bar{Y}_t)'$ with known variances $D_1, \dots, D_t$. Suppose also that a k -vector of benchmark variables $x_i = (x_{i1}, \dots, x_{ik})$ is available for each small area $i = 1, \dots, t$ and that the $\bar{Y}_{i.}$ are independent random variables with means $x_i \beta$ and unknown variance A , where β is a k -vector of unknown parameters. Under normality for both $\bar{Y}_{i.}$ (given $\bar{Y}_{i.}$) and $\bar{Y}_{i.}$, an empirical Bayes estimator of $\bar{Y}_{i.}$ is given by

$$\bar{Y}_i^* = \frac{A}{A + D_i} \bar{Y}_{i.} + \frac{D_i}{A + D_i} v_i, \quad (2.9)$$

where

$$v_i = x_i \hat{\beta}; \quad \hat{\beta} = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} \bar{Y},$$

$$X' = (x_1', \dots, x_t'); \quad \hat{V} = \text{diag} (\hat{A} + D_i), \quad 1 \leq i \leq t$$

and $\hat{A}$ is a solution of the non-linear equation

$$\sum_{i=1}^t \frac{(\bar{Y}_{i.} - \hat{v}_{i.})^2}{\hat{A} + D_i} = t - k.$$

Note that v_i depends on $\hat{A}$.

The Fay-Herriot model can be restated in terms of the following linear model:

$$\bar{Y}_{i.} = \bar{Y}_{i.} + e_i \quad (2.10)$$

$$\bar{Y}_{i.} = x_i \beta + v_i, \quad i = 1, \dots, t,$$

where e_i and v_i are stochastically independent and normally distributed with zero means and variances D_i and A respectively.

The best linear unbiased predictor (BLUP) for a general mixed linear model is presented in the next section without normality assumption for the random errors. All the three models discussed in this section are special cases of the general mixed linear model.

2.2. Prediction Under a General Mixed Linear Model (Known Variance Components).

A general mixed linear model is given by

$$\begin{aligned} \underline{y} &= \underline{X}\underline{\beta} + \underline{Z}\underline{v} + \underline{e} \\ &= \underline{X}\underline{\beta} + \underline{u}, \end{aligned} \quad (2.11)$$

where $\underline{y}$ is an $n \times 1$ vector of observations, $\underline{X}$ and $\underline{Z}$ are known $n \times k$ and $n \times t$ matrices, $\underline{\beta} = (\beta_1, \dots, \beta_k)'$, a $k \times 1$ vector of unknown parameters. Further, $\underline{v}$ and $\underline{e}$ are unobservable random $t \times 1$ and $n \times 1$ vectors respectively with mean 0 and covariance matrix

$$\text{var} \begin{pmatrix} \underline{v} \\ \underline{e} \end{pmatrix} = \begin{pmatrix} \underline{G} & \underline{0} \\ \underline{0} & \underline{R} \end{pmatrix}, \quad (2.12)$$

where $\underline{G}$ and $\underline{R}$ are both non-singular matrices depending on some parameters $\underline{\theta}$. The elements of $\underline{\theta}$ are called variance components. The small area means can be represented in terms of the model (2.11) as a linear combination of fixed and random effects:

$$\mu_i = \underline{k}_i' \underline{\beta} + \underline{m}_i' \underline{v} \quad (2.13)$$

where $\underline{k}_i$ and $\underline{m}_i$ are $k \times 1$ and $t \times 1$ known vectors.

Henderson (1975) and Harville (1976) provide some general results on estimating $\underline{k}'\underline{\beta} + \underline{m}'\underline{y}$. Their results are summarized in Theorem 2.1 below. Before giving the theorem, we give some definitions which are borrowed from Kackar (1979).

Definition 2.1. An estimator (predictor) $t(\underline{y})$ is said to be linear if $t(\underline{y})$ is a linear function of the data vector, that is, $t(\underline{y}) = \underline{c} + \underline{\ell}'\underline{y}$, where $\underline{c}$ is a constant and $\underline{\ell}$ is a vector of constants.

Definition 2.2. A predictor $t(\underline{y})$ is said to be an unbiased predictor of $\underline{k}'\underline{\beta} + \underline{m}'\underline{y}$ if

$$E(t(\underline{y}) - \underline{k}'\underline{\beta} - \underline{m}'\underline{y}) = 0.$$

Definition 2.3. A linear combination $\underline{k}'\underline{\beta} + \underline{m}'\underline{y}$, is said to be estimable if it admits a linear unbiased estimator.

Definition 2.4. If $t(\underline{y})$ is a predictor of $\underline{k}'\underline{\beta} + \underline{m}'\underline{y}$, then the quantity $E(t(\underline{y}) - \underline{k}'\underline{\beta} - \underline{m}'\underline{y})^2$ will be referred to as the mean square error (MSE) of $t(\underline{y})$.

Definition 2.5. A linear unbiased predictor $t(\underline{y})$ of $\underline{k}'\underline{\beta} + \underline{m}'\underline{y}$ is called the best linear unbiased predictor (BLUP), if $MSE(t(\underline{y}))$ is less than or equal to the MSE of any other linear unbiased predictor $t^*(\underline{y})$ of $\underline{k}'\underline{\beta} + \underline{m}'\underline{y}$, with equality holding iff $t(\underline{y}) = t^*(\underline{y})$ with probability one.

Theorem 2.1. For the general mixed model (2.11) - (2.12), let the parameters $\underline{\theta}$ be known. Suppose $\underline{k}'\underline{\beta} + \underline{m}'\underline{y}$ is estimable and X is a full rank matrix, then the predictor

$$t(\theta, y) = k'\tilde{\beta} + m'GZ'V^{-1}(Y - X\tilde{\beta}) \quad (2.14)$$

is an essentially ~~unique~~ BLUP of $k'\beta + m'v$, where

$$\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (2.15)$$

and

$$V = R + ZGZ' \quad (2.16)$$

Furthermore, the mean square error (MSE) of $t(\theta, y)$ is given by

$$\begin{aligned} E(t(\theta, y) - k'\beta - m'v)^2 &= K'(X'V^{-1}X)^{-1}K \\ &\quad - k'(X'V^{-1}X)^{-1}X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}m \\ &\quad - m'(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X(X'V^{-1}X)^{-1}k \\ &\quad + m'(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X(X'V^{-1}X)^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}m. \end{aligned} \quad (2.17)$$

Also, if Y is a multivariate normal then $t(\theta, y)$ is an essentially unique best unbiased predictor of $k'\beta + m'v$.

Proof of Theorem 2.1. See Henderson (1975) and Harville (1976).

2.3 BLUPs Under The Small Area Models.

Using Theorem 2.1 the BLUPs and their mean square error are obtained for the three models considered in Section 2.1, when the variance components are known. The following Lemma on matrix inversion is needed.

Lemma 2.1. (Graybill (1969); Theorem 8.3.3 pp. 170).

Let C be a $k \times k$ matrix defined by $C = D + \alpha ab'$, where D is a non-singular diagonal matrix, a and b are each $k \times 1$ vectors, and α is a scalar such that $\alpha \neq -(\sum_{i=1}^k a_i b_i d_i^{-1})^{-1}$. Then the inverse of C

is given by

$$C^{-1} = D^{-1} + \gamma a^* b^*$$

where

$$\gamma = -\alpha(1 + \alpha \sum_{i=1}^k a_i b_i d_{ii}^{-1})^{-1}, \quad a_i^* = a_i d_{ii}^{-1}, \quad b_i^* = b_i d_{ii}^{-1}$$

and d_{ii} is i^{th} diagonal element of D .

Theorem 2.2. For the nested error regression model defined by (2.1) -

(2.3), let the variance components σ_e^2 and σ_v^2 be known. Then the BLUP of $\mu_i = \bar{X}_{i.} \beta + v_i$ is given by

$$t_i(\sigma^2, y) = \bar{X}_{i.} \beta + \gamma_i (\bar{y}_{i.} - \bar{X}_{i.} \beta) \quad (2.18)$$

where $\sigma^2 = (\sigma_e^2, \sigma_v^2)$ and

$$\gamma_i = \sigma_v^2 (\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-1}. \quad (2.19)$$

Furthermore,

$$E(t_i(\sigma^2, y) - \mu_i)^2 = (\bar{X}_{i.} - \gamma_i \bar{X}_{i.}) (X' V^{-1} X)^{-1} (\bar{X}_{i.} - \gamma_i \bar{X}_{i.})' + (1 - \gamma_i) \sigma_v^2, \quad (2.20)$$

where V is as defined by (2.4).

Proof of Theorem 2.2. Writing the nested error regression model (2.1) -

(2.3) in terms of the general mixed model (2.11) - (2.12), we have

$$Y = \text{Col}_{1 \leq i \leq t} \text{Col}_{1 \leq j \leq n_i} (y_{ij}), \quad X = \text{Col}_{1 \leq i \leq t} \text{Col}_{1 \leq j \leq n_i} (x_{ij}),$$

$$Z = \text{diag Col } (1), \quad \bar{v} = \text{Col } (v_i), \quad e = \text{Col } (e_{ij})$$

$$1 \leq i \leq t, 1 \leq j \leq n_i \quad 1 \leq i \leq t \quad 1 \leq i \leq t, 1 \leq j \leq n_i$$

and $G = \sigma_v^2 I_t$, $R = \sigma_e^2 I_n$, $\theta = (\sigma_e^2, \gamma_i)$. Now, apply Lemma 2.1 with $C = V_i$, $D = \sigma_e^2 I_{n_i}$, $\alpha = \sigma_v^2$ and $a = b = 1_{n_i}$. We obtain

$$V^{-1} = \text{diag } (v_i^{-1})$$

$$1 \leq i \leq t$$

$$= \text{diag } [(\sigma_e^2)^{-1} I_{n_i} - \gamma_i n_i^{-1} (\sigma_e^2)^{-1} 1_{n_i} 1_{n_i}']$$

$$1 \leq i \leq t$$

Hence, taking $k = \bar{x}_i$ and $m = \text{Col } (\delta_{i\ell})$ in (2.14), where $\delta_{i\ell} = 1$ if $i = \ell$, $\delta_{i\ell} = 0$, otherwise, we get the BLUP $t_i(\sigma^2, y)$ as given by (2.18). Also, it can be shown that

$$Z'R^{-1}Z + G^{-1} = \text{diag } [(1 - \gamma_i)^{-1} (\sigma_v^2)^{-1}]$$

$$1 \leq i \leq t$$

$$Z'R^{-1}X = \text{Col } (\gamma_i \bar{x}_i)$$

$$1 \leq i \leq t$$

Substituting the above quantities in (2.17), we get the result (2.20) after simplifications.

Remark 2.3. Battese and Fuller (1982) also obtained the predictor (2.18) using heuristic arguments, but they have not demonstrated that it is in fact a BLUP.

In view of the complexity involved in obtaining an explicit expression for the inverse of the covariance matrix V under model 2 (see (2.8)), we restrict ourselves to the special case of one auxiliary variable. The BLUP for this special case is stated below.

Theorem 2.3. For the random regression model (2.7) with $k = 1$, and $x_{ij1} = x_{ij}$, $\beta_{i1} = \beta_i$, let the variance components $\sigma_v^2 = \text{var}(v_i)$ and σ_e^2 be known. Then the BLUP of $\mu_i = \bar{x}_i \beta_i = \bar{x}_i \beta + \bar{x}_i v_i$ is given by

$$t_i(\sigma^2, y) = \bar{x}_i \tilde{\beta} + \gamma_i (\bar{y}_i - \bar{x}_i \tilde{\beta}), \quad (2.21)$$

where

$$\gamma_i = \sigma_v^2 \left[\sigma_v^2 + \sigma_e^2 (\sum_j x_{ij}^2)^{-1} \right]^{-1},$$

$$= \left[\sum_{i=1}^t \gamma_i (\sum_j x_{ij} y_{ij}) (\sum_j x_{ij}^2)^{-1} \right] \left(\sum_{i=1}^t \gamma_i \right)^{-1}.$$

Further,

$$E(t_i(\sigma^2, y) - \mu_i)^2 = (\bar{x}_i - \gamma_i \bar{x}_i)^2 \left[\sum_j x_{ij}^2 (\sigma_e^2 + \sigma_v^2 \sum_j x_{ij}^2)^{-1} \right]^{-1} + (1 - \gamma_i) \sigma_v^2. \quad (2.22)$$

The proof of this theorem follows along the lines of Theorem 2.2, by taking X, Z, Y, k and m as $X = \text{Col}_{1 \leq i \leq t} \text{Col}_{1 \leq j \leq n_i} (x_{ij})$;
 $Z = \text{diag}_{1 \leq i \leq t} \text{Col}_{1 \leq j \leq n_i} (x_{ij})$; $V = \sigma_e^2 I_n + \sigma_v^2 \text{diag}_{1 \leq i \leq t} [(\text{Col}_{1 \leq j \leq n_i} (x_{ij})) (\text{Col}_{1 \leq j \leq n_i} (x_{ij}))']$;

$k = \bar{x}_i$ and $m = \text{Col}_{1 \leq l \leq t} (\bar{x}_i, \delta_{il})$ in (2.14) and (2.17).

Theorem 2.4. For the Fay-Herriot model defined by (2.10), let the variance components A and D_i ($i=1, \dots, t$) be known. Then the BLUP of $\bar{y}_i = x_i \beta + v_i$ is given by

$$t_i(A, \bar{y}) = x_i \tilde{\beta} \frac{D_i}{A + D_i} + \frac{A}{A + D_i} \bar{y}_i. \quad (2.23)$$

where $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$. Further,

$$E(t_i(A, \bar{Y}) - \bar{Y}_i)^2 = x_i' V(\tilde{\beta}) x_i + A D_i (A + B_i)^{-1}, \quad (2.24)$$

where $V(\tilde{\beta}) = \text{var}(\tilde{\beta})$.

The proof of this theorem follows along the lines of Theorem 2.2, by

taking $k' = \bar{X}_i$; $m = \text{Col}_{1 \leq l \leq t} (\delta_{il})$, $B = A I_t$; $R = \text{diag}(D_i)$; $1 \leq i \leq t$

$X = \text{Col}_{1 \leq i \leq t} (x_i)$ and $Z = I_t$ in (2.14) and (2.17).

Remark 2.4. The estimator (2.23) is a Bayes estimator under normality assumptions, as demonstrated by Fay and Herriot (1979).

CHAPTER 3

ESTIMATION OF VARIANCE COMPONENTS

All the results on the best linear unbiased predictor given in Chapter 2 are based on the assumption that the variance components, θ , are known. When the variance components are unknown, the BLUPs need to be estimated by replacing the unknown variance components by their respective asymptotically consistent estimators. Harville (1977) reviewed various methods of obtaining consistent estimators of the variance components for a general mixed model. Here we confine ourselves to quadratic estimates given by the well-known method of fitting of constants (called Henderson's method 3). Section 3.1 provides the relevant formulae for the estimates of variance components under the three small area models considered in Chapter 2.

The variances of these quadratic estimates and their estimators have been obtained under the normality assumption. However, these latter estimators are sensitive to non-normality. We have therefore obtained, in Section 3.3, robust estimators utilizing a "jackknife" procedure. The results are presented for the special case of a single auxiliary variable, to simplify the discussion.

The jackknife method is used in Section 3.4 to reduce the bias of estimates, $g(\hat{\theta})$, of nonlinear functions, $g(\theta)$, of the variance components, θ , and also to estimate the variance of $g(\hat{\theta})$, where $\hat{\theta}$ is a consistent estimator of θ . The jackknife variance estimator of $g(\hat{\theta})$ can be used to construct robust confidence intervals on $g(\theta)$. Previous

work (Arvesen (1969) and Arvesen and Schmitz (1970)) on the jackknife for variance components was restricted to the nested error model,

$$y_{ij} = \mu + v_i + e_{ij}, \text{ without a covariate.}$$

Section 3.4 establishes the bias reduction property of the jackknife estimate of $g(\theta)$ and the asymptotic consistency of the jackknife variance estimator, as the number of levels, t , of the random effect, v_i , tends to ∞ . Finally, the results of a simulation study on the jackknife method are reported in Section 3.5, under the nested error regression model with one auxiliary variable.

3.1 The Method of Fitting of Constants

Fuller and Battese (1973) outlined the procedure for obtaining variance components estimators under the nested error regression model using the method of fitting of constants. Their results are summarized below. In a similar way, the variance component estimators for the other two small area models are also provided in this section.

a. Nested Error Regression Model (Model 1)

The procedure involves two ordinary least squares regressions. At the first step, the estimated residuals $\{\hat{e}_{ij}; j=1, \dots, n_i; i=1, \dots, t\}$ are obtained from the ordinary least squares regression of the y-deviations, $y_{ij} - \bar{y}_{i.}$, on the x-deviations, $x_{ijs} - \bar{x}_{i.s}$, $s=1, 2, \dots, k$. A second set of residuals $\{\hat{u}_{ij}; j=1, \dots, n_i; i=1, \dots, t\}$ are obtained at the second step by performing the ordinary least squares regression of y_{ij} on x_{ij} ($\hat{e}_{ij}$ and $\hat{u}_{ij}$ are defined respectively by (3.22) and (3.23), see P.34).

From $\{\hat{e}_{ij}, \hat{u}_{ij}; j=1, \dots, n_i; i=1, \dots, t\}$ unbiased quadratic estimators

of $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$ are given by

$$\hat{\sigma}_e^2 = (n-t-k+\lambda)^{-1} \sum_i \sum_j e_{ij}^2, \quad (3.1)$$

and

$$\hat{\sigma}_v^2 = n^{-1} \left[\sum_i \sum_j \hat{u}_{ij}^2 - (n-k) \hat{\sigma}_e^2 \right], \quad (3.2)$$

where $\lambda = 0$ if the model (2.1) has no intercept term and $\lambda = 1$ otherwise,

$$n_* = n - \text{tr}[(X'X)^{-1} \sum_{i=1}^t n_i \bar{x}_i \bar{x}_i'] \quad (3.3)$$

and tr denotes the trace operator. It is possible for $\hat{\sigma}_v^2$, defined by (3.2), to be negative. In practice, if $\hat{\sigma}_v^2 < 0$, then one sets $\hat{\sigma}_v^2 = 0$. The estimator would no longer be unbiased, but it would be in the closure of the parameter space.

Under normality assumptions for the random errors v_i and e_{ij} , Battese and Fuller (1982) have shown that

$$\text{var}(\hat{\sigma}_e^2) = 2\sigma_e^4 (n-t-k+\lambda)^{-1}, \quad (3.5)$$

$$\text{var}(\hat{\sigma}_v^2) = 2n_*^{-2} [(n-t-k+\lambda)^{-1} (t-\lambda) \sigma_e^4 + 2n_* \sigma_e^2 \sigma_v^2 + n_{**} \sigma_v^4], \quad (3.6)$$

and

$$\text{cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2) = -(t-\lambda) n_*^{-1} \text{var}(\hat{\sigma}_e^2), \quad (3.7)$$

where

$$n_{**} = \sum_i n_i^2 - \text{tr}[(X'X)^{-1} \sum_i n_i^3 \bar{x}_i \bar{x}_i']. \quad (3.8)$$

The estimators $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$ defined by (3.1) and (3.2) can be expressed as quadratic forms (which are needed in Chapter 4) as shown below:

$$\hat{\sigma}_e^2 = y' C_1 y, \quad (3.9)$$

$$\hat{\sigma}_v^2 = y' C_2 y, \quad (3.10)$$

where

$$(n-t-k+1)C_1 = C'[I - CX(X'C'CX)^{-1}X'C']C, \quad (3.11)$$

$$C = I - \text{diag} [n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}'] \quad 1 \leq i \leq t, \quad (3.12)$$

$$n_* C_2 = I - X(X'X)^{-1}X' - (n-k)C_1. \quad (3.13)$$

b. Random Regression Model (Model 2)

For the random regression model defined by (2.7) with one auxiliary variable (i.e. $k=1$, $x_{ij1}=x_{ij}$, $\psi = \sigma_v^2$), unbiased estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ based on the method of fitting of constants, can be obtained as

$$\hat{\sigma}_e^2 = (n-t)^{-1} \sum_i \sum_j \hat{e}_{ij}^2, \quad (3.14)$$

where

$$\hat{e}_{ij} = y_{ij} - x_{ij} (\sum_j x_{ij} y_{ij}) / (\sum_j x_{ij}^2)^{1/2},$$

and

$$\hat{\sigma}_v^2 = \hat{n}_*^{-1} [\sum_i \sum_j \hat{u}_{ij}^2 - (n-1)\hat{\sigma}_e^2], \quad (3.15)$$

where

$$\hat{n}_* = \frac{\sum_i \sum_j x_{ij}^2 - 2(\sum_i \sum_j x_{ij}^2)^2 (\sum_i \sum_j x_{ij}^2)^{-1}}{\sum_i \sum_j x_{ij}^2} \quad \text{and} \quad \hat{u}_{ij} = y_{ij} - x_{ij} (\sum_j x_{ij} y_{ij}) / (\sum_j x_{ij}^2)^{1/2}$$

c. Fay-Herriot Model (Model 3)

For the Fay-Herriot model, the sampling variances D_1 are assumed to be known (see Fay and Herriot (1979)). An unbiased quadratic estimator of the variance component A can be obtained as

$$A = (t-1)^{-1} \sum_{i=1}^t \hat{u}_i^2 - \sum_{i=1}^t D_i (1 - x_i' (X'X)^{-1} x_i) \quad (3.16)$$

where

$$\hat{u}_i = \bar{y}_i - x_i' \hat{\beta} \quad (3.17)$$

and

$$\hat{\beta} = (X'X)^{-1} X' \bar{y} \quad (3.18)$$

The expressions for the variances of quadratic estimators can be obtained, under normality assumptions, using the following standard result (see Searle 1971, Sec. 2.6): If $\underline{y}$ is distributed as a multivariate normal with mean $\underline{0}$ and dispersion matrix Σ , then $\text{cov}(\underline{y}' A \underline{y}, \underline{y}' B \underline{y}) = 2 \text{tr} A \Sigma B \Sigma$, where A and B are any symmetric matrices.

3.2 The Jackknife Method: Introduction

The jackknife method was originally proposed by Quenouille (1949, 1956) as a means of bias reduction, while Tukey (1958) proposed the use of it in variance estimation and in constructing confidence intervals. To describe the jackknife procedure, suppose that $\hat{\theta}_n$ denotes an estimator of an unknown parameter θ based on n independent identically distributed (i.i.d.) observations. We divide the n observations at random into m groups each of k observations ($n = mk$). Let $\hat{\theta}_{m-1}^{(-i)}$ be the estimate obtained from the sample after omitting the i^{th} group of observations ($i = 1, \dots, m$), and let

$$\tilde{\theta}_i = m \hat{\theta}_m - (m-1) \hat{\theta}_{m-1}^{(-i)} \quad (3.19)$$

The $\tilde{\theta}_i$ were called "pseudo-values" by Tukey (1958). The jackknife

estimator of θ is given by

$$\hat{\theta}_J = m^{-1} \sum_{i=1}^m \hat{\theta}_i,$$

the average of the pseudo-values. The estimator $\hat{\theta}_J$ eliminates the n^{-1} term from the bias in $\hat{\theta}_n$, provided, $E(\hat{\theta}_n)$ can be expressed in the form $\theta + a_1/n + a_2/n^2 + \dots$, where $a_1, a_2, \dots$ do not depend on n .

Tukey proposed the jackknife variance estimator

$$v_J = (m-1)^{-1} m^{-1} \sum_i (\hat{\theta}_i - \hat{\theta}_J)^2 \quad (3.20)$$

and conjectured that $t = (\hat{\theta}_J - \theta) v_J^{-1/2}$ is approximately distributed as a t -variable with $m-1$ degrees of freedom, and hence also suggested the use of t as a pivotal statistics for getting confidence intervals on θ . The special case 'delete-one' jackknife (i.e., $k=1, m=n$) is commonly used.

However, when the observations are not i.i.d., (for example, in the case of our small area models and in a standard regression model), the extension of the above procedure is not immediate. Miller (1974), Hinkley (1977) and Wu (1984) extended the jackknife method to the standard regression model $y = X\beta + e$, where $E(e) = 0$. Suppose that the parameter of interest is $\theta = g(\beta)$, and that $\hat{\beta}$ denotes the ordinary least square estimator of β . The estimator $\hat{\beta}$ can be represented as

$$(1 - \frac{k}{n}) \hat{\beta} = \sum_{i=1}^n w_i \hat{\beta}(-i),$$

where $\hat{\beta}(-i)$ is the estimator obtained after omitting the i^{th} observation, and $w_i = 1 - x_i'(X'X)^{-1}x_i$, $X' = (x_1', \dots, x_n')$. Motivated by the above

representation, Wu (1984) proposed

$$v_{JW}(\hat{\beta}) = \sum_1 w_i (\hat{\beta}(-i) - \hat{\beta}) (\hat{\beta}(-i) - \hat{\beta})'$$

and

$$v_{JW}(\hat{\theta}) = \sum_1 w_i (\hat{\theta}(-i) - \hat{\theta}) (\hat{\theta}(-i) - \hat{\theta})'$$

as the jackknife variance estimators for $V(\hat{\beta})$ and $V(\hat{\theta})$ respectively,

where $\hat{\beta}(-i) = g(\hat{\beta}(-i))$. The jackknife variance estimator $v_{JW}(\hat{\beta})$ has

the following desirable properties: (i) If the errors are homoscedastic,

i.e., if the covariance matrix $V(e) = \sigma^2 I$, then $v_{JW}(\hat{\beta})$ is exactly

unbiased. (ii) $v_{JW}(\hat{\beta})$ is robust in the sense that it remains asymp-

totically consistent under violations of the homoscedastic assumption,

that is, when $V(e) = \text{diag}_{1 \leq i \leq n} (c_i^2)$ holds. Also motivated by the above

representation of $\hat{\beta}$ as a weighted linear combination of $\hat{\beta}(-i)$, Hinkley

(1977) constructed a variance-covariance estimator similar to (3.20), based on weighted pseudo values:

$$Q_i = \hat{\theta} + nw_i (\hat{\theta} - \hat{\theta}(-i)), \quad i=1, \dots, n.$$

In the linear case, $\hat{\theta} = \hat{\beta}$, the mean of weighted pseudo values is

exactly equal to $\hat{\beta}$, but the variance-covariance estimator does not

enjoy the property (i) mentioned above.

3.3 A Jackknife Procedure for Small Area Models

In this section, we describe a jackknife procedure in detail for the estimation of variance components under a nested error regression model.

However, for the other small area models we only give necessary expressions for the jackknife estimates.

a. The Nested Error Regression Model.

For simplicity, we confine ourselves to the model (2.1) with one auxiliary variable and an intercept term ($k=2$, $x_{ij1} = 1$, $x_{ij2} = x_{ij}$, $\beta_1 = \alpha$, $\beta_2 = \beta$). Before discussing our jackknife procedure, we need to modify the estimator $\hat{\sigma}_v^2$ in order to construct weighted pseudo-values, similar to Hinkley's (1977). The modified estimator is obtained along the lines of Arvesen (1969) as follows:

$$\hat{\sigma}_v^2 = n^{-1} \left[\sum_i \sum_j \hat{u}_{ij}^2 - \sum_i \frac{\lambda_i}{w_i} \sum_j \hat{e}_{ij}^2 \right], \quad (3.21)$$

where w_i and λ_i are as defined in (3.26) and (3.30) below,

$$\hat{e}_{ij} = y_{ij} - \bar{y}_{i.} - \beta^*(x_{ij} - \bar{x}_{i.}), \quad (3.22)$$

$$\hat{u}_{ij} = y_{ij} - \hat{\alpha} - \hat{\beta}x_{ij}, \quad (3.23)$$

and

$$\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 \beta^* = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})(x_{ij} - \bar{x}_{i.})$$

$$\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 \hat{\beta} = \sum_i \sum_j (y_{ij} - \bar{y}_{..})(x_{ij} - \bar{x}_{..})$$

$$\hat{\alpha} = \bar{y}_{..} - \bar{x}_{..} \hat{\beta}, \quad \bar{x}_{..} = n^{-1} \sum_i n_i \bar{x}_{i.}, \quad \bar{y}_{..} = n^{-1} \sum_i n_i \bar{y}_{i.}$$

The estimator $\hat{\sigma}_v^2$ is also unbiased which follows from the following results:

$$E\left[\sum_j \hat{e}_{ij}^2\right] = w_i \sigma_e^2, \quad (3.24)$$

$$E\left[\sum_j \hat{u}_{ij}^2\right] = n_i^* \sigma_v^2 + \lambda_i \sigma_e^2, \quad (3.25)$$

where

$$w_i = n_i - 1 - B_1 \sum_j (x_{ij} - \bar{x}_{i.})^2, \quad (3.26)$$

$$B_1 = \left(\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 \right)^{-1}, \quad (3.27)$$

$$\begin{aligned} n_1^* = n_1 - 2n^{-1}n_1^2 + n^{-1}n_1 \sum_i n_i^2 + B_2 \sum_j (x_{ij} - \bar{x}_{..})^2 \sum_i n_i^2 (\bar{x}_{i.} - \bar{x}_{..})^2 \\ + 2B_2^2 n_1 (\bar{x}_{i.} - \bar{x}_{..}) n^{-1} \sum_i n_i^2 (\bar{x}_{i.} - \bar{x}_{..})^2, \end{aligned} \quad (3.28)$$

with

$$B_2 = \left(\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 \right)^{-1}, \quad (3.29)$$

and

$$\lambda_i = n_i n^{-1} (n-1) - B_2 \sum_j (x_{ij} - \bar{x}_{..})^2. \quad (3.30)$$

Also, note that $\sum_i n_i^* = n_*$ (n_* is defined by (3.3)) and

$$\sum_i w_i = n - t - 1.$$

The estimator $\hat{\sigma}_e^2$ remains unchanged, but we denote it as $\hat{\sigma}_e^2$ to keep the notation consistent with $\hat{\sigma}_v^2$.

Our jackknife procedure is essentially based on omitting the residuals $\{e_{ij}, \hat{u}_{ij}; j=1, \dots, n_i\}$ corresponding to i^{th} small area and then computing the estimates of $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$ for $i=1, \dots, t$. The resulting unbiased estimators of $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$, denoted by $\hat{\sigma}_e^2(-i)$ and $\hat{\sigma}_v^2(-i)$, are given by

$$\hat{\sigma}_e^2(-i) = (n-t-1-w_i)^{-1} \left(\sum_j \hat{e}_{ij}^2 - \sum_j \hat{e}_{ij}^2 \right), \quad (3.31)$$

and

$$\hat{\sigma}_v^2(-i) = (n_*^* - n_i^*)^{-1} \left[\left(\sum_j \hat{u}_{ij}^2 - \sum_j \frac{\lambda_i}{w_i} \hat{e}_{ij}^2 \right) - \left(\sum_j \hat{u}_{ij}^2 - \frac{\lambda_i}{w_i} \sum_j \hat{e}_{ij}^2 \right) \right]. \quad (3.32)$$

The estimators $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$ both can be represented as weighted sums of $\hat{\sigma}_e^2(-i)$ and $\hat{\sigma}_v^2(-i)$ respectively, similar to Wu's representation for the standard regression model, as follows:

$$\hat{\sigma}_e^2 = (t-1)^{-1} \sum_i a_i \hat{\sigma}_e^2(-i), \quad (3.33)$$

and

$$\hat{\sigma}_v^2 = (t-1)^{-1} \sum_i b_i \hat{\sigma}_v^2(-i), \quad (3.34)$$

where

$$a_i = 1 - (n-t-1)^{-1} w_i, \quad b_i = 1 - n_i^{-1} n_i^*. \quad (3.35)$$

Following Hinkley (1977), we define weighted pseudo values for estimating $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$ respectively by

$$Q_i(e) = \hat{\sigma}_e^2 + t a_i (\hat{\sigma}_e^2 - \hat{\sigma}_e^2(-i)), \quad (3.36)$$

$$Q_i(v) = \hat{\sigma}_v^2 + t b_i (\hat{\sigma}_v^2 - \hat{\sigma}_v^2(-i)). \quad (3.37)$$

The jackknife estimators, $\bar{Q}(e) = t^{-1} \sum_{i=1}^t Q_i(e)$ and $\bar{Q}(v) = t^{-1} \sum_{i=1}^t Q_i(v)$, the mean of the weighted pseudo values, satisfy the reproducing property, i.e. $\bar{Q}(e) = \hat{\sigma}_e^2$, $\bar{Q}(v) = \hat{\sigma}_v^2$.

Turning to variance-covariance estimation, we propose the following, as in the standard jackknife procedure:

$$v_J(\hat{\sigma}_e^2) = (t(t-1))^{-1} \sum_i (Q_i(e) - \bar{Q}(e))^2 \quad (3.38)$$

$$= t(t-1)^{-1} \sum_i a_i^2 (\hat{\sigma}_e^2(-i) - \hat{\sigma}_e^2)^2$$

$$v_J(\hat{\sigma}_v^2) = t^{-1}(t-1)^{-1} \sum_i (Q_i(v) - \bar{Q}(v))^2 \quad (3.39)$$

$$= t(t-1)^{-1} \sum_i b_i^2 (\hat{\sigma}_v^2(-i) - \hat{\sigma}_v^2)^2$$

and

$$\begin{aligned} \text{cov}_J(\hat{\sigma}_v^2, \hat{\sigma}_e^2) &= t^{-1}(t-1)^{-1} \sum_i (Q_i(e) - \bar{Q}(e)) (Q_i(v) - \bar{Q}(v)) \\ &= t(t-1)^{-1} \sum_i a_i b_i (\sigma_e^2(-i) - \bar{\sigma}_e^2) (\sigma_v^2(-i) - \bar{\sigma}_v^2). \end{aligned} \quad (3.40)$$

Remark 3.1. The estimator defined by (3.40) was motivated by the fact that

$$2 \text{ cov}(\hat{\sigma}_v^2, \hat{\sigma}_e^2) = V(\hat{\sigma}_v^2 + \hat{\sigma}_e^2) - V(\hat{\sigma}_v^2) - V(\hat{\sigma}_e^2).$$

The asymptotic consistency of the jackknife variance estimators $v_J(\hat{\sigma}_e^2)$ and $v_J(\hat{\sigma}_v^2)$ defined by (3.38) and (3.39) is established below in Theorem 3.1.

Theorem 3.1. Assume the following regularity conditions for the Battese-Fuller model (model 1) with one auxiliary variable ($x_{ij1} \equiv 1$, $x_{ij2} = x_{ij}$, $\beta_1 = \alpha$, $\beta_2 = \beta$):

$$(C1) \quad \sup_{1 \leq i \leq t} E \left[\sum_j (e_{ij} - \bar{e}_{i.})^2 \right]^{2+\delta} < \infty \quad \text{for some } \delta > 0,$$

$$(C2) \quad \sup_{1 \leq i \leq t} E(\sum_j u_{ij}^2)^{2+\delta} < \infty \quad \text{for some } \delta > 0,$$

$$(C3) \quad \sup_{1 \leq i \leq t} \sum_j (x_{ij} - \bar{x}_{i.})^2 \leq d,$$

$$(C4) \quad c_1 \leq t^{-1} \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 \leq c_2 \quad \text{for all } t,$$

$$(C5) \quad \{n_i\}_{i=1}^t \text{ is such that } n_i \geq 2 \text{ for all } i, \text{ and } \sup_{1 \leq i \leq t} n_i < \infty,$$

where d , c_1 and c_2 are some positive constants. Then

$$t[v_J(\tilde{\sigma}_e^2) - v(\tilde{\sigma}_e^2)] \xrightarrow{p} 0 \quad (3.41)$$

and

$$t[v_J(\tilde{\sigma}_v^2) - v(\tilde{\sigma}_v^2)] \xrightarrow{p} 0. \quad (3.42)$$

The result (3.41) is established from lemmas 3.1 and 3.2, stated below. In Lemma 3.1 a result similar to (3.41) is established for

$$v_J(\bar{\sigma}_e^2) = t(t-1)^{-1} \sum_i a_i^2 (\bar{\sigma}_e^2(-i) - \bar{\sigma}_e^2)^2, \quad (3.43)$$

where w_i and λ_i are as defined in (3.26) and (3.30) below,

$$\bar{\sigma}_e^2 = (n-t-1)^{-1} \sum_i \sum_j \tilde{e}_{ij}^2, \quad (3.44)$$

$$\bar{\sigma}_e^2(-i) = (n-t-1-w_i)^{-1} [\sum_j \sum \tilde{e}_{ij}^2 - \sum_j \tilde{e}_{ij}^2] \quad (3.45)$$

and

$$\tilde{e}_{ij} = (y_{ij} - \bar{y}_{i.}) - \beta(x_{ij} - \bar{x}_{i.}) = e_{ij} - \bar{e}_{i.}. \quad (3.46)$$

Note that $\bar{\sigma}_e^2$, $\bar{\sigma}_e^2(-i)$ and $v_J(\bar{\sigma}_e^2)$ are obtained respectively by replacing β^* in $\tilde{\sigma}_e^2$, $\tilde{\sigma}_e^2(-i)$ and $v_J(\tilde{\sigma}_e^2)$ by its expected value. Also, the random variables $\sum_j \tilde{e}_{ij}^2$ are independent over i and as a result standard law of large numbers (Lemma 3.3) can be employed.

Lemma 3.1. Let $\bar{\sigma}_e^2$, $v_J(\bar{\sigma}_e^2)$ be defined as in (3.44) and (3.43). Then

$$t[v_J(\bar{\sigma}_e^2) - v(\bar{\sigma}_e^2)] \xrightarrow{p} 0. \quad (3.47)$$

Lemma 3.2 relates $\bar{\sigma}_e^2$ and $v_J(\bar{\sigma}_e^2)$ respectively to $\tilde{\sigma}_e^2$ and $v_J(\tilde{\sigma}_e^2)$.

Lemma 3.2. Let the regularity conditions (C1) and (C3 - C5) of Theorem 3.1 be true. Then

$$(a) \quad \sigma_e^2 = \bar{\sigma}_e^2 + O_p(t^{-1})$$

$$(b) \quad v(\sigma_e^2) = v(\bar{\sigma}_e^2) + O(t^{-2})$$

$$(c) \quad v_J(\sigma_e^2) = v_J(\bar{\sigma}_e^2) + O_p(t^{-3/2})$$

$$(d) \quad Ev_J(\bar{\sigma}_e^2) = v(\bar{\sigma}_e^2) + O(t^{-2})$$

The proof of Lemma 3.2 is somewhat lengthy and hence it is delegated to Appendix A.1.

Lemma 3.3. (Loève (1963), Theorem B(i), p. 275).

Let $S_n = \sum_{i=1}^n \xi_i$ be a sequence of independent random variables centered at their expectations. If $n^{-(1+\delta)} \sum_i E|\xi_i|^{1+\delta} \rightarrow 0$, for a positive $\delta \leq 1$, then $n^{-1} S_n \xrightarrow{p} 0$.

Proof of Lemma 3.1. It can be shown, on simplification, that

$$v_J(\bar{\sigma}_e^2) = t(t-1)^{-1}(n-t-1)^{-2} \left[\left(\sum_{i,j} \tilde{e}_{ij}^2 \right)^2 - 2 \left(\sum_i w_i \sum_j \tilde{e}_{ij}^2 \right) \bar{\sigma}_e^2 + \left(\sum_i w_i^2 \right) \bar{\sigma}_e^4 \right]$$

Hence, to establish (3.47) it is sufficient to show that

$$\begin{aligned} t^{-1} \sum_{i,j} \tilde{e}_{ij}^2 &= 2t^{-1} \bar{\sigma}_e^2 \sum_i w_i \sum_j \tilde{e}_{ij}^2 + \bar{\sigma}_e^4 t^{-1} \sum_i w_i^2 \\ &= (n-t-1)^2 t^{-1} v(\bar{\sigma}_e^2) \xrightarrow{p} 0. \end{aligned} \quad (3.48)$$

since $t(n-t-1)^{-1}$ is bounded (cf. condition C5). Now applying Lemma 3.3 to the sequences: $\{(\sum_j \tilde{e}_{ij}^2)^2 = E(\sum_j e_{ij}^2)^2\}$, $\{\sum_j \tilde{e}_{ij}^2 - E \sum_j e_{ij}^2\}$ and

$\{w_i \sum_j \tilde{e}_{ij}^2 - w_i E \sum_j \tilde{e}_{ij}^2\}$ we get

$$t^{-1} \sum_{i,j} (\sum_j \tilde{e}_{ij}^2)^2 - (n-t-1)^2 t^{-1} v(\bar{\sigma}_e^2) - t^{-1} \sum_i w_i^2 \sigma_e^4 \xrightarrow{p} 0, \quad (3.49)$$

$$(\bar{\sigma}_e^2 - \sigma_e^2) \xrightarrow{p} 0, \quad (3.50)$$

and

$$t^{-1} (\sum_i w_i \sum_j \tilde{e}_{ij}^2 - \sigma_e^2 \sum_i w_i^2) \xrightarrow{p} 0, \quad (3.51)$$

leading to the result (3.48).

Proof of Theorem 3.1. The result (3.41) follows immediately by noting that

$$\begin{aligned} t[v_J(\bar{\sigma}_e^2) - v(\bar{\sigma}_e^2)] &= t[v_J(\bar{\sigma}_e^2) - v_J(\bar{\sigma}_e^2)] \\ &\quad + t[v_J(\bar{\sigma}_e^2) - v(\bar{\sigma}_e^2)] + t[v(\bar{\sigma}_e^2) - v(\bar{\sigma}_e^2)], \end{aligned}$$

and then applying Lemmas 3.1 and 3.2 to the terms on the right.

Turning to the proof of (3.42), we only sketch the essential steps:

Step 1. A result similar to (3.47) is obtained by defining,

$$\bar{\sigma}_v^2 = n_*^{-1} (\sum_{i,j} u_{ij}^2 - \sum_i \frac{\lambda_i}{w_i} \sum_j \tilde{e}_{ij}^2),$$

$$\bar{\sigma}_v^2(-i) = (n_* - n_i^*)^{-1} (n_* \bar{\sigma}_v^2 - \sum_j u_{ij}^2 + \frac{\lambda_i}{w_{ij}} \sum_j \tilde{e}_{ij}^2),$$

and

$$v_J(\bar{\sigma}_V^2) = t(t-1)^{-1} \sum_i (\bar{\sigma}_V^2(-i) - \bar{\sigma}_V^2)^2 b_1^2.$$

That is, we first obtain

$$t(v_J(\bar{\sigma}_V^2) - v(\bar{\sigma}_V^2)) \xrightarrow{P} 0. \quad (3.52)$$

Here also note that $\bar{\sigma}_V^2$, $\bar{\sigma}_V^2(-i)$ and $v_J(\bar{\sigma}_V^2)$ are obtained by replacing b , a and β^* , in the expressions for $\bar{\sigma}_V^2$, $\bar{\sigma}_V^2(-i)$ and $v_J(\bar{\sigma}_V^2)$ by their expected values respectively. The proof of (3.52) follows along the lines of proof of (3.47), utilizing the additional condition (C2). Here we apply Lemma 3.3 to the sequences $\{T_1^2\}$, $\{T_i\}$ and $\{n_i^* T_i\}$, where

$$T_i = \sum_j u_{ij}^2 - \frac{1}{w_i} \sum_j \tilde{e}_{ij}^2$$

are independent random variables.

Step 2. We obtain results similar to Lemma 3.2, viz:

$$v(\tilde{\sigma}_V^2) = v(\bar{\sigma}_V^2) + O(t^{-2}) \quad (3.53)$$

and

$$v_J(\tilde{\sigma}_V^2) = v_J(\bar{\sigma}_V^2) + O_P(t^{-3/2}). \quad (3.54)$$

Step 3. Finally, the asymptotic consistency of $v_J(\tilde{\sigma}_V^2)$ i.e., the result (3.42) follows from the results (3.52) - (3.54).

Turning to the jackknife covariance estimator $\text{cov}_J(\tilde{\sigma}_e^2, \tilde{\sigma}_V^2)$ defined by (3.40), its asymptotic consistency is obtained below, as a Corollary to Theorem 3.1.

Corollary 3.1. Under the regularity conditions of Theorem 3.1,

$$t[\text{cov}_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) - \text{Cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)] \xrightarrow{p} 0. \quad (3.55)$$

Proof of Corollary 3.1. In view of Remark 3.1, to establish (3.55) it is sufficient to prove the result

$$t[v_J(\tilde{\sigma}_e^2 + \tilde{\sigma}_v^2) - v(\tilde{\sigma}_e^2 + \tilde{\sigma}_v^2)] \xrightarrow{p} 0 \quad (3.56)$$

which follows along the lines of (3.42) by noting that $n_*(n-t-1)^{-1}$ is bounded and

$$\tilde{\sigma}_e^2 + \tilde{\sigma}_v^2 = n_*^{-1} \left[\sum_i \sum_j \hat{u}_{ij}^2 - \sum_i \left(\frac{\lambda_i}{w_i} + \frac{n_*}{n-t-1} \right) \sum_j \hat{e}_{ij}^2 \right].$$

Remark 3.2. Since $\tilde{\sigma}_e^2$ and $\tilde{\sigma}_v^2$ are represented as weighted sums of $\hat{\sigma}_e^2(-i)$ and $\hat{\sigma}_v^2(-i)$ respectively (see (3.33) and (3.34)), we can get alternative jackknife variance estimators, following Wu (1984):

$$v_J^*(\tilde{\sigma}_e^2) = \sum_i (1-w_i(n-t-1)^{-1}) (\hat{\sigma}_e^2(-i) - \tilde{\sigma}_e^2)^2$$

and

$$v_J^*(\tilde{\sigma}_v^2) = \sum_i (1-n_i^* n_*^{-1}) (\hat{\sigma}_v^2(-i) - \tilde{\sigma}_v^2)^2.$$

The asymptotic consistency of these jackknife variance estimators follows by noting that

$$v_J^*(\tilde{\sigma}_e^2) = v_J(\tilde{\sigma}_e^2) (1 + O_p(t^{-1})),$$

and

$$v_J^*(\tilde{\sigma}_v^2) = v_J(\tilde{\sigma}_v^2) (1 + O_p(t^{-1})).$$

However, the optimal choice of a and b and the comparison of these alternative estimators with $v_J(\hat{\sigma}_e^2)$ requires further study.

b. Random Regression Model

For the random regression model (2.7) we again only modify the estimator $\hat{\sigma}_v^2$ given by (3.15). The modified estimator is

$$\hat{\sigma}_v^2 = \hat{n}_*^{-1} \sum_i \sum_j \hat{u}_{ij}^2 - \sum_i \tilde{\lambda}_i (n_i - 1)^{-1} \sum_j \hat{e}_{ij}^2,$$

where u_{ij} , e_{ij} and $\hat{n}_*$ are as defined in Section 3.1b and

$\tilde{\lambda}_i = n_i - \sum_{j=1}^{n_i} x_{ij}^2 (\sum_i \sum_j x_{ij}^2)^{-1}$. Our jackknife variance estimators are given by

$$v_J(\hat{\sigma}_e^2) = t(t-1)^{-1} \sum_i \{1 - (n_i - 1)(n-t)^{-1}\}^2 (\hat{\sigma}_e^2(-i) - \hat{\sigma}_e^2)^2$$

and

$$v_J(\hat{\sigma}_v^2) = t(t-1)^{-1} \sum_i \{1 - \hat{n}_i^* \hat{n}_*^{-1}\}^2 (\hat{\sigma}_v^2(-i) - \hat{\sigma}_v^2)^2,$$

where

$$\hat{\sigma}_e^2(-i) = (n-t-(n_i-1))^{-1} [(n-t)\hat{\sigma}_e^2 - \sum_j \hat{e}_{ij}^2]$$

and

$$\hat{\sigma}_v^2(-i) = (\hat{n}_* - \hat{n}_i^*)^{-1} [\hat{n}_* \hat{\sigma}_v^2 - (\sum_j \hat{u}_{ij}^2 - \tilde{\lambda}_i (n_i - 1)^{-1} \sum_j \hat{e}_{ij}^2)],$$

with

$$\hat{n}_i^* = \sum_j x_{ij}^2 - (\sum_j x_{ij}^2)^2 (\sum_i \sum_j x_{ij}^2)^{-1}, \text{ and } \hat{\sigma}_e^2 \text{ is the same as } \hat{\sigma}_e^2.$$

c. Fay-Herriot Model.

We use the estimator $\hat{A}$, given by (3.16), without modification.

In the case of a single auxiliary variable, $\hat{A}$ reduces to

$$\hat{A} = (t-1)^{-1} \left[\sum_i \hat{u}_i^2 - \sum_i D_i (1 - x_i^2 (\sum_i x_i^2)^{-1}) \right],$$

where

$$\hat{u}_i = \bar{y}_i - \bar{x}_i \hat{\beta}; \quad \hat{\beta} = (\sum_i x_i \bar{y}_i) (\sum_i x_i^2)^{-1}$$

Our jackknife variance estimator of $\hat{A}$ is given by

$$v_J(\hat{A}) = t(t-1)^{-1} \sum_{i=1}^t (\hat{A}(-i) - \hat{A})^2,$$

where

$$\hat{A}(-i) = \{t-1 - (1 - x_i^2 (\sum_i x_i^2)^{-1})\}^{-1} \{(t-1)\hat{A} - z_i\}$$

with

$$z_i = \hat{u}_i^2 - D_i - x_i^2 (\sum_i x_i^2 D_i) (\sum_i x_i^2)^{-2} + 2D_i x_i^2 (\sum_i x_i^2)^{-1}$$

3.4. The Jackknife for Non-Linear Functions $g(\theta)$

In many practical situations, we are interested in estimating a non-linear function of the variance components θ , say $g(\theta)$, where $g(\cdot)$ is a non-linear function. One such immediate application is the estimation of the non-linear parameter γ_i in the BLUP for small area models, (see Chapter 4; Section 4.1). In this section, a jackknife estimator of $g(\theta)$ and the corresponding jackknife variance estimator are obtained.

To simplify the discussion, we consider only the nested error regression model with one auxiliary variable ($x_{ij1} \equiv 1$, $x_{ij2} = x_{ij}$, $\beta_1 = \alpha$ and $\beta_2 = \beta$). For this model, our jackknife estimator of $g(\sigma_e^2, \sigma_v^2)$ is given by

$$g_J(\sigma_e^2, \sigma_v^2) = t^{-1} \sum_{i=1}^t Q_i(g); \quad (3.57)$$

where

$$Q_i(g) = g(\sigma_e^2, \sigma_v^2) - t \{ g[\sigma_e^2 + a_i(\tilde{\sigma}_e^2(-i) - \sigma_e^2), \sigma_v^2 + b_i(\tilde{\sigma}_v^2(-i) - \sigma_v^2)] - g(\sigma_e^2, \sigma_v^2) \} \quad (3.58)$$

are the weighted pseudo values and $\tilde{\sigma}_e^2$, $\tilde{\sigma}_v^2$, $\tilde{\sigma}_e^2(-i)$, $\tilde{\sigma}_v^2(-i)$, a_i and b_i are as defined in the previous section.

The motivation for the above choice of pseudo values $Q_i(g)$ is two-fold:

(i) When the function $g(\cdot, \cdot)$ is linear in σ_e^2 and σ_v^2 , i.e.

$$g(\sigma_e^2, \sigma_v^2) = \ell_1 \sigma_e^2 + \ell_2 \sigma_v^2, \text{ for some known constants } \ell_1 \text{ and } \ell_2,$$

then the jackknife estimator $g_J(\sigma_e^2, \sigma_v^2)$ reduces to the unbiased estimator $\ell_1 \sigma_e^2 + \ell_2 \sigma_v^2$.

(ii) If $g(\cdot, \cdot)$ is non-linear, then the jackknife estimator has a bias of lower order than $g(\sigma_e^2, \sigma_v^2)$, (see Theorem 3.2 given below).

The jackknife variance estimator based on the pseudo values $Q_i(g)$ is given by

$$\begin{aligned} v_J(g) &= t^{-1}(t-1)^{-1} \sum_{i=1}^t (Q_i(g) - g_J(\sigma_e^2, \sigma_v^2))^2 \\ &= t(t-1)^{-1} \sum_{i=1}^t [g(\sigma_e^2 + \tilde{\sigma}_e^2(i), \sigma_v^2 + \tilde{\sigma}_v^2(i)) - g(\sigma_e^2, \sigma_v^2)]^2, \end{aligned} \quad (3.59)$$

where

$$\tilde{\sigma}_1^2(i) = a_i(\tilde{\sigma}_e^2(-i) - \tilde{\sigma}_e^2) \quad (3.60)$$

and

$$\tilde{\sigma}_2^2(i) = b_i(\tilde{\sigma}_v^2 - \tilde{\sigma}_v^2) \quad (3.61)$$

The asymptotic consistency of $v_J(g)$ is also established in Theorem 3.2.

Theorem 3.2. Let $g(\cdot, \cdot)$ be a scalar function having third order continuous derivatives around (σ_e^2, σ_v^2) . Then under the regularity conditions of Theorem 3.1, for the nested error regression model with one auxiliary variable,

$$Eg_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) - g(\sigma_e^2, \sigma_v^2) = o(t^{-1}) \quad (3.62)$$

and

$$t[v_J(g) - v_g] \xrightarrow{p} 0 \quad (3.63)$$

where $v_g = (\nabla g)' V(\tilde{\sigma}^2) (\nabla g)$ is the asymptotic variance of $g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$, with (∇g) denoting the vector of partial derivatives of $g(\sigma_e^2, \sigma_v^2)$ with respect to σ_e^2 and σ_v^2 , and $V(\tilde{\sigma}^2)$ denoting the covariance matrix of $\tilde{\sigma}^2 = (\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$.

Proof of Theorem 3.2. The Taylor's expansions of $g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ and

$g(\tilde{\sigma}_e^2 + \tilde{\sigma}_1^2(i), \tilde{\sigma}_v^2 + \tilde{\sigma}_2^2(i))$ respectively about the points (σ_e^2, σ_v^2) and $(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ leads to

$$\begin{aligned} g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) - g(\sigma_e^2, \sigma_v^2) &= (\tilde{\sigma}^2 - \sigma^2)' (\nabla g) + 2^{-1} (\tilde{\sigma}^2 - \sigma^2)' \nabla^2 g(\sigma_e^2, \sigma_v^2) (\tilde{\sigma}^2 - \sigma^2) \\ &\quad + O_p(t^{-3/2}) \end{aligned} \quad (3.64)$$

and

$$\begin{aligned} & \sum_{i=1}^t \{g(\tilde{\sigma}_e^2 + \sigma_1^2(i), \tilde{\sigma}_v^2 + \tilde{\sigma}_2^2(i)) - g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)\} \\ &= 2^{-1} \sum_{i=1}^t \{(\tilde{\sigma}_1^2(i), \tilde{\sigma}_2^2(i)) (\nabla^2 g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)) (\tilde{\sigma}_1^2(i), \tilde{\sigma}_2^2(i))'\} + 6^{-1} \sum_{i=1}^t \eta(i), \quad (3.65) \end{aligned}$$

where $\nabla^2 g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ denotes the matrix of Second order derivatives of $g(\sigma_e^2, \sigma_v^2)$ with respect to (σ_e^2, σ_v^2) evaluated at $(\sigma_e^2, \sigma_v^2) = (\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ and

$$\eta(i) = \sum_{\ell=1}^2 \sum_{k=1}^2 \sum_{m=1}^2 \tilde{\sigma}_\ell^2(i) \tilde{\sigma}_k^2(i) \tilde{\sigma}_m^2(i) \nabla_{\ell k m}^3 g(\xi_1(i), \xi_2(i)) \quad (3.66)$$

with $\nabla_{\ell k m}^3 g(\xi_1(i), \xi_2(i))$ denoting third order derivatives of $g(\sigma_e^2, \sigma_v^2)$ evaluated at $\xi_1(i) = \tilde{\sigma}_e^2 + \lambda \tilde{\sigma}_1^2(i)$ and $\xi_2(i) = \tilde{\sigma}_v^2 + \mu \tilde{\sigma}_2^2(i)$; $0 \leq \lambda, \mu \leq 1$.

Now, using (3.64), (3.65) and noting that $\sum_i \eta(i) = O_p(t^{-3/2})$ (proved in Lemma 3.4) we rewrite $g_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ as

$$\begin{aligned} g_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) - g(\sigma_e^2, \sigma_v^2) &= (\tilde{\sigma}^2 - \sigma^2)' (\nabla g) + 2^{-1} (\tilde{\sigma}^2 - \sigma^2)' (\nabla^2 g(\sigma_e^2, \sigma_v^2)) (\tilde{\sigma}^2 - \sigma^2) \\ &\quad - 2^{-1} \text{tr}[(\nabla^2 g(\sigma_e^2, \sigma_v^2)) v_J(\tilde{\sigma}^2)] + O_p(t^{-3/2}), \quad (3.67) \end{aligned}$$

where

$$v_J(\tilde{\sigma}^2) = \begin{bmatrix} v_J(\tilde{\sigma}_e^2) & \text{cov}_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) \\ \text{cov}_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) & v_J(\tilde{\sigma}_v^2) \end{bmatrix}$$

with $v_J(\tilde{\sigma}_e^2)$, $v_J(\tilde{\sigma}_v^2)$ and $\text{cov}_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ as defined by (3.38) - (3.40). Hence, the result (3.62) follows from (3.67), by using the result $EO_p(t^{-3/2}) = O(t^{-2})$ (for which the required conditions e.g. Theorem 5.4.3 of Fuller (1976), have been verified) and noting that

$$E v_J(\tilde{\sigma}_e^2) = v(\tilde{\sigma}_e^2) + O(t^{-2})$$

$$E v_J(\tilde{\sigma}_v^2) = v(\tilde{\sigma}_v^2) + O(t^{-2})$$

and hence

$$E \text{cov}_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) = \text{cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) + O(t^{-2})$$

The latter result is easily obtained from Lemma 3.2.

Turning to the proof of (3.63), we again use the Taylor expansion for $g(\tilde{\sigma}_e^2 + \tilde{\sigma}_1^2(i), \tilde{\sigma}_v^2 + \tilde{\sigma}_2^2(i))$ about the point $(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ to get

$$\begin{aligned} v_J(g) &= t(t-1)^{-1} \sum_{i=1}^t \{ (\tilde{\sigma}_1^2(i), \tilde{\sigma}_2^2(i)) (\nabla g) \\ &\quad + \frac{1}{2} (\tilde{\sigma}_1^2(i), \tilde{\sigma}_2^2(i))' (\nabla^2 g(\xi_1(i), \xi_2(i))) (\tilde{\sigma}_1^2(i), \tilde{\sigma}_2^2(i))' \}^2 \end{aligned} \quad (3.68)$$

where $\xi_1(i)$ and $\xi_2(i)$ are as defined above. We rewrite (3.68)

(after simplification using (3.38) - (3.40) and (3.60) and (3.61)) as

$$v_J(g) = (\nabla g)' v_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) (\nabla g) + t(t-1)^{-1} \sum_i \eta^*(i), \quad (3.69)$$

where $\eta^*(i)$ is the sum of squares of the second term and the cross-product terms in $\{ \}$ of (3.68). The desired result (3.63) now follows from Theorem 3.1 and its corollary 3.1, provided $\sum_{i=1}^t \eta^*(i) = O_p(t^{-3/2})$.

The proof of the latter result follows along the lines of Lemma 3.4.

Lemma 3.4. Under the conditions of Theorem 3.2,

$$\sum_{i=1}^t \eta(i) = O_p(t^{-3/2})$$

where $\eta(i)$ is defined by (3.66),

Proof of Lemma 3.4. The proof of Lemma 3.4 is along the lines of Lemma 5 of Wu (1984).

Since $V^3_{\ell km} g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ is bounded in a neighborhood of (σ_e^2, σ_v^2) , it can be shown that,

$$\left| \sum_i \eta(i) \right| < M \sum_{i=1}^t \sum_{\ell=1}^2 \sum_{k=1}^2 \sum_{m=1}^2 |\tilde{\sigma}_\ell^2(i) \tilde{\sigma}_k^2(i) \tilde{\sigma}_m^2(i)|, \quad (3.70)$$

where $M < \infty$ and independent of t . Also, following the proof of Lemma 3.3 of Miller (1974), it can be shown that

$$\sup_{1 \leq i \leq t} |\tilde{\sigma}_\ell^2(i)| = O_p(t^{-1/2}); \quad \ell = 1, 2.$$

Using the above result in (3.70), we get

$$\left| \sum_i \eta(i) \right| < O_p(t^{-1/2}) \sum_{i=1}^t \sum_{\ell=1}^2 \sum_{k=1}^2 |\tilde{\sigma}_\ell^2(i) \tilde{\sigma}_k^2(i)|. \quad (3.71)$$

Now, for $\ell = k$,

$$\sum_{i=1}^t \sum_{\ell=1}^2 (\tilde{\sigma}_\ell^2(i))^2 = v_J(\tilde{\sigma}_e^2) + v_J(\tilde{\sigma}_v^2) = O_p(t^{-1}), \quad (3.72)$$

and for $\ell \neq k$,

$$\sum_{i=1}^t \sum_{\ell \neq k=1}^2 |\tilde{\sigma}_\ell^2(i) \tilde{\sigma}_k^2(i)| = O_p(t^{-1}),$$

which follows from (3.72) and the Cauchy-Schwarz inequality. Hence

$$\left| \sum_{i=1}^t \eta(i) \right| = O_p(t^{-3/2}). \quad \square$$

Remark 3.3. Since different weights are used in representing $\tilde{\sigma}_e^2$ and $\tilde{\sigma}_v^2$ respectively as weighted linear combinations of $\tilde{\sigma}_e^2(-i)$ and

$\tilde{\sigma}_v^2(-i)$ (see (3.33) and (3.34)). The method of obtaining jackknife variance estimators for $g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ along the lines of Wu (1984) (see Remark 3.2 and Section 3.2)) is not clear at this stage.

Turning to the use of our jackknife procedure in constructing a confidence interval on $g(\sigma_e^2, \sigma_v^2)$, a theorem similar to Theorem 1 of Miller (1974) can be established. We are only stating the result below.

Theorem 3.3. Under the regularity conditions of Theorem 3.2,

$$\left[t^{1/2} (g_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) - g(\sigma_e^2, \sigma_v^2)) / (v_g)^{1/2} \right] \xrightarrow{d} N(0,1),$$

where v_g is defined in Theorem 3.2 (see text following (3.63)).

As a consequence of Theorems 3.2 and 3.3, our jackknife procedure can be used to construct confidence intervals on $g(\sigma_e^2, \sigma_v^2)$.

3.5. Monte Carlo Study

In this section we report findings from a Monte Carlo study of the finite sample performance of our jackknife confidence intervals on a non-linear parametric function, under the nested error regression model. Though the parameters of primary interest are the intraclass correlation coefficient $\rho = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2)$ and the variance ratio σ_v^2 / σ_e^2 , we have instead followed Arvesen and Schmitz (1970) and chose to work with the function $\theta = g(\sigma_e^2, \sigma_v^2) = \log(1 + n_* \sigma_v^2 / \sigma_e^2)$, where n_* is given by (3.3). Confidence interval on ρ or the variance ratio can be readily obtained from intervals on $g(\sigma_e^2, \sigma_v^2)$ using the invariance property of confidence

intervals. As demonstrated by Arvesen and Schmitz (1970), confidence intervals on a logarithmic function tend to approximate their predictive behaviour more closely than they do with confidence intervals based directly on σ_v^2/σ_e^2 .

The jackknife confidence interval on $\hat{\theta} = g(\sigma_e^2, \sigma_v^2)$ is given by

$$[\hat{\theta}_J - t_{\alpha/2: (t-1)} (v_J(\hat{\theta}))^{1/2}, \hat{\theta}_J + t_{\alpha/2: (t-1)} (v_J(\hat{\theta}))^{1/2}], \quad (3.73)$$

where $t_{\alpha/2: (t-1)}$ is the upper α -point of the t -distribution with $(t-1)$ degrees of freedom.

We also consider an interval derived under the assumption of normality, given by

$$[\hat{\theta} - t_{\alpha/2: (t-1)} (v_N(\hat{\theta}))^{1/2}, \hat{\theta} + t_{\alpha/2: (t-1)} (v_N(\hat{\theta}))^{1/2}]. \quad (3.74)$$

Here

$$v_N(\hat{\theta}) = \{(\nabla g)' v_N(\hat{\sigma}^2) (\nabla g)\} / \hat{\sigma}^2,$$

where

$$v_N(\hat{\sigma}^2) = \begin{bmatrix} v(\hat{\sigma}_e^2) & \text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2) \\ \text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2) & v(\hat{\sigma}_v^2) \end{bmatrix}, \quad (3.75)$$

$$v(\hat{\sigma}_e^2) = 2\sigma_e^4 / (n-t-1), \quad (3.76)$$

$$v(\hat{\sigma}_v^2) = 2n_*^{-2} \left[\sigma_e^4 \left\{ (t-1) + 2 \sum_{i=1}^t \frac{(w_i - \bar{w})^2}{w_i} \right\} + 2n_* \sigma_e^2 \sigma_v^2 + n_* \sigma_v^4 \right], \quad (3.77)$$

and

$$\text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2) = - (t-1) n_*^{-1} v(\hat{\sigma}_e^2), \quad (3.78)$$

with λ_i , w_i and n_i as defined respectively by (3.30), (3.26) and (3.3).

Our motivation for proposing the jackknife procedure is that it should perform at least as well as the normality based procedure when the data are normal, and that it should be robust to deviation from normality. To examine these conjectures, we have computed the following three measures for comparing the properties of (i) confidence intervals based on the jackknife estimators $\hat{\theta}_J$ and $v_J(\hat{\theta})$, and based on the estimators $\hat{\theta}$ and $v_N(\hat{\theta})$; (ii) $\hat{\theta}$, $v_N(\hat{\theta})$ and $\hat{\theta}_J$ and $v_J(\hat{\theta})$:

- (a) Percentage coverage rates P_J and P_N of the $(1-\alpha)$ confidence intervals defined respectively by (3.73) and (3.74).
 (b) Percentage relative bias of the estimators of $\theta = g(\sigma_e^2, \sigma_v^2)$:
 $R.B.(\hat{\theta}_J), R.B.(\hat{\theta})$, where

$$R.B.(\hat{\theta}_J) = 100\{(E_S \hat{\theta}_J - \theta)/\theta\}$$

and

$$R.B.(\hat{\theta}) = 100\{(E_S \hat{\theta} - \theta)/\theta\}$$

with E_S denoting Monte Carlo expectation.

- (c) Percentage relative bias of the variance estimators:

$R.B.(v_J(\hat{\theta})), R.B.(v_N(\hat{\theta}))$, where

$$R.B.(v_J(\hat{\theta})) = 100\{(E_S v_J(\hat{\theta}) - v_S(\hat{\theta}))/v_S(\hat{\theta})\}$$

and

$$R.B.(v_N(\hat{\theta})) = 100\{(E_S v_N(\hat{\theta}) - v_S(\hat{\theta}))/v_S(\hat{\theta})\}$$

with v_S denoting Monte Carlo variance.

For our Monte Carlo study, we have considered the following nested error regression model:

$$y_{ij} = 5.5 + 0.388x_{ij} + v_i + e_{ij}, \quad (3.79)$$

where $i = 1, \dots, 10$; $j = 1, \dots, n_i$; with n_i and x_{ij} values as reported in Table A (See Appendix A.2). A detailed description of this model can be found in Section 4.3, in the context of the estimation of mean square errors of small area predictors. In the Monte Carlo experiment 1000 sets of observations $\{y_{ij}; i=1, \dots, 10; j=1, \dots, n_i\}$ were obtained by generating 1000 sets of independent pseudo random deviates $\{v_i\}$ and $\{e_{ij}\}$, $i=1, \dots, 10$; $j=1, \dots, n_i$ according to the following four distributions: the standard normal, standard exponential, standard double exponential and the uniform (0,1). The random deviates $\{v_i\}$ and $\{e_{ij}\}$ thus generated were transformed to give deviates having mean zero and specified variances σ_v^2 and σ_e^2 respectively. The value of σ_e^2 was set equal to 1 for all the Monte Carlo runs, while σ_v^2 was varied to take the values 1, 2.25 and 4.0. From the centered deviates $\{v_i\}$ and $\{e_{ij}\}$, the observations $\{y_{ij}\}$ were obtained from the model (3.79). The main findings of the Monte Carlo experiment utilizing the measures defined in (a), (b) and (c) are as follows: (see Table 3.1a, p. 54).

Table 3.1a reports the coverage rates for the confidence intervals defined by (3.73) and (3.74), for nominal levels 95% (90%), the percentage biases of the estimators of θ and of their variance estimators.

Coverage Rates: From Table 3.1a, it can be observed that normality based intervals defined by (3.74) provide coverage probabilities close to the nominal levels only for normally distributed data. Coverage rates fall significantly short of the nominal levels under exponential

Table 3.1a: Coverage rates of the confidence intervals for the nominal level 95% (90%) and percentage relative biases of θ , $\hat{\theta}_J$, $v_N(\hat{\theta})$ and $v_J(\tau)$ (with $t = 10$)

Distribution	σ_{v/τ_e}	P_N	P_J	R.B. ($\hat{\theta}$)	R.B. ($\hat{\theta}_J$)	R.B. ($v_N(\hat{\theta})$)	R.B. ($v_J(\tau)$)
Normal	1.0	93(89)	89(87)	-3.4	-1.4	-20.2	-21.0
	1.5	93(89)	89(87)	-3.4	-1.3	-20.2	-21.5
	2.0	93(89)	89(87)	-3.3	-1.1	-20.5	-22.3
Exponential	1.0	78(74)	85(81)	-8.4	-6.7	-49.2	-18.8
	1.5	78(74)	85(81)	-7.9	-5.3	-57.8	-20.8
	2.0	78(74)	85(81)	-6.5	-3.9	-62.3	-24.0
Double Exponential	1.0	85(80)	88(82)	-8.1	-6.7	-40.2	-18.0
	1.5	85(80)	88(82)	-7.0	-5.0	-51.7	-19.7
	2.0	85(80)	88(82)	-6.1	-3.7	-59.3	-23.0
Uniform	1.0	98(93)	93(88)	-3.2	-0.9	28.5	-9.8
	1.5	98(93)	93(88)	-1.9	-.24	32.3	-9.8
	2.0	98(93)	93(88)	-1.5	-.19	40.9	-8.8

Note: $\theta = \log(1 + n \sigma_{v/\tau_e}^2)$.

and double exponential distributions. Actual coverage rates are as low as 78% (74%) under exponential and 85% (80%) under double exponential distributions compared to nominal levels of 95% (90%). For the uniform distribution the trend is in the other direction, with coverage rates exceeding nominal levels 98% (93%) compared to nominal levels of 95% (90%). Turning to our jackknife interval defined by (3.73), we see that coverage rates under normally distributed data are lower than the nominal rate (89% and 87% versus 95% and 90% respectively). For exponential and double exponential distributions, the coverage rates are again low compared to nominal levels, though there is considerable improvement compared to the coverage rates of the normality based interval, e.g. 85% compared to 78% under exponential data, and 88% compared to 85% under double exponential data. For uniform data, the jackknife interval performs well yielding a coverage rate only two percentage points less than the nominal.

Remark 3.4: 1. Note that the Monte Carlo accuracy of P_N and P_J , which is obtained from the binomial standard error, is approximately equal to .007 (.01) for 95% (90%) nominal level.

2. The values reported in Table 3.1a corresponding to P_N and P_J are the values obtained by rounding the corresponding actual values to the nearest integers. There were small variations in the actual values of P_N and P_J corresponding to different ratios $C/\sqrt{n}$.

Bias of estimates $\hat{\theta}$ and $\hat{\theta}_J$: The values corresponding to columns of R.B. ($\hat{\theta}$) and R.B. ($\hat{\theta}_J$) in Table 3.1a clearly exhibits the bias

reduction property of our jackknife procedure under all distributions. However, magnitude of the relative bias (4) of $\hat{\theta}_J$ is still appreciable under exponential and double exponential distributions, the worst case being -6.2% under exponential distribution and -6.9% under double exponential distribution, when the true variance ratio is equal to one.

Bias of variance estimators $v_N(\hat{\theta})$ and $v_J(\hat{\theta})$: It is seen from Table 3.1a that both $v_N(\hat{\theta})$ and $v_J(\hat{\theta})$ underestimate the true variance $V(\theta)$ (i.e. Monte Carlo variance) under all distributions but uniform. In the latter case, $v_N(\hat{\theta})$ overestimate by as much as 41% compared to underestimate of about 10% by $v_J(\hat{\theta})$. For the exponential distribution $v_N(\hat{\theta})$ attains a maximum underestimate of 62%, while for the double exponential distribution case, the maximum underestimation of $v_N(\hat{\theta})$ is 59%. However, for $v_J(\hat{\theta})$, the corresponding underestimation results in lower relative biases of 24% and 23% respectively.

Effect of increasing t . In the experiment described above, t , the number of groups was set at 10. To examine the effect of increasing t on the performance of the two confidence intervals and on the properties of the estimators of θ and $V(\hat{\theta})$, a second Monte Carlo experiment was performed by generating another ten groups, having identical values of n_i and x_{ij} , and then obtaining new sets $\{y_{ij}; i=1, \dots, 20; j=1, \dots, n_i\}$ from the model (3.79). The quantities corresponding to Table 3.1a are reported in Table 3.1b (see page 57).

It is seen from Table 3.1b that when $t = 20$, the coverage probability rates of the jackknife intervals are much closer to their

Table 3.1b: Coverage rates of the confidence intervals for the nominal level 95%(90%) and percentage relative biases of θ , $\hat{\theta}_J$, $v_N(\hat{\theta})$ and $v_J(\hat{\theta})$ (with $t = 20$).

Distribution	$\sqrt{v}/\sigma$	P_N	P_J	R.B. (θ)	R.B. ($\hat{\theta}_J$)	R.B. ($v_N(\hat{\theta})$)	R.B. ($v_J(\hat{\theta})$)
Normal	1.0	95 (90)	93 (89)	-1.3	-0.3	-2.1	-3.0
	1.5	95 (90)	93 (89)	-0.9	-0.3	-2.1	-3.0
	2.0	95 (90)	93 (89)	-0.7	-0.1	-2.3	-2.4
Exponential	1.0	79 (75)	89 (84)	-5.2	-2.3	-57.6	-11.9
	1.5	78 (75)	89 (84)	-4.7	-2.0	-63.8	-11.0
	2.0	78 (75)	89 (84)	-4.3	-2.0	-72.3	-11.2
Double Exponential	1.0	88 (80)	92 (85)	-5.2	-2.3	-45.9	-11.0
	1.5	88 (80)	92 (85)	-4.2	-1.8	-51.4	-11.0
	2.0	88 (90)	92 (85)	-3.6	-1.5	-53.3	-10.5
Uniform	1.0	99 (93)	96 (91)	-1.8	-0.2	41.1	1.1
	1.5	99 (93)	96 (91)	-1.0	-0.0	66.6	2.8
	2.0	99 (93)	96 (91)	-0.7	-0.0	80.6	3.8

Note: $\sigma = \log(1 + n \cdot \frac{2}{v} \cdot \frac{2}{\sigma^2})$.

nominal level than when $t = 10$. For example, when $t = 20$, coverage rates are 89% (84%) as compared to 85% (78%) for $t = 10$, under exponential data, and 92% (85%) as compared to 88% (81%) under double exponential data. But the same is not true for the normality based intervals. In this case, the coverage rates are 79% (75%) for $t = 20$ compared to 78% (73%) for $t = 10$, under exponential distribution, and 88% (80%) for $t = 20$ compared to 85% (80%) for $t = 10$ under double exponential distribution. Thus the increase in t improves the performance of the jackknife interval much more than the normality based intervals. Table 3.1b also indicates that relative bias of the estimators $\hat{\theta}_J$ and $v_J(\hat{\theta})$ is considerably reduced when $t = 20$. The maximum relative bias of $\hat{\theta}_J$ with $t = 20$ occurs under exponential and double exponential distributions, corresponding to the case $\sigma_V/\sigma_e = 1.0$. In the worst case situation of exponential and double exponential data, the relative bias of $v_J(\hat{\theta})$ when $t = 20$ is approximately one half of its value when $t = 10$. For the more favourable cases of normal and uniform distributions, the improvement is even greater; under the normal distribution case, biases have been reduced from about -20% to -3% while for uniform distribution a low reduction in magnitude from about -10% to around 4%. However, for the estimators $\hat{\theta}$ and $v_N(\hat{\theta})$, high relative biases are still found even when $t = 20$. For $\hat{\theta}$, biases are low under normal and uniform distribution cases, of the order of -1.1%; under exponential and double exponential distributions, biases read -5%. For $v_N(\hat{\theta})$, biases are much greater, ranging from -72% to -81% for all distributions except the normal distribution.

In the latter case $v_N(\cdot)$ performs well, having a relative bias of only -2%.

In summary, it is seen that the jackknife estimators $\hat{\mu}_J$ and $\hat{v}_J(\cdot)$ track their true values reasonably well, though they do exhibit a tendency to underestimate except in the uniform distribution case. This underestimation of the variance results in a slight underestimation of coverage probability for the jackknife intervals. Increasing the number of groups, t , may overcome this problem.

CHAPTER 4

PREDICTION OF SMALL AREA MEANS WHEN THE VARIANCE COMPONENTS ARE UNKNOWN

All the results, Theorems 2.2 - 2.4, given in Chapter 2, are based on the assumption that the variance components θ are known. In practice, some or all of the variance components are unknown. The BLUPs are then estimated by replacing the unknown variance components by their asymptotically consistent estimators considered in Chapter 3. The estimated BLUPs $t_1(\hat{\theta}, y)$ are no longer unbiased.

Turning to the mean square error (MSE) of the estimated BLUP, Pixeto (1982) has obtained an exact expression for the MSE, only under a special case of model 1: $k=1$, $x_{ij1} \equiv 1$, $n_1=r$, along with the assumption of normality for errors y and e . However, in general (for example, unbalanced models), an exact expression for the MSE is not tractable. Hence, it is desirable to obtain an approximation to MSE by letting t , the number of small areas, tend to infinity (with n_1 's uniformly bounded). Along this direction an approximate expression for the MSE of the estimated BLUP $t_1(\hat{\theta}, y)$ is obtained in Section 4.1, under the normality assumption. The jackknife procedure developed in Sections 3.3 and 3.4 is also utilized in reducing the bias of the estimated BLUP. In Section 4.2 an estimate of the approximate MSE is obtained to the same order of approximation. Robust estimators, based on our jackknife procedure, to estimate approximate MSE are also provided. Section 4.3 reports the results of a simulation study on the accuracy of these approximations. The effect of relaxing the normality assumption is also studied in

Section 4.3. In this Chapter, the details of all the results are provided only for the nested error regression model. For the other two small area models only necessary expressions are given.

4.1. Approximate Mean Square Error

a. Nested-Error Regression Model.

The estimated BLUP and its approximate mean square error (MSE) under the nested error regression model (model 1) are given below in Theorem 4.1. The following regularity conditions are employed in establishing Theorem 4.1:

(C1) The elements of the matrix X are assumed to be uniformly bounded such that

$$(a) \quad \lim_{t \rightarrow \infty} t^{-1} (X' V^{-1} X) = G \quad \text{and} \quad G^{-1} \text{ exists}$$

$$(b) \quad \lim_{t \rightarrow \infty} t^{-1} (X' V^{-1} X) = S \quad \text{and} \quad S^{-1} \text{ exists almost surely}$$

and

$$(c) \quad \lim_{t \rightarrow \infty} t^{-1} X' \frac{\partial V^{-1}}{\partial \theta} X = H_{\theta}$$

where V is obtained from V by replacing θ by $\hat{\theta}$, and the elements of H_{θ} are continuous functions of θ .

(C2) The n_i 's are uniformly bounded.

Theorem 4.1. For the nested error regression model defined by (2.1) -

(2.3), let the regularity conditions C1 - C2 and normality assumptions for the random errors be true. Then the MSE of the estimated BLUP

$$t_i(\hat{\beta}^2, y) = \bar{x}_i' \hat{\beta} + \hat{y}_i (\bar{y}_i - \bar{x}_i' \hat{\beta}), \quad (4.1)$$

can be approximated as follows:

$$\begin{aligned} \text{MSE}(t_i(\sigma^2, y)) &= E\{t_i(\hat{\sigma}^2, y) - u_i\}^2 \\ &= E\{t_i(\hat{\sigma}^2, y) - u_i\}^2 + \\ &\quad n_i^{-2}(\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-3} \text{Var}(\hat{\sigma}_e^2 \sigma_v^2 - \hat{\sigma}_v^2 \sigma_e^2) + o(t^{-1}), \end{aligned} \quad (4.2)$$

where

$$t_i = \hat{\sigma}_v^2(\hat{\sigma}_v^2 + \hat{\sigma}_e^2 n_i^{-1})^{-1}, \quad (4.3)$$

$$E = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} Y, \quad (4.4)$$

with

$$\hat{V} = \hat{\sigma}_e^2 I + \hat{\sigma}_v^2 \text{diag} \left(\frac{1}{n_i}, \frac{1}{n_i} \right)_{1 \leq i \leq t}$$

and $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$, $E\{t_i(\hat{\sigma}^2, y) - u_i\}^2$ are as defined respectively by (3.1), (3.2) and (2.19).

The proof of Theorem 4.1 involves several lemmas given below.

Lemma 4.1. Let u be distributed as $(\sim)$ a p -variate normal with mean vector 0 and dispersion matrix I (i.e. $u \sim N_p(0, I)$) and A_i ($i=1, 2$) be non-stochastic symmetric matrices of order p , then

$$(i) \quad E[u(u' A_s u)u'] = (\text{tr } A_s)I + 2A_s; \quad s=1, 2,$$

$$(ii) \quad E\left[\prod_{s=1}^2 u' A_s u\right] = 2 \text{tr } A_1 A_2 + \text{tr } A_1 \text{tr } A_2$$

and

$$\begin{aligned} (iii) \quad E\left[u\left(\prod_{s=1}^2 u' A_s u\right)u'\right] &= (\text{tr } A_1)(\text{tr } A_2)I + 2(\text{tr } A_1)A_2 + 2(\text{tr } A_1 A_2)I \\ &\quad + 2(\text{tr } A_2)A_1 + 4A_1 A_2 + 4A_2 A_1, \end{aligned}$$

where 'tr' denotes the trace of a matrix.

Proof of Lemma 4.1. See Srivastava and Tiwari (1976).

Lemma 4.2. Let $u \sim N_p(0, \Sigma)$, where Σ is a positive definite matrix.

Also, let A_s ($s=1,2$), be non-stochastic matrices of order p . Then

$$(i) \quad E[u(u'A_s u)u'] = (\text{tr } A_s \Sigma) \Sigma + 2 \Sigma A_s \Sigma; \quad s=1,2,$$

$$(ii) \quad E\left[\sum_{s=1}^2 (u'A_s u)\right] = 2 \text{tr } A_1 \Sigma A_2 \Sigma + \text{tr } A_1 \Sigma \text{tr } A_2 \Sigma$$

and

$$(iii) \quad E\left[\sum_{s=1}^2 u(u'A_s u)u'\right] = (\text{tr } A_1 \Sigma)(\text{tr } A_2 \Sigma) \Sigma + 2(\text{tr } A_1 \Sigma) \Sigma A_2 \Sigma + 2(\text{tr } A_2 \Sigma) \Sigma A_1 \Sigma \\ + 2(\text{tr } A_1 \Sigma A_2 \Sigma) \Sigma + 4 \Sigma A_1 \Sigma A_2 \Sigma + 4 \Sigma A_2 \Sigma A_1 \Sigma.$$

Proof of Lemma 4.2. We only establish (iii), while (i) and (ii) follow (in a simpler way) along the same lines. Since Σ is positive definite matrix, there exists a non-singular matrix C such that $CC' = \Sigma$ or $I = C^{-1} \Sigma C'^{-1}$. Let

$$\bar{u} = C^{-1} u,$$

then

$$\bar{u} \sim N_p(0, I).$$

The result (iii) now follows by first noting that

$$E\left[\sum_{s=1}^2 u(u'A_s u)u'\right] = CE\left[\sum_{s=1}^2 \bar{u}(\bar{u}' \bar{A}_s \bar{u})\bar{u}'\right]C'$$

where $\bar{A}_s = C'A_s C$ and then applying Lemma 4.1. $\square$

Lemma 4.3. Let $u \sim N_p(0, \Sigma)$, $Z_i = \lambda'(i)u$, and $\hat{\sigma}_i^2 = u'A_i u$, $i=1,2$,

where $\lambda(i)$ and A_i are non-stochastic matrices of order $p \times 1$ and $p \times p$ respectively, and Σ is a positive definite matrix. Then

$$E[Z_i Z_j (\hat{\sigma}_i^2 - E\hat{\sigma}_i^2)(\hat{\sigma}_j^2 - E\hat{\sigma}_j^2)] \\ = \lambda'(i) \Sigma \lambda(j) \text{cov}(\hat{\sigma}_i^2, \hat{\sigma}_j^2) + 4\lambda'(i) \Sigma A_i \Sigma A_j \Sigma \lambda(j) + 4\lambda'(i) \Sigma A_j \Sigma A_i \lambda(j).$$

Proof of Lemma 4.3. The desired result follows immediately by using Lemma 4.2, after noting the following:

$$E \hat{\sigma}_1^2 = \text{tr } A_1 \Sigma,$$

$$E \hat{\sigma}_1^2 \hat{\sigma}_j^2 = 2 \text{tr } A_1 \Sigma A_j \Sigma$$

and

$$\begin{aligned} & E[Z_i Z_j (\hat{\sigma}_1^2 - E\hat{\sigma}_1^2) (\hat{\sigma}_j^2 - E\hat{\sigma}_j^2)] \\ &= \lambda'(i) E[u(u'A_i u - u'A_j u)u'] \lambda(j) - (E\hat{\sigma}_1^2) \lambda'(i) E[u(u'A_j u)u'] \lambda(j) \\ &\quad - (E\hat{\sigma}_j^2) \lambda'(i) E[u(u'A_i u)u'] \lambda(j) + E(\hat{\sigma}_1^2) E(\hat{\sigma}_j^2) \lambda'(i) \Sigma \lambda(j). \quad \square \end{aligned}$$

Lemma 4.4. Let $\{n_i\}_{i=1}^t$ be a sequence of integers such that

$\sup_{1 \leq i \leq t} n_i < \infty$. Also, let

$$(a) \quad \Sigma = \text{diag} [\Sigma_i]_{n \times n, 1 \leq i \leq t},$$

$$(b) \quad G = [O(t^{-1})]_{n \times n},$$

$$(c) \quad C = \text{diag} [O(t^{-1})]_{n_1 \times n_1} + [O(t^{-2})]_{n \times n}$$

and

$$(d) \quad P(i) = \text{col}_{1 \leq h \leq t} \text{col}_{1 \leq j \leq n_h} (O(t^{-1})) + \text{col}_{1 \leq h \leq t} \text{col}_{1 \leq j \leq n_h} (\delta_{ih} O(1)),$$

where Σ_i is a $n_i \times n_i$ matrix with bounded elements and $[O(t^{-s})]_{a \times b}$ is an $a \times b$ matrix with all the elements uniformly of order $O(t^{-s})$.

Then

$$(i) \quad \Sigma G = [O(t^{-1})]_{n \times n}$$

$$(ii) \quad \Sigma C \Sigma C \Sigma = [O(t^{-2})]_{n \times n}$$

and

$$(iii) \quad T'(1) \text{LCZCZP}(1) = O(t^{-2}).$$

Proof of Lemma 4.4. By computing the product of the matrices in (i), (ii) and (iii) using (a) - (d) with the condition that $\sup_{1 \leq i \leq t} n_i < \infty$, all the results can be verified.

We need the following definition before giving Lemma 4.5.

Definition 4.1. An estimator $T(y)$ is said to be translation invariant with respect to the general mixed model (2.11) iff

$$T(y) = T(y - xa)$$

for all non-stochastic vectors $a' = (a_1, \dots, a_k)$.

Lemma 4.5. For the general mixed model (2.11), let $\hat{\theta}$ be translation invariant estimator of θ , the vector of variance components. Also, let $t(\hat{\theta}, y) = t(\theta, y)$ evaluated at $\theta = \hat{\theta}$, where $t(\theta, y)$ is defined by (2.14). Then with the assumption of normality for y and e , the random quantities $(t(\hat{\theta}, y) - k'e - m'y)$ and $(t(\hat{\theta}, y) - t(\theta, y))$ are independently distributed.

Proof of Lemma 4.5. See Section 2.1 of Kackar and Harville (1984).

The following lemma is a well-known result on matrix differentiation.

Lemma 4.6. If A and B are two matrices whose elements are differentiable functions of θ , then

$$\frac{\partial}{\partial \theta} AB = \left(\frac{\partial}{\partial \theta} A \right) B + A \left(\frac{\partial}{\partial \theta} B \right)$$

and further, if B is non-singular for all θ , then

$$\frac{\partial}{\partial \theta} B^{-1} = -B^{-1} \left(\frac{\partial}{\partial \theta} B \right) B^{-1}$$

Proof of Theorem 4.1. For simplicity, we use $\hat{Q}^2$, $d(i)$ and $D(i)$ to denote respectively $(\hat{\sigma}_e^2, \hat{\sigma}_v^2)'$, the vector of first order derivatives of $t_i(\hat{Q}^2, y)$ with respect to $\hat{Q}^2$ evaluated at $\hat{Q}^2 = Q^2$ and the matrix of second order derivatives of $t_i(\hat{Q}^2, y)$ evaluated at $\hat{Q}^2 = Q^2 + \lambda'(\hat{Q}^2 - Q^2)$, where $\hat{\sigma}_e^2, \hat{\sigma}_v^2$ are defined by (3.9) - (3.13) and $\lambda' = (\lambda_1, \lambda_2)$ with $0 < \lambda_1, \lambda_2 < 1$.

It can be shown that $\hat{Q}^2$ is a translation invariant estimator of Q^2 with respect to the nested error regression model (see Definition 4.1). Hence, applying Lemma 4.5, we get

$$E(t_i(\hat{Q}^2, y) - \mu_i)^2 = E[t_i(\hat{Q}^2, y) - t_i(Q^2, y)]^2 + E[t_i(Q^2, y) - \mu_i]^2. \quad (4.5)$$

Thus to establish (4.2) it is sufficient to show that

$$\begin{aligned} & E[t_i(\hat{Q}^2, y) - t_i(Q^2, y)]^2 \\ &= n_i^{-2} \left(\sigma_v^2 + \sigma_e^2 n_i^{-1} \right)^{-3} \text{var}(\hat{\sigma}_e^2 \hat{\sigma}_v^2 - \sigma_e^2 \sigma_v^2) + o(t^{-1}). \end{aligned} \quad (4.6)$$

Now, we establish the above result (4.6) in the following steps:

Step 1. By noting that $\hat{Q}^2 - Q^2 = O_p(t^{-1/2})$ (from (3.5)-(3.7)) and that the elements of the matrix $D(i)$ are of order $O_p(1)$ (from the regularity conditions C1 and C2, and Lemma 4.6), the Taylor's expansion of $t_i(\hat{Q}^2, y)$ about the point Q^2 leads to

$$E[t_i(\hat{Q}^2, y) - t_i(Q^2, y)]^2 = E[(\hat{Q}^2 - Q^2)' d(i)]^2 + o(t^{-1}). \quad (4.7)$$

In establishing (4.7), the result $E[o_p(t^{-1})] = o(t^{-1})$ is used, for

which the required conditions, (e.g. Theorem 5.4.3 of Fuller, 1976) have been verified.

Step 2. We evaluate $E[(\sigma^2 - \sigma^2) d(i)]^2$ using Lemma 4.3. To apply Lemma 4.3, we first write $d_1(i)$ and $d_2(i)$ (using Lemma 4.6) as

$$d_1(i) = \frac{\partial}{\partial \sigma^2} t_i(\sigma^2, y) \Big|_{\sigma^2 = \sigma^2} = P_1(i) u \quad (4.8)$$

and

$$d_2(i) = \frac{\partial}{\partial \sigma^2} t_i(\sigma^2, y) \Big|_{\sigma^2 = \sigma^2} = P_2(i) u, \quad (4.9)$$

where

$$P_1(i) = -q_1(\sigma^2) - \sigma_{v i}^{2-1} (\sigma_v^2 + \sigma_{e i}^{2-1})^{-2} \Delta_i \quad (4.10)$$

and

$$P_2(i) = -q_2(\sigma^2) + \sigma_{e i}^{2-1} (\sigma_v^2 + \sigma_{e i}^{2-1})^{-2} \Delta_i \quad (4.11)$$

with

$$q_1(\sigma^2) = (\bar{X}_i - \gamma_i \bar{x}_i) (X' V^{-1} X)^{-1} X' V^{-1} V^{-1} A + \sigma_{e i}^{2-1} (\sigma_v^2 + \sigma_{e i}^{2-1})^{-2} \bar{x}_i (X' V^{-1} X)^{-1} X' V^{-1} \quad (4.12)$$

$$q_2(\sigma^2) = (\bar{X}_i - \gamma_i \bar{x}_i) (X' V^{-1} X)^{-1} X' V^{-1} (\text{diag } 1_{n_i} 1_{n_i}') V^{-1} A - \sigma_{v i}^{2-1} (\sigma_v^2 + \sigma_{e i}^{2-1})^{-2} \bar{x}_i (X' V^{-1} X)^{-1} X' V^{-1} \quad (4.13)$$

$$A = I - X(X' V^{-1} X)^{-1} X' V^{-1} \quad (4.14)$$

and

$$\Delta_i = \text{col}_{1 \leq l \leq t} \text{col}_{1 \leq j \leq n_l} (\delta_{il}) \quad (4.15)$$

Also, it can be shown (using the conditions C1 - C2) that

$$q_1(\sigma^2) = \text{col}_{1 \leq h \leq t} \text{col}_{1 \leq j \leq n_h} (O(t^{-1})) \quad (4.16)$$

and

$$q_2(\sigma^2) = \text{col}_{1 \leq h \leq t} \text{col}_{1 \leq j \leq n_h} (O(t^{-1})), \quad (4.17)$$

where the bounds in defining $O(t^{-1})$ are independent of h .

Next, we apply Lemma 4.3 with $A_s = C_s$ and $Z_s = P'_s(i)u$, ($s=1,2$), where $u \sim N(0, V)$ and V and C_1, C_2 are given by (2.4) and (3.11) - (3.13) respectively. This leads to

$$E\{(\hat{\sigma}^2 - \sigma^2) d(i)\}^2 = \sum_{q=1}^2 \sum_{s=1}^2 \{ \text{cov}(\hat{\sigma}_q^2, \hat{\sigma}_s^2) P'_q(i) V P'_s(i) + 8 P'_q(i) V C_q V C_s V P'_s(i) \}, \quad (4.18)$$

where, without loss of generality, we have denoted $(\hat{\sigma}_e^2, \hat{\sigma}_v^2)$ by $(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$.

Step 3. The order of the terms in (4.18) is now obtained. Using Lemma 4.4, we get

$$P'_q(i) V C_q V C_s P'_s(i) = o(t^{-1}); \quad q, s = 1, 2 \quad (4.19)$$

and

$$P'_q(i) V P'_s(i) = (-1)^{q+s} \sigma_q^2 \sigma_s^2 n_i^{-2} (\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-3} + O(t^{-1}), \quad (4.20)$$

by noting, from (4.16), (4.17) and (3.11)-(3.13), that $P_q(i)$ and C_q ($q=1,2$) are respectively of the form $P(i)$ and C defined in

Lemma 4.4. The desired result (4.2) now follows from (4.5), (4.18)-

(4.20) and by noting that $\text{cov}(\hat{\sigma}_q^2, \hat{\sigma}_s^2) = O(t^{-1})$; $q, s = 1, 2$.

Remark 4.1. 1. Harter (1983) derived an approximation to MSE of

$t_1(\sigma^2, y)$, but it can be shown that some of the neglected terms are of order $O(t^{-1})$. Battese and Fuller (1982) obtained an approximation to MSE identical to our expression (4.2), but they do not justify that the approximation is in fact correct to terms of order $O(t^{-1})$.

2. The second term in the right side of (4.5) reflects the increase in MSE of the i^{th} small area predictors when the variance components σ_e^2, σ_v^2 are estimated to compute $t_1(\sigma^2, y)$.

3. Theorem 4.1 also holds true when σ^2 is estimated by any other method which permits us to write $\hat{\sigma}_i^2$ as $\hat{\sigma}_i^2 = u'C_i u$, or $\hat{\sigma}_i^2 = u'C_i u + \frac{1+\delta}{p}(t^{-1})$, where $\delta > 0$, and C_i is of the form C defined in Lemma 4.4. The minimum norm quadratic unbiased estimators (MINQUE) of σ_i^2 satisfy the above conditions. The computation of MINQUE is discussed in Chapter 6, Section 6.3.

4. If σ_v^2 defined by (3.21) is used instead of $\hat{\sigma}_v^2$ in estimating the BLUP, then an approximation to $\text{MSE}(t_1(\hat{\sigma}_v^2, y))$ can be obtained by replacing $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$ respectively by $\tilde{\sigma}_e^2$ and $\tilde{\sigma}_v^2$ in the right side of (4.2).

Using our jackknife procedure described in Section 3.4, we give below an approximately unbiased estimator of γ_i . By taking $g(\sigma_e^2, \sigma_v^2) = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 n_i^{-1})$ in (3.58), we get jackknife estimator

$$\gamma_i(J) = \tilde{\gamma}_i - t^{-1} \sum_{\ell=1}^t \{ \tilde{\sigma}_v^2(\ell) / (\tilde{\sigma}_v^2(\ell) + \tilde{\sigma}_e^2(\ell) n_i^{-1}) - \tilde{\gamma}_i \} \quad (4.20)$$

where

$$\tilde{\sigma}_e^2(\ell) = \tilde{\sigma}_e^2 + a_\ell(\tilde{\sigma}_e^2(-\ell) - \tilde{\sigma}_e^2)$$

$$\tilde{\sigma}_v^2(\ell) = \tilde{\sigma}_v^2 + b_\ell(\tilde{\sigma}_v^2(-\ell) - \tilde{\sigma}_v^2)$$

with σ_e^2 , σ_v^2 , $\sigma_e^2(-\ell)$, $\sigma_v^2(-\ell)$, a_ℓ and b_ℓ ; $\ell=1, \dots, t$ as defined in Section 3.3.

From Theorem 3.2, it follows that

$$E(\hat{\gamma}_i(J) - \gamma_i) = o(t^{-1}),$$

whereas

$$E(\hat{\gamma}_i - \gamma_i) = O(t^{-1}).$$

The predictor $t_i^*(\hat{\sigma}^2, y)$, based on $\hat{\gamma}_i(J)$ is then given by

$$t_i^*(\hat{\sigma}^2, y) = \bar{x}_{i.} \hat{\beta} + \hat{\gamma}_i(J) (\bar{y}_{i.} - \bar{x}_{i.} \hat{\beta}). \quad (4.21)$$

Theorems 4.2 and 4.3 below establish results analogous to Theorem 4.1 respectively for the random regression model with one auxiliary variable, and the Fay-Herriot model.

Theorem 4.2. For the random-regression model (2.7) with one auxiliary variable (i.e., $k=1$, $x_{ij1} = x_{ij}$, $\psi = \sigma_v^2$), let the regularity conditions C1 - C2 of Theorem 4.1 and the normality assumption for the random errors y and e be true. Then the MSE of the estimated BLUP

$$t_i(\hat{\sigma}^2, y) = \bar{x}_{i.} \hat{\beta} + \hat{\gamma}_i(\bar{y}_{i.} - \bar{x}_{i.} \hat{\beta}), \quad (4.22)$$

with $\hat{\gamma}_i = \hat{\sigma}_v^2 [\hat{\sigma}_v^2 + \hat{\sigma}^2 (\sum_j x_{ij}^2)^{-1}]^{-1}$ can be approximated as follows:

$$\begin{aligned} \text{MSE}(t_i(\hat{\sigma}^2, y)) &= E[t_i(\hat{\sigma}^2, y) - \mu_i]^2 \\ &= E[t_i(\hat{\sigma}^2, y) - \mu_i]^2 \\ &\quad + \bar{x}_{i.}^2 (\sum_j x_{ij}^2) (\sigma_v^2 \sum_j x_{ij}^2 + \sigma_e^2)^{-3} \text{Var}(\hat{\sigma}_e^2 \hat{\sigma}_v^2 - \hat{\sigma}_v^2 \hat{\sigma}_e^2) \\ &\quad + o(t^{-1}), \end{aligned} \quad (4.23)$$

where $E[t_1(\bar{y}, y) - \mu_1]^2$ is defined by (2.21).

Theorem 4.3. For the Fay-Herriot model (2.10), let the following regularity conditions and normality assumption for the error components be true:

(a) $\sup_{1 \leq i \leq t} D_i < \infty$

(b) The elements of the matrix X are uniformly bounded such that

(i) $\lim_{t \rightarrow \infty} t(X'V^{-1}X)^{-1} = G$ and G^{-1} exist

(ii) $\lim_{t \rightarrow \infty} tX'[\text{diag}(A + D_i)^{-1}]X = S$ and S^{-1} exist almost surely

and

(iii) $\lim_{t \rightarrow \infty} tX' \frac{\partial V^{-1}}{\partial A} X = H$,

where the elements of H are continuous function of A , then the MSE of the estimated BLUP (empirical Bayes estimator)

$$t_1(A, \bar{y}) = \bar{X}_{1.} \hat{\beta} \frac{D_1}{A + D_1} + \frac{A}{A + D_1} \bar{y}_{1.} \quad (4.24)$$

can be approximated as follows:

$$\begin{aligned} \text{MSE}(t_1(A, \bar{y})) &= E[t_1(A, \bar{y}) - \bar{y}_{1.}]^2 \\ &= \bar{X}_{1.} \text{Var}(\hat{\beta}) \bar{X}_{1.}' + A D_1 (A + D_1)^{-1} \\ &\quad + D_1^2 \text{Var}(\hat{A}) (A + D_1)^{-3} + o(t^{-1}), \end{aligned} \quad (4.25)$$

where

$$\hat{\beta} = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} \bar{y}$$

with

$$\hat{V} = \text{diag}(\hat{A} + D_i)_{1 \leq i \leq t}$$

and $\hat{A}$ as defined by (3.16).

4.2. Estimation of Mean Square Error

a. Nested error regression model

For the MSE of the estimated BLUP $t_i(\hat{\sigma}^2, y)$ derived in the previous section, we now obtain an estimator $mse(t_i(\hat{\sigma}^2, y))$, which has a bias of order lower than $O(t^{-1})$, that is

$$E\{mse(t_i(\hat{\sigma}^2, y))\} - MSE(t_i(\hat{\sigma}^2, y)) = o(t^{-1}). \quad (4.26)$$

To derive $mse(t_i(\hat{\sigma}^2, y))$ for the nested error regression model, from the equation (4.2) we write an expression for $MSE^*(t_i(\hat{\sigma}^2, y))$, the approximation to $MSE(t_i(\hat{\sigma}^2, y))$, as

$$\begin{aligned} MSE^*(t_i(\hat{\sigma}^2, y)) &= (\bar{x}_{i.} - \gamma_i \bar{x}_{i.})(X'V^{-1}X)^{-1}(\bar{x}_{i.} - \gamma_i \bar{x}_{i.})' \\ &\quad + (1 - \gamma_i)\sigma_v^2 + n_i^{-2}(\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-3}[\sigma_v^4 \text{Var}(\hat{\sigma}_e^2) \\ &\quad + \sigma_e^4 \text{Var}(\hat{\sigma}_v^2) - 2 \text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2)\sigma_e^2 \sigma_v^2]. \end{aligned} \quad (4.27)$$

Without loss of generality, we shall denote the terms on the right side of (4.27) respectively by $g_1(\hat{\sigma}^2)$, $g_2(\hat{\sigma}^2)$ and $g_3(\hat{\sigma}^2)$. Since $g_1(\hat{\sigma}^2)$ and $g_3(\hat{\sigma}^2)$ are of order $O(t^{-1})$, it follows that

$$Eg_1(\hat{\sigma}^2) - g_1(\hat{\sigma}^2) = o(t^{-1}) \quad (4.28)$$

and

$$Eg_3(\hat{\sigma}^2) - g_3(\hat{\sigma}^2) = o(t^{-1}), \quad (4.29)$$

where

$$g_1(\hat{\sigma}^2) = (\bar{x}_{i.} - \hat{\gamma}_i \bar{x}_{i.})(X'V^{-1}X)^{-1}(\bar{x}_{i.} - \hat{\gamma}_i \bar{x}_{i.})'$$

and

$$\begin{aligned} g_3(\hat{\sigma}^2) &= n_i^{-2}(\hat{\sigma}_v^2 + \hat{\sigma}_e^2 n_i^{-1})^{-3}[\{\sigma_e^4 \text{Var}(\hat{\sigma}_v^2) + \sigma_v^4 \text{Var}(\hat{\sigma}_e^2) \\ &\quad - 2\sigma_e^2 \sigma_v^2 \text{Cov}(\hat{\sigma}_v^2, \hat{\sigma}_e^2)\}_{\hat{\sigma}^2 = \hat{\sigma}^2}]. \end{aligned}$$

with $\text{Var}(\hat{\sigma}_e^2)$, $\text{Var}(\hat{\sigma}_v^2)$ and $\text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2)$ as defined by (3.5) - (3.7).

However, for the second term $g_2(\hat{\sigma}^2)$ which is of order $O(1)$, we have

to estimate the bias in

$$g_2(\hat{\sigma}^2) = (1 - \frac{1}{n_i}) \hat{\sigma}_v^2 = \frac{\hat{\sigma}_e^2 \hat{\sigma}_v^2}{\sigma_e^2 \sigma_v^2} (n_i \hat{\sigma}_v^2 + \hat{\sigma}_e^2) \cdot 1$$

to order $O(t^{-1})$.

The Taylor's expansion of $g_2(\hat{\sigma}^2)$ about the point, σ^2 , gives

$$g_2(\hat{\sigma}^2) = g_2(\sigma^2) + (\hat{\sigma}^2 - \sigma^2) \nabla g_2(\sigma^2) + \frac{1}{2} (\hat{\sigma}^2 - \sigma^2)^2 \nabla^2 g_2(\sigma^2) + O_p(t^{-3/2}), \quad (4.30)$$

where $\nabla g_2(\sigma^2)$ and $\nabla^2 g_2(\sigma^2)$ are the first and second order derivatives of $g_2(\cdot)$ with respect to σ^2 . After noting that the required conditions (as in Theorem 5-4.3 of Fuller (1976)) are satisfied for $E[O_p(t^{-3/2})] = O(t^{-2})$ to hold, it follows that

$$\begin{aligned} E g_2(\hat{\sigma}^2) &= g_2(\sigma^2) + \frac{1}{2} \text{tr}[\text{Var}(\hat{\sigma}^2) \cdot \nabla^2 g_2(\sigma^2)] + O(t^{-2}) \\ &= g_2(\sigma^2) - n_i^{-2} (\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-3} [\sigma_e^4 \text{Var}(\hat{\sigma}_v^2) \\ &\quad + \sigma_v^4 \text{Var}(\hat{\sigma}_e^2) - 2 \text{Cov}(\hat{\sigma}_v^2, \hat{\sigma}_e^2) \sigma_e^2 \sigma_v^2], \quad (4.31) \end{aligned}$$

By observing that

$$\nabla^2 g_2(\sigma^2) = -2 n_i^{-2} (\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-3} \begin{bmatrix} \sigma_e^4 & -\sigma_e^2 \sigma_v^2 \\ -\sigma_e^2 \sigma_v^2 & \sigma_v^4 \end{bmatrix}$$

Thus, with

$$\begin{aligned} g_2(\hat{\sigma}^2) &= g_2(\sigma^2) + n_i^{-2} (\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-3} [\sigma_e^4 \text{Var}(\hat{\sigma}_v^2) \\ &\quad + \sigma_v^4 \text{Var}(\hat{\sigma}_e^2) - 2 \text{Cov}(\hat{\sigma}_v^2, \hat{\sigma}_e^2) \sigma_e^2 \sigma_v^2] \end{aligned}$$

$$= g_2(\hat{\sigma}^2) + g_3(\hat{\sigma}^2),$$

we get

$$Eg_2(\hat{\sigma}^2) = g_2(\sigma^2) + o(t^{-1}). \quad (4.32)$$

Hence, from (4.23), (4.24) and (4.27) an expression for $mse(t_1(\hat{\sigma}^2, y))$, satisfying (4.26) is given by

$$\begin{aligned} mse(t_1(\hat{\sigma}^2, y)) &= (\bar{X}_{i.} - \hat{\gamma}_i \bar{x}_{i.}) (X' \hat{V}^{-1} X)^{-1} (\bar{X}_{i.} - \hat{\gamma}_i \bar{x}_{i.})' \\ &\quad + (1 - \hat{\gamma}_i) \hat{\sigma}_v^2 + 2n_i^{-2} (\hat{\sigma}_v^2 + \hat{\sigma}_e^2 n_i^{-1})^{-3} \left[\{ \hat{\sigma}_e^4 \text{Var}(\hat{\sigma}_v^2) \right. \\ &\quad \left. + \hat{\sigma}_v^4 \text{Var}(\hat{\sigma}_e^2) - 2\hat{\sigma}_e^2 \hat{\sigma}_v^2 \text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2) \} \right]_{\hat{\sigma}_e^2 = \hat{\sigma}_v^2}, \end{aligned} \quad (4.33)$$

where $\text{Var}(\hat{\sigma}_e^2)$, $\text{Var}(\hat{\sigma}_v^2)$ and $\text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2)$ are as defined by (3.5)-(3.7).

If $\hat{\sigma}_v^2$ defined by (3.21) is used instead of $\hat{\sigma}_v^2$ in estimating the BLUP, then an estimator of $MSE(t_1(\hat{\sigma}^2, y))$ (see Remark 4.1.4) is given by

$$\begin{aligned} mse(t_1(\hat{\sigma}^2, y)) &= (\bar{X}_{i.} - \hat{\gamma}_i \bar{x}_{i.}) (X' \hat{V}^{-1} X)^{-1} (\bar{X}_{i.} - \hat{\gamma}_i \bar{x}_{i.})' \\ &\quad + (1 - \hat{\gamma}_i) \hat{\sigma}_v^2 + 2n_i^{-2} (\hat{\sigma}_v^2 + \hat{\sigma}_e^2 n_i^{-1})^{-3} \left[\{ \hat{\sigma}_e^4 \text{Var}(\hat{\sigma}_v^2) \right. \\ &\quad \left. + \hat{\sigma}_v^4 \text{Var}(\hat{\sigma}_e^2) - 2\hat{\sigma}_e^2 \hat{\sigma}_v^2 \text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2) \} \right]_{\hat{\sigma}_e^2 = \hat{\sigma}_v^2}. \end{aligned} \quad (4.34)$$

It can also be verified that

$$E mse(t_1(\hat{\sigma}^2, y)) - MSE(t_1(\hat{\sigma}^2, y)) = o(t^{-1}).$$

Remark 4.2. From our Monte Carlo Study in Section 3.5, it has been observed that the estimator of $\text{Var}(\hat{\sigma}^2)$, defined by (3.75)-(3.78) is sensitive to

non-normality. When these estimates are used in (4.34) to compute $mse(t_1(\hat{\sigma}^2, y))$ the resulting estimate may become a serious under estimate or over estimate of the $MSE^*(t_1(\hat{\sigma}^2, y))$ under non-normally distributed random errors.

In view of the above remark, it would be useful to develop robust estimators of $MSE^*(t_1(\hat{\sigma}^2, y))$. Using the jackknife estimators of $Var(\hat{\sigma}_v^2)$, $Var(\hat{\sigma}_e^2)$ and $Cov(\hat{\sigma}_e^2, \hat{\sigma}_v^2)$ given in Section 3.3, a robust estimator of $MSE^*(t_1(\hat{\sigma}^2, y))$ is given by

$$\begin{aligned} mse_1^*(t_1(\hat{\sigma}^2, y)) &= (\bar{X}_1 - \bar{Y}_1 \bar{X}_1)(X'V^{-1}X)^{-1}(\bar{X}_1 - \bar{Y}_1 \bar{X}_1)' \\ &\quad + (1 - \bar{Y}_1) \hat{\sigma}_v^2 + 2n_i^{-2}(\hat{\sigma}_v^2 + \hat{\sigma}_e^2 n_i^{-1})^{-3} [\hat{\sigma}_{eJ}^4 \hat{\sigma}_v^2 \\ &\quad + \hat{\sigma}_{vJ}^4 \hat{\sigma}_e^2 - 2\hat{\sigma}_e^2 \hat{\sigma}_v^2 Cov_J(\hat{\sigma}_e^2, \hat{\sigma}_v^2)]. \end{aligned} \quad (4.35)$$

An alternate robust estimator can be obtained by jackknifing $g_2(\hat{\sigma}^2)$ directly:

$$\begin{aligned} mse_2^*(t_1(\hat{\sigma}^2, y)) &= (\bar{X}_1 - \bar{Y}_1 \bar{X}_1)(X'V^{-1}X)^{-1}(\bar{X}_1 - \bar{Y}_1 \bar{X}_1)' \\ &\quad + g_{2J}(\hat{\sigma}^2) + n_i^{-2}(\hat{\sigma}_v^2 + \hat{\sigma}_e^2 n_i^{-1})^{-3} [\hat{\sigma}_{eJ}^4 \hat{\sigma}_v^2 \\ &\quad + \hat{\sigma}_{vJ}^4 \hat{\sigma}_e^2 - 2\hat{\sigma}_e^2 \hat{\sigma}_v^2 Cov_J(\hat{\sigma}_e^2, \hat{\sigma}_v^2)], \end{aligned} \quad (4.36)$$

where

$$\begin{aligned} g_{2J}(\hat{\sigma}^2) &= \hat{\sigma}_e^2 \hat{\sigma}_v^2 (n_i \hat{\sigma}_v^2 + \hat{\sigma}_e^2)^{-1} \\ &\quad - t^{-1} \sum_{\ell=1}^t \{ \hat{\sigma}_e^2(\ell) \hat{\sigma}_v^2(\ell) (n_i \hat{\sigma}_v^2(\ell) + \hat{\sigma}_e^2(\ell))^{-1} \\ &\quad - \hat{\sigma}_e^2 \hat{\sigma}_v^2 (n_i \hat{\sigma}_v^2 + \hat{\sigma}_e^2)^{-1} \} \end{aligned} \quad (4.37)$$

with $\hat{\sigma}_e^2(l)$, $\hat{\sigma}_v^2(l)$ as defined in Section 4.1 (below 4.20).

The formula for $\text{mse}(t_1(\hat{\theta}, y))$ similar to (4.34) under the other two small area models are stated below:

b. Random Regression Model: For the random regression model (2.7), an estimator for $\text{MSE}(t_1(\hat{\theta}^2, y))$ defined by (4.23), is given by

$$\begin{aligned} \text{mse}(t_1(\hat{\theta}^2, y)) = & (\bar{X}_{1.} - \bar{t}_1 \bar{x}_{1.})^2 \left[\sum_{i,j} x_{1j}^2 (\hat{\sigma}_e^2 + \hat{\sigma}_v^2 \sum_{j,j} x_{1j}^2)^{-1} - 1 + (1 - \bar{t}_1) \hat{\sigma}_v^2 \right. \\ & + 2\bar{X}_{1.}^2 \left(\sum_{j,j} x_{1j}^2 \right) (\hat{\sigma}_e^2 + \hat{\sigma}_v^2 \sum_{j,j} x_{1j}^2)^{-3} \left[\frac{4}{e} \text{Var}(\hat{\sigma}_v^2) + \frac{4}{v} \text{Var}(\hat{\sigma}_e^2) \right. \\ & \left. \left. - 2 \text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2) \frac{\hat{\sigma}_e^2 \hat{\sigma}_v^2}{\hat{\sigma}_e^2 \hat{\sigma}_e^2, \hat{\sigma}_v^2 \hat{\sigma}_v^2} \right] \right] \end{aligned}$$

where $\bar{t}_1$ is as defined in Theorem 4.2 and $\text{Var}(\hat{\sigma}_e^2)$, $\text{Var}(\hat{\sigma}_v^2)$ and $\text{Cov}(\hat{\sigma}_e^2, \hat{\sigma}_v^2)$ are the variances and covariances of $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$ (as defined by (3.14) and (3.15)) under the normality assumption.

c. Fay-Herriot Model: For the Fay-Herriot model (2.10) an estimator for $\text{MSE}(t_1(\hat{A}, \bar{y}))$, defined by (4.25), is given by

$$\begin{aligned} \text{mse}(t_1(\hat{A}, \bar{y})) = & \bar{X}_{1.} (X' V^{-1} X)^{-1} \bar{X}_{1.}' + \bar{A} D_i (\bar{A} + D_i)^{-1} \\ & + 2D_i^2 \text{Var}(\hat{A}) (\bar{A} + D_i)^{-3} \end{aligned}$$

where V is as defined in Theorem 4.3 and $\text{Var}(\hat{A}) = \frac{1}{A} \text{Var}(A)$, with $\text{Var}(A)$ as the variance of A under the normality assumption.

4.3. A Monte Carlo Study of the Nested Error Model

In order to determine the finite-sample properties of our proposed estimation strategies, a Monte Carlo study was undertaken, based on a

nested error regression model having one auxiliary variable ($k = 2$, $x_{ij1} = 1$, $x_{ij2} = x_{ij}$, $\beta_1 = \alpha$ and $\beta_2 = \beta$). Detailed objectives, a description of the Monte Carlo experiment, and a discussion of the findings are furnished in the following subsections.

a. Objectives of the Monte Carlo Study. The objectives of our Monte Carlo study are four fold. In the first phase, we examine the efficiency of the estimated BLUP $t_1(\sigma^2, y)$ relative to the direct estimator, the synthetic estimator and the 'adjusted survey predictor'. These estimators are defined below (for direct and synthetic estimators see also Sect. 1.1).

Direct estimator: The direct estimator (expansion estimator) which utilizes only the sample y -data in the i^{th} small area is given by

$$t_1(1) = \bar{y}_i. \quad (4.38)$$

Synthetic estimator: A simple synthetic estimator is obtained by first assuming that

$$y_{ij} = \alpha + x_{ij}\beta + e_{ij}, \quad i=1, \dots, t; \quad j=1, \dots, n_i,$$

i.e., by taking $\sigma_v^2 = 0$ in the nested error regression model and then deriving the BLUP. This results in a synthetic regression estimator given by

$$t_1(2) = \bar{x}_{i..}\hat{\beta} = \bar{x}_{i..}\hat{\beta}, \quad (4.39)$$

where $\hat{\beta} = \sum_{i,j} (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) / \sum_{i,j} (x_{ij} - \bar{x}_{..})^2$ and $\hat{\alpha} = \bar{y}_{..} - \bar{x}_{..}\hat{\beta}$,

with $\bar{y}_{..} = n^{-1} \sum_{i,j} y_{ij}$ and $\bar{x}_{..} = n^{-1} \sum_{i,j} x_{ij}$.

Adjusted survey predictor: This estimator is obtained under the assumption that $n_i \rightarrow \infty$ or $\sigma_e^2 \rightarrow 0$, and can be expressed as

$$t_1(3) = \bar{y}_{i.} + (\bar{x}_{i.} - \bar{x}_{i.})\beta, \quad (4.40)$$

where β is as defined above and $\bar{x}_{i.}$ is the sample mean of the i th small area for the variable x . Battese and Fuller (1982) called this estimator the 'adjusted survey predictor'.

The relative efficiencies of the estimated BLUP with respect to other small area predictors $t_1(1)$, $t_1(2)$ and $t_1(3)$ are defined by

$$RE(t_1(\tilde{\sigma}^2; y) : t_1(k)) = 100 \times MSE(t_1(k)) / MSE(t_1(\tilde{\sigma}^2, y)); k=1,2,3, \quad (4.41)$$

where $MSE(t_1(k)); k=1,2,3$ are the mean squares of the predictors $t_1(1)$, $t_1(2)$ and $t_1(3)$ respectively.

The second objective of this Monte Carlo experiment is to determine the accuracy of the approximated mean square error of the estimated BLUP, $MSE^*(t_1(\tilde{\sigma}^2, y))$, by computing its percentage relative error. This is defined by

$$\begin{aligned} R.E. MSE^* &= R.E. (MSE^*(t_1(\tilde{\sigma}^2, y))) \\ &= 100 \{ MSE^*(t_1(\tilde{\sigma}^2, y)) - MSE(t_1(\tilde{\sigma}^2, y)) \} / MSE(t_1(\tilde{\sigma}^2, y)), \end{aligned} \quad (4.42)$$

where (from Remark 4.1.4)

$$\begin{aligned} MSE^*(t_1(\tilde{\sigma}^2, y)) &= (\bar{x}_{i.} - \gamma_1 \bar{x}_{i.}) (X^T V^{-1} X)^{-1} (\bar{x}_{i.} - \gamma_1 \bar{x}_{i.}) \\ &\quad + (1 - \gamma_1) \sigma_v^2 + n_1^{-2} (\sigma_v^2 + \sigma_e^2 - 1)^{-3} \{ \sigma_e^4 \text{Var}(\tilde{\sigma}_v^2) \\ &\quad + \sigma_v^4 \text{Var}(\tilde{\sigma}_e^2) - 2 \sigma_e^2 \sigma_v^2 \text{Cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) \}, \end{aligned} \quad (4.43)$$

with $\text{Var}(\tilde{\sigma}_v^2)$, $\text{Var}(\tilde{\sigma}_e^2)$ and $\text{Cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ as defined by (3.76)-(3.78).

Note that $MSE(t_1(\tilde{\sigma}^2, y))$ will be a Monte Carlo estimate.

Our third objective is to study the percentage relative bias of the normality based estimator $mse^*(t_1(\hat{\sigma}^2, y))$ and the jackknife estimator $mse_1^*(t_1(\hat{\sigma}^2, y))$ (defined respectively by (4.34) and (4.35)) with respect to the true MSE and the approximate MSE, $MSE^*(t_1(\hat{\sigma}^2, y))$. With respect to the true MSE, these relative biases are denoted by $R.Bias_N$ and $R.Bias_J$ respectively. With respect to the approximate MSE, they are denoted by $R.Bias_N^*$ and $R.Bias_J^*$ respectively. Actual definitions are given below:

$$R.Bias_N = 100[E\{mse^*(t_1(\hat{\sigma}^2, y))\} - MSE(t_1(\hat{\sigma}^2, y))] / MSE(t_1(\hat{\sigma}^2, y)), \quad (4.44)$$

$$R.Bias_J = 100[E\{mse_1^*(t_1(\hat{\sigma}^2, y))\} - MSE(t_1(\hat{\sigma}^2, y))] / MSE(t_1(\hat{\sigma}^2, y)), \quad (4.45)$$

$$R.Bias_N^* = 100[E\{mse^*(t_1(\hat{\sigma}^2, y))\} - MSE^*(t_1(\hat{\sigma}^2, y))] / MSE^*(t_1(\hat{\sigma}^2, y)), \quad (4.46)$$

and

$$R.Bias_J^* = 100[E\{mse_1^*(t_1(\hat{\sigma}^2, y))\} - MSE^*(t_1(\hat{\sigma}^2, y))] / MSE^*(t_1(\hat{\sigma}^2, y)). \quad (4.47)$$

Our fourth and the final objective is to determine the percentage relative bias of the estimator, which is being used in practice:

$$mse_D = mse(t_1(\hat{\sigma}^2, y)) = \{(\bar{X}_1 - \gamma_1 \bar{x}_1)(X'V^{-1}X)^{-1}(\bar{X}_1 - \gamma_1 \bar{x}_1) + (1 - \gamma_1)\sigma_v^2\} \quad (4.48)$$

$$\sigma_e^2 = \sigma_e^2, \sigma_v^2 = \sigma_v^2$$

with respect to the true MSE. This relative bias, denoted $R.Bias_D$ is defined below

$$R.Bias_D = 100[E\{mse(t_1(\hat{\sigma}^2, y))\} - MSE(t_1(\hat{\sigma}^2, y))] / MSE(t_1(\hat{\sigma}^2, y)). \quad (4.49)$$

b. Description of the Monte Carlo experiment. For our study, we selected the nested error regression model used by Battese and Fuller (1982) to predict corn hectares (y) for 12 countries (small areas) in Iowa, using the number of pixels (x), classified as having corn. They fitted the nested error regression model with x as the auxiliary variable and obtained the following estimates: $\beta = 5.5$, $\beta = 0.388$, $\sigma_e^2 = 292$ and $\sigma_v^2 = 64$. In their data set, $n_i = 1$ for three of the small areas. We pooled these three small areas into one small area to satisfy the requirement $n_i \geq 2$ for our jackknife method. Table A (in the Appendix) reports the values of n_i , x_{ij} and the population mean $\bar{x}_i$, for each of the ten small areas. The average of the population means for the three small areas having $n_i = 1$ is used as the population mean for the aggregated small area.

Using the estimates $\sigma_e^2 = 292$ and $\sigma_v^2 = 64$ as parameter values, we generated 1000 independent sets of pseudo random deviates $\{e_{ij}; j = 1, \dots, n_i, i = 1, \dots, 10\}$ and $\{v_i; i = 1, \dots, 10\}$ having normal distributions with means zero and variance $\sigma_e^2 = 292$ and $\sigma_v^2 = 64$ respectively. From these generated sets $\{v_i\}$, $\{e_{ij}\}$, and using the x_{ij} values reported in Table A, the observations $\{y_{ij}; j = 1, \dots, n_i, i = 1, \dots, 10\}$ were obtained from the model

$$y_{ij} = 5.5 + 0.388x_{ij} + v_i + e_{ij}. \quad (4.50)$$

c. Results. Monte Carlo estimates of the mean squared errors of the four small area predictors $t_1(\sigma^2, y)$, $t_1(1)$, $t_1(2)$ and $t_1(3)$ were obtained. Relative efficiencies of the estimated BLUP, with respect

to the three alternative estimators were computed according to (4.41) and are shown in Table 4.1.

Table 4.1: Relative Efficiency (F.E.)(%) of the estimated BLUP with respect to (w.r. t.) $t_1(1)$, $t_1(2)$ and $t_1(3)$.

Small area	n_1	(%R.E. of $t_1(\hat{\sigma}^2, y)$ w.r. t.		
		$t_1(1)$	$t_1(2)$	$t_1(3)$
1	2	300	130	281
2	3	586	146	199
3	3	365	145	224
4	3	635	153	206
5	3	217	147	217
6	3	624	150	216
7	4	189	174	188
8	5	553	169	161
9	5	263	185	165
10	6	913	199	156

Remark 4.3. In computing the predictor $t_1(\hat{\sigma}^2, y)$, $\hat{\sigma}_v^2$ is taken to be zero, whenever $\hat{\sigma}_v^2 < 0$.

It is seen from Table 4.1 that the estimated BLUP is much more efficient than the other predictors. The percentage relative efficiency of $t_1(\hat{\sigma}^2, y)$ ranges from 189 to 913 with respect to $t_1(1)$, from 130 to 199

with respect to $t_i(2)$ and from 161 to 281 with respect to $t_i(3)$. The relative efficiency of the estimated BLUP with respect to $t_i(3)$, the adjusted survey predictor, decreases as n_1 increases. With respect to the synthetic estimator $t_i(2)$, this trend is reversed, the relative efficiency of the estimated BLUP, $t_i(\hat{c}^2, y)$, now increasing as n_1 increases. From Table 4.1 it can be seen that both the synthetic ($t_i(2)$) and the adjusted survey predictor ($t_i(3)$) are more efficient than the direct estimator ($t_i(1)$).

The percentage relative error of MSE^* , the approximation to the $MSE(t_i(\hat{c}^2, y))$ and the percentage relative biases of the MSE estimators: $mse^*(t_i(\hat{c}^2, y))$, $mse_1^*(t_i(\hat{c}^2, y))$ and mse_b^* were computed using (4.42), (4.44), (4.45) and (4.46), and then averaged over small areas having the same n_1 value. These averaged values are reported in Table 4.2a (see p. 87). Monte Carlo estimates of the relative biases (%) of $mse^*(t_i(\hat{c}^2, y))$ and $mse_1^*(t_i(\hat{c}^2, y))$ with respect to the approximation MSE^* (obtained from (4.46) and (4.47)) are also shown in Table 4.2a.

The following are the main findings of Table 4.2a:

- (i) MSE^* values exceed the true MSE values uniformly across all the small areas, the differences ranging from 7% to 15%.
- (ii) The values of $Bias_N^*$ show that the normality based estimator $mse^*(t_i(\hat{c}^2, y))$ over estimates the approximate MSE for $n_1 \geq 3$, the over estimation increasing with increasing n_1 , leading to the values as high as 19% (for $n_1 = 5$) and 29% (for $n_1 = 6$). From the values of $Bias_J^*$, it can be seen that the jackknife estimator $mse_1^*(t_i(\hat{c}^2, y))$

over estimates the approximate MSE (for all $n_1 \geq 3$), though it does perform a little better than $mse^*(t_1(\hat{\sigma}^2, y))$ (over estimation of 11% for $n_1 = 5$ and 16% for $n_1 = 6$).

(iii) Since $MSE^*(t_1(\hat{\sigma}^2, y))$ values exceed the true MSE values, the relative bias of both the MSE estimators with respect to the true MSE (measured by $Bias_N$ and $Bias_J$) is further increased, the over estimation in these cases reaches 37% ($n_1 = 5$) for $mse^*(t_1(\hat{\sigma}^2, y))$ and 28% ($n_1 = 6$) for $mse^*(t_1(\hat{\sigma}^2, y))$.

(iv) The estimator mse_B defined by (4.48) seriously under estimates the true MSE, as expected. The percentage of under estimation ranges from 27% to 32%.

Effect of increasing t : To examine the effect of increasing t , the number of small areas, on the performance of the MSE estimates, another Monte Carlo experiment was undertaken. By duplicating the n_1 , x_{1j} and $\bar{x}_1$, and obtaining new sets of y_{1j} , $i=1, \dots, 20$; $j=1, \dots, n_1$ from the model (4.50), an experiment involving 20 small areas was constructed. Monte Carlo estimates of relative errors, corresponding to those of Table 4.2a, are reported in Table 4.2b. It can be seen from Table 4.2b that $MSE^*(t_1(\hat{\sigma}^2, y))$, the approximation to MSE, performs much better for $t = 20$ than for $t = 10$. The maximum difference is reduced from 15% to 4%. It can be seen that the maximum relative biases of $mse^*(t_1(\hat{\sigma}^2, y))$ and the jackknife estimator $mse^*(t_1(\hat{\sigma}^2, y))$ with respect to $MSE^*(t_1(\hat{\sigma}^2, y))$ (measured by $Bias_N^*$ and $Bias_J^*$) are reduced to 8% and 9% respectively, while with respect to the true MSE, the corresponding values (measured

by Bias_N and Bias_J are reduced to 11% and 12% respectively. Again, it can be seen that mse_b badly under estimates the true MSE, though the underestimation is less than for the case $t = 10$.

Effect of non-normality: $\text{MSE}^*(t_1(\tilde{\sigma}^2, y))$, the approximation to the true MSE, was derived under the assumption of normality. To examine the effect of relaxing this assumption on the performance of $\text{MSE}^*(t_1(\tilde{\sigma}^2, y))$, and on the performance of the MSE estimators, a Monte Carlo study paralleling the previous study was performed for $t = 10$ and $t = 20$. In this Monte Carlo study, 1000 independent sets of pseudo random deviates $\{e_{ij}; j=1, \dots, n_i, i=1, \dots, t\}$ and $\{v_i; i=1, \dots, t\}$ were generated both according to three non-normal distributions viz: standard double exponential (symmetric long-tail), uniform (0,1) (short-tail) and standard exponential (positively skewed). Thus generated $\{e_{ij}\}$ and $\{v_i\}$ were appropriately centered and scaled to give zero means and variances 292 and 64 respectively as required. As before, pseudo observations $\{y_{ij}; i=1, \dots, t; j=1, \dots, n_i\}$ were obtained from the model (4.50). Results are given in Tables 4.3 (a and b), 4.4 (a and b) and 4.5 (a and b) for the exponential, double exponential and uniform distributions respectively. For each of the pairs of tables 'a' denotes results for $t = 10$, and 'b' denotes results for $t = 20$.

The main findings are summarized below (see pages 88-90 for the tables).

- (i) Under both exponential and double exponential distributions, the approximate mean square error $(\text{MSE}^*(t_1(\tilde{\sigma}^2, y)))$ exceeds the Monte Carlo estimate of the true MSE uniformly across all the small areas for

Both settings of t . The differences are highest when $t = 10$, and are particularly bad under the exponential distribution, ranging from 29 to 45%. For $t = 20$, this relative error is decreased, but still ranges between 13 and 22%. In the double exponential case, the percentage excess of $MSE^*(t_1(\hat{\sigma}^2, y))$ over true MSE ranges from 20 to 33 for $t = 10$, and from 7 to 18 for $t = 20$. In contrast to the above result, $MSE^*(t_1(\hat{\sigma}^2, y))$, the approximation to MSE is quite close to the true MSE values under the uniform distribution, for both $t = 10$ and $t = 20$.

(ii) The values of $Bias_J^*$ and $Bias_N^*$ show that, in estimating the MSE approximation $MSE^*(t_1(\hat{\sigma}^2, y))$, the jackknife estimator $mse_1^*(t_1(\hat{\sigma}^2, y))$ outperforms the normality based estimator $mse^*(t_1(\hat{\sigma}^2, y))$, under all three distributions, and for both settings of t . Under exponential data, $mse^*(t_1(\hat{\sigma}^2, y))$ underestimates by as much as 12% ($n_1 = 2$, $t = 20$) compared to a maximum underestimate of only 4% by the jackknife estimator $mse_1^*(t_1(\hat{\sigma}^2, y))$. When $t = 10$, the biases exhibited by both estimator are larger, being underestimates, for small n_1 , and overestimates for large n_1 . Again, however, the jackknife estimator is the better of the two. Under double exponential data similar trends are observed. In this case, when $t = 20$, the bias in $mse^*(t_1(\hat{\sigma}^2, y))$ ranges from -7% to 4%, while the jackknife competitor, $mse_1^*(t_1(\hat{\sigma}^2, y))$ exhibits an almost constant bias around -3%. Somewhat different trends are visible from Tables 4.5a and 4.5b for the uniform distribution. With few exceptions, the biases of both MSE^* estimators are positive. When $t = 10$, these biases increase monotonically from 1% and -1% to 30% and 17% for $mse^*(t_1(\hat{\sigma}^2, y))$ and $mse_1^*(t_1(\hat{\sigma}^2, y))$ respectively. When $t = 20$,

the bias in $mse^*(t_1(c^2, y))$ goes from -1 at $n_1 = 2$ to 9 at $n_1 = 6$, while the bias of its jackknife competitor remains small for all n_1 's.

(iii) From the values of $Bias_N$ and $Bias_J$ it can be seen that both estimators of MSE perform poorly under exponential and double exponential data. For exponential distribution case, the normality based estimator $mse^*(t_1(\tilde{c}^2, y))$ is slightly better than the jackknife estimator $mse^*(t_1(c^2, y))$ when $t = 20$, in which case, the bias ranges from 2% to 18% for $mse^*(t_1(\tilde{c}^2, y))$, and from 6% to 21% for $mse^*(t_1(c^2, y))$. These results are consistent with our earlier observation that the approximate MSE (MSE^*) exceeds the true MSE, and that the normality based MSE estimator underestimates MSE^* , more than does the jackknife estimator. From Tables 5.a and 5.b, we see that both estimators perform better for uniform data. When $t = 10$, the positive biases are still unacceptably large for $n_1 = 5$ and 6. However, when $t = 20$, both MSE estimators perform well, the magnitude of these biases being less than 5% for both estimators.

(iv) The last columns of Tables 4.3 (a and b), 4.4 (a and b) and 4.5 (a and b) display the bias of mse_D , the estimator of the variance of the BLUP. As in the case of normal data, we again see that this estimator seriously underestimates the variance, for all three distributions, for both $t = 10$ and $t = 20$.

2 OF/DE 2

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

Table 4.2a: Relative Error (%) of MSE approximation (MSE^*) and Relative Biases of MSE estimators with respect to MSE^* and MSE: ($t=10$, Normal distribution).

n_i	R.E. MSE^*	Bias $_N^*$	Bias $_J^*$	Bias $_N$	Bias $_J$	Bias $_b$
2	7	-2	0	5	7	-27
3	12	3	3	15	14	-28
4	7	10	8	18	15	-32
5	15	19	11	37	28	-30
6	7	29	16	37	24	-29

Table 4.2b: Relative Error (%) of MSE approximation (MSE^*) and Relative Biases of MSE estimators with respect to MSE^* and MSE: ($t = 20$, Normal distribution).

n_i	R.E. MSE^*	Bias $_N^*$	Bias $_J^*$	Bias $_N$	Bias $_J$	Bias $_b$
2	2	-2	0	0	2	-19
3	2	-1	2	1	4	-16
4	4	2	4	6	8	-16
5	-1	4	6	5	7	-23
6	3	8	9	11	12	-19

Table 4.3a: Relative Error (%) of MSE approximation (MSE^*) and Relative Biases of MSE estimators with respect to MSE^* and MSE: ($t = 10$, Exponential distribution).

n_1	R.E. MSE^*	$Bias_N^*$	$Bias_J^*$	$Bias_N$	$Bias_J$	$Bias_b$
2	29	-14	-6	11	21	-23
3	36	-10	-5	23	30	-25
4	45	-3	-2	41	43	-20
5	44	6	2	53	47	-23
6	36	16	6	58	44	-20

Table 4.3b: Relative Error (%) of MSE approximation (MSE^*) and Relative Biases of MSE estimators with respect to MSE^* and MSE: ($t = 20$, Exponential distribution).

n_1	R.E. MSE^*	$Bias_N^*$	$Bias_J^*$	$Bias_N$	$Bias_J$	$Bias_b$
2	13	-12	-4	2	6	-13
3	15	-10	-1	4	13	-14
4	19	-7	0	11	17	-14
5	22	-4	1	18	21	-18
6	18	0	1	18	19	-17

Table 4.4a: Relative Error (%) of MSE approximation (MSE^*) and Relative Biases of MSE estimators with respect to MSE^* and MSE: ($t = 10$, Double Exponential Distribution).

n_i	R.E.MSE*	Bias _N *	Bias _J *	Bias _N	Bias _J	Bias _b
2	20	-9	-3	10	17	-25
3	25	-5	-1	19	24	-25
4	33	2	3	37	38	-20
5	29	11	7	43	38	-27
6	25	21	11	51	39	-28

Table 4.4b: Relative Error (%) of MSE approximation (MSE^*) and Relative Biases of MSE estimators with respect to MSE^* and MSE: ($t = 20$, Double exponential Distribution).

n_i	R.E.MSE*	Bias _N *	Bias _J *	Bias _N	Bias _J	Bias _b
2	7	-7	-3	-1	3	-13
3	10	-6	-3	4	7	-15
4	18	-3	-3	13	13	-14
5	9	0	-4	11	7	-19
6	13	4	-3	17	10	-16

Table 4.5a: Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: ($t = 10$, Uniform Distribution).

n_i	R.E.MSE*	Bias _N *	Bias _J *	Bias _N	Bias _J	Bias _b
2	1	1	-1	2	0	-29
3	5	6	2	11	6	-32
4	-1	13	6	12	5	-35
5	7	21	11	31	19	-30
6	-1	30	17	29	15	-33

Table 4.5b: Relative Error (%) of MSE approximation (MSE*) and Relative Biases of MSE estimators with respect to MSE* and MSE: ($t = 20$, Uniform Distribution).

n_i	R.E.MSE*	Bias _N *	Bias _J *	Bias _N	Bias _J	Bias _b
2	1	-1	1	2	1	-20
3	-2	1	-1	-1	-1	-26
4	1	3	0	4	1	-26
5	-5	6	0	1	-5	-22
6	-3	9	1	5	-3	-22

CHAPTER 5

EXTENSION OF NESTED ERROR REGRESSION MODEL TO STRATIFICATION

All the results given in the previous chapters are based on the assumption that the samples drawn are simple random samples. However, in some cases, precision of the estimates can be increased through stratification or post-stratification (see Cochran, 1977).

In this Chapter, we extend the results, described in Chapters 2 and 4 on the nested error regression model to the case when the sample is a stratified (or post-stratified) sample. Before giving the mathematical model, we report below two such applications in the context of small area estimation, given by Walker and Sigman (1984) and Hidioglou et al. (1984).

The data set considered by Walker and Sigman (1984) is in the context of utilizing Landsat satellite data on the number of pixels (x) for predicting mean hectares (y) under different crops for six counties (small areas) in Eastern South Dakota. The sample of segments was drawn using stratified random sampling with land use (percentage of cultivated land) as the criterion of stratification. In the second data set, considered by Hidioglou et al. (1984), the object was to provide estimates for total wages and salaries (y) for different industrial groups for 18 census divisions (small areas) of Nova Scotia province, using the supplementary information on gross business income (x) (available on a 100% basis). The units (unincorporated tax filers)

in the sample were post-stratified according to gross business income groups (\$25K - \$50K, \$50K - \$150K and \$150K - \$500K). In the above examples, the strata can be thought of as being crossed with the small areas.

The nested error regression model with stratification is described in Section 5.1. In Section 5.2, the BLUP for the small areas under this model are obtained. The approximation to MSE of estimated BLUP is studied in Section 5.3, along the lines of Theorem 4.1.

5.1. The Model and Assumptions.

We consider the following model as an extension of the nested error regression model ((2.1) and (2.2)) to stratification:

$$y_{hij} = x_{hij}' \beta_h + v_{hi} + e_{hij} \quad (5.1)$$

$$v_{hi} = a_i + b_{hi}, \quad h=1, \dots, L; \quad i=1, \dots, t; \quad j=1, \dots, n_{hi}$$

where, for the j th sampled unit in the i th small area and the h th stratum, y_{hij} denotes the observed value of the study characteristic, $x_{hij} = \text{Col } (x_{hijk})$ is a vector of unknown parameters in the h th stratum, $1 \leq k \leq K$. Further, a_i , b_{hi} and e_{hij} are assumed to be mutually independent and all are independently distributed with means zero and variances σ_a^2 , $\sigma_h^2(b)$ and $\sigma_h^2(e)$ respectively. Under the above assumptions, it follows that the error covariance structure of the error term $u_{hij} = a_i + b_{hi} + e_{hij}$ is given by

$$E(u_{hij}, u_{h'i'j'}) = \begin{cases} \sigma_a^2 + \sigma_h^2(b) + \sigma_h^2(e), & \text{if } h=h', i=i', j=j' \\ \sigma_a^2 + \sigma_h^2(b), & \text{if } h=h', i=i', j \neq j' \\ \sigma_a^2, & \text{if } h \neq h', i=i', j \neq j' \\ 0, & \text{if } i \neq i'. \end{cases} \quad (5.2)$$

The model defined by (5.1) and (5.2) can be written as a general mixed linear model

$$y = x\beta + z\gamma + e \quad (5.3)$$

where

$$y = \text{Col}_{1 \leq h \leq L} \text{Col}_{1 \leq i \leq t} \text{Col}_{1 \leq j \leq n_{hi}} (y_{hij}) \quad (5.4)$$

$$x = \text{diag}_{1 \leq h \leq L} \text{Col}_{1 \leq i \leq t} \text{Col}_{1 \leq j \leq n_{hi}} (x_{hij}) \quad (5.5)$$

$$\beta = \text{Col}_{1 \leq h \leq L} (\beta_h); \quad \gamma = \text{Col}_{1 \leq h \leq L} \text{Col}_{1 \leq i \leq t} (\gamma_{hi}) \quad (5.6)$$

$$z = \text{diag}_{1 \leq h \leq L} \text{diag}_{1 \leq i \leq t} \text{Col}_{1 \leq j \leq n_{hi}} (1) \quad (5.7)$$

$$e = \text{Col}_{1 \leq h \leq L} \text{Col}_{1 \leq i \leq t} \text{Col}_{1 \leq j \leq n_{hi}} (e_{hij})$$

and

$$\text{Var}(y) = V = R + ZGZ' \quad (5.8)$$

with

$$R = E(ee') = \text{diag}_{1 \leq h \leq L} \text{diag}_{1 \leq i \leq t} \text{diag}_{1 \leq j \leq n_{hi}} (\sigma_h^2(e)) \quad (5.9)$$

$$G = \text{diag}_{1 \leq h \leq L} \text{diag}_{1 \leq i \leq t} (\sigma_h^2(b)) + \sigma_a^2 \text{diag}_{1 \leq i \leq t} (1, 1', \dots, 1, 1') \quad (5.10)$$

Now, the problem is to predict the small area means

$$\mu_i = \sum_h w_{hi} \mu_{hi}; \quad i=1, \dots, t, \quad (5.11)$$

where

$$w_{hi} = N_{hi} / N_{\cdot i}, \quad N_{\cdot i} = \sum_h N_{hi}$$

and

$$\mu_{hi} = \bar{x}_{hi} \beta_h + v_{hi} \quad (5.12)$$

with $\bar{x}_{hi} = N_{hi}^{-1} \sum_{j=1}^{N_{hi}} x_{hij}$ and also the population means $\bar{x}_{hi}$ and the sizes N_{hi} are assumed to be known. As in the simple nested error regression model, we first give in the next section the BLUP of μ_i . When the variance components $(\sigma_a^2, \sigma_h^2(b)$ and $\sigma_h^2(e))$ in (5.2) are known.

5.2. BLUP When the Variance Components are Known.

Using Theorem 2.1, for the general mixed linear model (2.11), we obtain below, the BLUP of μ_i under the model defined by (5.3)-(5.10), when the variance components, $\sigma_a^2, \sigma_h^2(b), \sigma_h^2(e)$ ($h=1, \dots, L$), are known.

Theorem 5.1. For the model defined by (5.1) and (5.2), let the variance components $\sigma^2 = (\sigma_a^2, \sigma_1^2(b), \dots, \sigma_L^2(b), \sigma_1^2(e), \dots, \sigma_L^2(e))$ be known. Then the BLUP of the i th small area mean

$$\mu_i = \sum_h w_{hi} \mu_{hi}$$

is given by

$$\tilde{\mu}_i(\sigma^2) = \sum_h w_{hi} \tilde{\mu}_{hi}(\sigma^2), \quad (5.13)$$

where

$$\begin{aligned} \tilde{\mu}_{hi}(\sigma^2) &= \bar{x}_{hi} \hat{\beta}_h + \gamma_{hi} (\bar{y}_{hi} - \bar{x}_{hi} \hat{\beta}_h) \\ &\quad + (1 - \gamma_{hi}) \sigma_a^2 (1 + \sigma_a^2 \sum_h \lambda_{hi})^{-1} (\bar{y}_{hi} - \bar{x}_{hi} \hat{\beta}_h), \end{aligned} \quad (5.14)$$

with

$$\hat{\beta} = \text{Col}_{1 \leq h \leq L} (\hat{\beta}_h) = (X' V^{-1} X)^{-1} X' V^{-1} y, \quad (5.15)$$

$$\lambda_{hi} = (\sigma_h^2(b) + n_{hi}^{-1} \sigma_h^2(e))^{-1}; \quad \gamma_{hi} = \sigma_h^2(b) \lambda_{hi}. \quad (5.16)$$

Furthermore,

$$E[\tilde{\mu}_i(\sigma^2) - \mu_i]^2 = \sum_{h=1}^L \sum_{h'=1}^L w_{hi} w_{h'i} E(\mu_{hi}(\sigma^2) - \mu_{hi})(\mu_{h'i}(\sigma^2) - \mu_{h'i}), \quad (5.17)$$

where

$$\begin{aligned} E(\mu_{hi}(\sigma^2) - \mu_{hi})(\mu_{h'i}(\sigma^2) - \mu_{h'i}) &= \{ \bar{x}_{hi} - \gamma_{hi} \bar{x}_{hi} - \alpha_i (1 - \gamma_{hi}) \sum_{h=1}^L \lambda_{hi} \bar{x}_{hi} \} \text{Cov}(\hat{\beta}_h, \hat{\beta}_{h'}) \{ \bar{x}_{h'i} - \gamma_{h'i} \bar{x}_{h'i} - \alpha_i (1 - \gamma_{h'i}) \sum_{h=1}^L \lambda_{h'i} \bar{x}_{h'i} \}' \\ &\quad + \delta_{hh'} (1 - \gamma_{hi}) \sigma_h^2(b) \\ &\quad + \alpha_i \gamma_{hi} \gamma_{h'i} n_{hi}^{-1} n_{h'i}^{-1}, \end{aligned} \quad (5.18)$$

with

$$\text{Var}(\hat{\beta}) = (X' V^{-1} X)^{-1}$$

$$a_i = \frac{1}{a} (1 + \frac{1}{a} \lambda_{hi})^{-1} \quad (5.19)$$

and $hh' = 1$ if $h = h'$, $hh' = 0$ otherwise.

Proof of Theorem 5.1. The following are the key steps involved in the proof:

Step 1: For the matrix V defined by (5.8), V^{-1} is obtained by using the result

$$V^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}, \quad (5.20)$$

where R and Z are respectively defined by (5.9) and (5.10), and then applying Lemma 2.1 to compute G^{-1} .

The result (5.20) can be verified by noting that $VV^{-1} = I$.

Step 2: The small area means are expressed in the form

$$\mu_{1i} = k_i' \bar{x} + m_i' \bar{v}, \quad (5.21)$$

where

$$k_i = \text{Col}_{1 \leq h \leq L} (\bar{x}_{hi}' w_{hi}), \quad m_i = \text{Col}_{1 \leq h \leq L} \text{Col}_{1 \leq i' \leq t} (\bar{v}_{hi}' w_{hi}). \quad (5.22)$$

Step 3: The desired results (5.13) and (5.17) follow respectively from (2.14) and (2.17), by taking $\bar{y}$, X , Z , R , G , k_i and m_i as respectively defined by (5.4), (5.5), (5.7), (5.9), (5.10), and (5.22), and using the steps 1 and 2.

5.3. Approximation to MSE of the Estimated BLUP.

When the variance components are unknown the BLUP $\hat{\mu}_i(\sigma^2)$ is estimated by replacing the unknown variance components σ^2 by their respective

asymptotically consistent estimators $\hat{\sigma}^2$. Theorem 5.2 below provides an expression for the estimated BLUP $\hat{\mu}_i = \hat{\mu}_i(\hat{\sigma}^2)$ and its approximate MSE, under the assumption of normality for the random errors.

The following regularity conditions are required to establish

Theorem 5.2:

(C1) The elements of the matrix X are assumed to be uniformly bounded such that

- (a) $\lim_{t \rightarrow \infty} t(X'V^{-1}X)^{-1} = G$ and G^{-1} exists,
- (b) $\lim_{t \rightarrow \infty} t(X'\hat{V}^{-1}X)^{-1} = S$ and S^{-1} exists almost surely

and

- (c) $\lim_{t \rightarrow \infty} t(X' \frac{\partial V^{-1}}{\partial \sigma^2} X) = H(\sigma^2)$,

where $\hat{V}^{-1}$ is obtained by replacing σ^2 in V^{-1} by $\hat{\sigma}^2$, and the elements of $H(\sigma^2)$ are continuous functions of σ^2 .

(C2) The sequence $\{n_{hi}\}$ is such that $\sup_{h,i} n_{hi} < \infty$, where

$$n_{.1} = \sum_{h=1}^L n_{hi}, \text{ and } L \text{ is bounded.}$$

(C3) For the variance components σ_a^2 , $\sigma_h^2(b)$ and $\sigma_h^2(e)$, translation invariant and asymptotically consistent estimators $\hat{\sigma}_a^2$, $\hat{\sigma}_h^2(b)$ and $\hat{\sigma}_h^2(e)$ are of the form:

$$\hat{\sigma}_a^2 = y' C_a y$$

$$\hat{\sigma}_h^2(b) = y' C_h(b) y$$

and

$$\hat{\sigma}_h^2(e) = y' C_h(e) y$$

with C_a , $C(b)$ and $C(e)$ of the form C defined in Lemma 4.4(c).

Remark 5.2. It can be shown that the method of MINQUE (described in section 6.2) yields estimators of variance components satisfying the condition (C3).

Theorem 5.2: For the model defined by (5.1) and (5.2), let the regularity conditions C1-C3, and normality assumption for the errors ϵ_{hi} , b_{hi} and e_{hi} be true. Then the mean square error of the estimated BLUP

$$u_i = \frac{1}{n} \sum_{h=1}^n w_{hi} u_{hi} \quad (5.23)$$

can be approximated by

$$\begin{aligned} \text{MSE}(u_i) &= E(u_i - \hat{u}_i)^2 \\ &= E(u_i - \hat{u}_i)^2 + \frac{1}{n} \text{tr}(\gamma_{hh} \gamma_{hh}^{-1}(i)) + o(t^{-1}), \end{aligned} \quad (5.24)$$

where $E(u_i - \hat{u}_i)^2$ is defined by (5.17) and

$$\gamma_{hh} = E(\epsilon_h^2 - \sigma_h^2)(\epsilon_h^2 - \sigma_h^2)' \quad (5.25)$$

$$\gamma_{hh}^{-1}(i) = E\left(\frac{\partial \hat{u}_{hi}}{\partial \sigma_h^2}\right)\left(\frac{\partial \hat{u}_{hi}}{\partial \sigma_h^2}\right)' \quad (5.26)$$

with

$$\sigma_h^2 = (\sigma_a^2, \sigma_b^2, \sigma_e^2) \quad (5.27)$$

Since the proof of Theorem 5.2 is similar to the proof of Theorem 4.1, only key steps are outlined below.

Key steps of the Proof of Theorem 5.2:

Step 1. From normality assumption and hence using Lemma 4.5, we have

$$\begin{aligned} & E[\hat{\mu}_{hi}(\sigma^2) - \mu_{hi}][\hat{\mu}_{h,i}(\hat{\sigma}^2) - \mu_{h,i}] \\ &= E[\hat{\mu}_{hi}(\sigma^2) - \mu_{hi}][\mu_{h,i}(\sigma^2) - \mu_{h,i}] \\ &+ E[\mu_{hi}(\sigma^2) - \hat{\mu}_{hi}(\sigma^2)][\hat{\mu}_{h,i}(\hat{\sigma}^2) - \hat{\mu}_{h,i}(\sigma^2)] \quad (5.27) \end{aligned}$$

Step 2: By Taylor's expansion of $\hat{\mu}_{hi}(\hat{\sigma}^2)$ about the point σ^2 , we get

$$\begin{aligned} & E[\hat{\mu}_{hi}(\hat{\sigma}^2) - \hat{\mu}_{hi}(\sigma^2)][\hat{\mu}_{h,i}(\hat{\sigma}^2) - \hat{\mu}_{h,i}(\sigma^2)] \\ &= E[(\hat{\sigma}_h - \sigma_h)' d_{hi}][(\hat{\sigma}_h - \sigma_h)' d_{h,i}] \\ &+ o(t^{-1}), \quad (5.28) \end{aligned}$$

where $\hat{\sigma}_h = (\hat{\sigma}_a^2, \hat{\sigma}_h^2(b), \hat{\sigma}_h^2(e))$ and

$$d_{hi}(\sigma^2, y) = \left(\frac{\partial \hat{\mu}_{hi}(\sigma^2)}{\partial \sigma_a^2}, \frac{\partial \hat{\mu}_{hi}(\sigma^2)}{\partial \sigma_h^2(b)}, \frac{\partial \hat{\mu}_{hi}(\sigma^2)}{\partial \sigma_h^2(e)} \right)$$

Step 3: The first order derivatives of $\hat{\mu}_{hi}(\sigma^2)$ with respect to σ_a^2 , $\sigma_h^2(b)$ and $\sigma_h^2(e)$ are expressed respectively as $P_{hi}'u$, $P_{hi}'(b)u$ and $P_{hi}'(e)u$, where

$$u = \begin{pmatrix} \text{Col}_{1 \leq h \leq L} & \text{Col}_{1 \leq i \leq t} & \text{Col}_{1 \leq j \leq n_{hi}} \end{pmatrix} (u_{hij})$$

with $u_{hij} = a_i + b_{hi} + e_{hij}$

and $P_{hi}(b)$ and $P_{hi}(e)$ are of the form $P(i)$ defined in Lemma 4.4(d).

• Step 4: Using Lemmas 4.3 and 4.4, it follows

$$E[(\hat{e}_h - e_h) d_{hi}(g^2, y)] E[(\hat{e}_{h'} - e_{h'}) d_{h'1}(g^2, y)] \\ = \text{tr } \Psi_{hh'} + o(t^{-1}) \quad (5.30)$$

where

$$\Psi_{hh'} = E[(\hat{e}_h - e_h)(\hat{e}_{h'} - e_{h'})'] \\ 2L \times 2L$$

$$\Psi_{hh'}(i) = E(d_{hi} d_{h'1}') \quad \square \\ 2L \times 2L$$

CHAPTER 6

ESTIMATION OF CORRELATED RESPONSE VARIANCE (CRV)

This chapter deals with the optimal estimation of the interviewer error variance, in a model based framework. We confine ourselves to the case where the interviewers are the only source of response errors, which are in general correlated. Hence, throughout this chapter, the interviewer error variance will be referred to as the correlated response variance (CRV). The results of this chapter, however, can be extended to the case where other sources of errors (e.g., coders errors) are also present.

Since the survey process cannot usually be repeated, a technique of interpenetration proposed by Mahalanobis (1946) is widely used for estimating CRV. In this technique, the sample of size n under consideration is divided at random into K subsamples of equal sizes, where K is the number of interviewers assigned. These subsamples are then assigned to the K interviewers so that there is no correlation between errors of measurement for different interviewers. Owing to the cost consideration, a part of the sample will be subjected to interpenetration. For example, in the Canadian 1971 census response variance Study, interpenetration was used in approximately 375 pairs of contiguous enumeration areas (EAs) chosen from approximately 40,000 enumeration areas across the country. Hansen, Hurwitz and Bershad (1959) provide an estimator for CRV, based only on the data obtained through interpenetration. Later, Fellegi (1974) obtained two estimators of CRV:

The first estimator is similar to the Hansen, Hurwitz and Berstad (1959) estimator based only on the interpenetrating data set, while the second estimator is based on the work of all the interviewers (those working in the interpenetrated assignments as well as the other interviewers).

This chapter is structured as follows: In Section 6.1, a design for interviewer allocation and a linear additive model for interviewer errors is described. The method of minimum norm quadratic unbiased estimation (MINQUE), which is used to derive the optimal estimators of CRV, is described in Section 6.2. Section 6.3 provides the optimal estimators of CRV based on the interpenetration data set only. In Section 6.4 the optimal estimator of CRV based on the entire data set is derived. The results of a simulation study on the relative efficiency of different estimators are reported in Section 6.5. The increase in efficiency by using the non-interpenetrated data for estimating the CRV is also evaluated.

6.1 Interviewer Allocation Scheme and the Model

A. Interviewer allocation scheme: Let S denote a sample of $2P$ EAs selected according to some sample design. Within each selected EA, let $2n$ households (h.h.) be selected according to simple random sampling (SRS) without replacement. From the population of enumerators, a sample of $2P$ enumerators is selected randomly, and each enumerator is initially assigned to enumerate one EA. Then the neighbouring assignments are arbitrarily paired into P pairs. For interpenetration, a simple random sample of p pairs denoted by C is selected from the

P pairs: interpenetration is carried out by splitting the workload equally in each EA of the pair and then assigning them to the two enumerators, who were initially assigned. Thus, for the interpenetrating pairs, interviewers are assigned as in a cross-over design. For the remaining $P-p$ pairs of EAs (denoted by C^c), the initial interviewer assignments is retained. A schematic representation of the interviewer assignment, along the lines of Dodds and Smith (1973), is given below.

EA pair in C				EA pair in C^c			
Inter-viewer	EA		Total number of h.hs	Inter-viewer	EA		Total number of h.hs
	1	2			1	2	
1	n	n	$2n$	1	$2n$	0	$2n$
2	n	n	$2n$	2	0	$2n$	$2n$
	$2n$	$2n$	$4n$		$2n$	$2n$	$4n$

The data corresponding to all EAs in C and C^c are respectively termed as Data Set A and Data Set B respectively for convenience.

b. The Model: For the m^{th} pair ($m=1, \dots, P$), let y_{mijk} be the k^{th} observation in the j^{th} EA ($j=1,2$) recorded by the i^{th} interviewer ($i=1,2$). Also, for the i^{th} interviewer assigned to the m^{th} pair, let a_{mi} , ϵ_{mijk} and η_{mijk} be the systematic interviewer error common to all the units assigned to the interviewer, the elementary interviewer error associated with the k^{th} unit of the j^{th} EA and the respondent error of the k^{th} unit of the j^{th} EA. It is assumed (as in Hartley and

Rao (1978) that (i) $\{a_{mi}; i=1,2; m=1,\dots,P\}$ are i.i.d. random variables with zero mean and variance σ_a^2 (ii) the errors a_{mi} are independent of e_{mijk} where $e_{mijk} = \epsilon_{mijk} + \eta_{mijk}$ are i.i.d. random variables with mean zero and variance σ_e^2 . We also denote μ_{mj} as the true value for the j th EA of the m th pair. Under the above assumptions, an additive linear model is given by

$$y_{mijk} = \mu_{mj} + a_{mi} + e_{mijk} \quad (6.1)$$

where

$$E(a_{mi}) = 0; \quad E(a_{mi}^2) = \sigma_a^2$$

and

$$E(e_{mijk}) = 0; \quad E(e_{mijk}^2) = \sigma_e^2$$

with $i, j = 1, 2$; $k=1,\dots,n$ for $m=1,\dots,p$ and $i = j = 1, 2$;

$k = 1, \dots, 2n$ for $m = p+1, \dots, P$. Note that in the non-interpenetrated pairs, the interviewers labelled 1 and 2 are assigned to the EAs labelled 1 and 2 respectively.

6.2. Minimum Norm Quadratic Unbiased Estimation (MINQUE)

Rao and Kleffe (1980) provide a detailed description of the MINQUE procedure. A brief summary of MINQUE method is provided here. Consider a general mixed linear model written in the form

$$Y = X\beta + U_1 e_1 + \dots + U_{k-1} e_{k-1} + e_k \quad (6.2)$$

where Y is the $n \times 1$ vector of observations, $X, U_1, \dots, U_{k-1}$ are the

matrices of known constants, β is a $p \times 1$ vector of unknown parameters

and $e_1, \dots, e_k$ are the vectors of independent random variables with

zero means and variances $\sigma_1^2, \dots, \sigma_k^2$ respectively. The objective is to estimate the variance components $\{\sigma_i^2\}$. The MINQUE estimator $\hat{\sigma}^2 = \text{col}_{1 \leq i \leq k} (\hat{\sigma}_i^2)$ is obtained as a solution of "MINQUE-equations":

$$Z \hat{\sigma}^2 = q, \quad (6.3)$$

where the (ℓ, ℓ') th element $z_{\ell\ell'}$ of the $k \times k$ matrix Z and the ℓ th element q_ℓ of the $k \times 1$ vector q are respectively defined below.

$$z_{\ell\ell'} = ||U'_\ell (M V M)^+ U_{\ell'}||^2, \quad (6.4)$$

with $U_k = I$ and

$$q_\ell = ||U'_\ell (M V M)^+ y||^2, \quad \ell, \ell' = 1, \dots, k, \quad (6.5)$$

where $||B||^2$ = the sum of squares of all the elements of the matrix B ,

X_0^+ is the Moore-Penrose inverse of X (see Rao and Mitra, 1971),

$M = I - XX^+$ and

$$V = \alpha_1 U_1 U_1' + \dots + \alpha_{k-1} U_{k-1} U_{k-1}' + I, \quad (6.6)$$

with $\{\alpha_i\}$ are the initial guesses or prior values for the ratios

$$\sigma_i^2 / \sigma_k^2, \quad i=1, \dots, k-1.$$

Regarding the choices of $\alpha' = (\alpha_1, \dots, \alpha_{k-1})'$ we can take α_i as the initial guesses of the ratios σ_i^2 / σ_k^2 (following Rao and Kleffe (1980)). This leads to a locally minimum variance unbiased estimator $\hat{\sigma}^2$ (denoted by MINQUE $(\alpha, 1)$) in the class of all quadratic unbiased estimators provided y is normally distributed and $\alpha' = (\alpha_1, \dots, \alpha_{k-1})'$

is equal to $(\sigma_1^2 / \sigma_k^2, \dots, \sigma_{k-1}^2 / \sigma_k^2)$. Further, the variance of the locally best quadratic unbiased estimator $\hat{\sigma}_i^2$, under normality, is given by

$$\text{var}(\hat{\beta}_1^2) = 2\hat{\beta}_1^{(i)'} Z^{-1} \hat{\beta}_1^{(i)} \quad (6.7)$$

where $\hat{\beta}_1^{(i)} = \text{Col}_{1 \leq h \leq k} (\hat{\beta}_{hi})$.

A major problem that has dissuaded many potential MINQUE users is the fact that its computation involves inverting a $n \times n$ matrix, where n is the sample size. Kleffe and Seifert (1983) give a simple algorithm for the computation of MINQU estimates, which requires only inversion and storage of square matrices of order C_i , C_i being the number of columns in U_i and independent of n . These computations are further reduced if $U_i U_i'$ is diagonal.

6.3. Optimal estimator of σ_a^2 : Data Set A

In this section, we obtain MINQUE $(\alpha, 1)$ of σ_a^2 from Data Set A only with $\alpha = \sigma_a^2 / \sigma_e^2$. The following steps are involved in the derivation:

Step 1: First we write the model (6.1), only for the p pairs which are in C , in matrix notation as

$$y^{(1)} = X^{(1)} \alpha^{(1)} + U_1 a^{(1)} + e^{(1)} \quad (6.8)$$

where

$$y^{(1)} = \begin{matrix} \text{Col} & \text{Col} & \text{Col} & \text{Col} \\ 1 \leq m \leq p & 1 \leq i \leq 2 & 1 \leq j \leq 2 & 1 \leq k \leq n \end{matrix} (y_{mijk}),$$

$$X^{(1)} = \begin{matrix} \text{diag} & \text{Col} & \text{Col} & \text{Col} \\ 1 \leq m \leq k & 1 \leq i \leq 2 & 1 \leq j \leq 2 & 1 \leq k \leq n \end{matrix} (1),$$

$$U_1 = \begin{matrix} \text{diag} & \text{diag} & \text{Col} & \text{Col} \\ 1 \leq m \leq p & 1 \leq i \leq 2 & 1 \leq j \leq 2 & 1 \leq k \leq n \end{matrix} (1),$$

$$a^{(1)} = \underset{1 \leq m \leq p}{\text{Col}} \underset{1 \leq j \leq 2}{\text{Col}} (\alpha_{mj})$$

$$a^{(1)} = \underset{1 \leq m \leq p}{\text{Col}} \underset{1 \leq i \leq 2}{\text{Col}} (a_{mi})$$

and

$$e^{(1)} = \underset{1 \leq m \leq p}{\text{Col}} \underset{1 \leq i \leq 2}{\text{Col}} \underset{1 \leq j \leq 2}{\text{Col}} \underset{1 \leq k \leq n}{\text{Col}} (e_{mijk}).$$

Also, we have,

$$E(a^{(1)} a^{(1)}) = \sigma_a^2 I_{2p}$$

and

$$E(e^{(1)} e^{(1)}) = \sigma_e^2 I_{4np}$$

Step 2: By taking $k = 2$, $U_2 = I$, $M = I - X^{(1)} X^{(1)'} / V$, $V = \alpha U_1' U_1 + 1$,

we obtain the "MINQUE-equations"

$$z_{11} \sigma_a^2 + z_{12} \sigma_e^2 = q_1$$

$$z_{21} \sigma_a^2 + z_{22} \sigma_e^2 = q_2$$

(6.9)

where $z_{\ell\ell'}$ and q_ℓ , $\ell, \ell' = 1, 2$ are as defined by (6.4) and (6.5).

To compute $z_{\ell\ell'}$ and q_ℓ , we employ the algorithm given by Kleffe and Seifert (1983). Since the details of the computations are somewhat lengthy, only the final expressions are provided below:

$$\begin{aligned} q_1 &= ||U_1' (MVM)^+ y^{(1)}||^2 \\ &= 4n^2 (1+2n\alpha)^{-2} \sum_{m=1}^p \sum_{i=1}^2 (\bar{y}_{mi...} - \bar{y}_{m...})^2, \end{aligned}$$

$$\begin{aligned}
 q_2 &= ||U_2'(MVM)^+ y^{(1)}||^2 \\
 &= \sum_{m=1}^p \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^n \{y_{mijk} - \bar{y}_{m..j} - \\
 &\quad 2n\alpha(1+2n\alpha)^{-1}(\bar{y}_{mi..} - \bar{y}_{m...})\}^2,
 \end{aligned}$$

$$\begin{aligned}
 z_{11} &= ||U_1'(MVM)^+ U_1||^2 \\
 &= p(1+2n\alpha)^{-2} \{1 + (4n-3)(1+2n\alpha)^2\},
 \end{aligned}$$

$$z_{12} = z_{21} = ||U_1'(MVM)^+ U_2||^2 = 2np(1+2n\alpha)^{-2},$$

and

$$z_{22} = ||U_2'(MVM)^+ U_2||^2 = 4n^2 p(1+2n\alpha)^{-2}.$$

where $\bar{y}_{mi..}$, $\bar{y}_{m..j}$ and $\bar{y}_{m...}$ are the average values of y over the missing subscripts.

Step 3: Substituting the above quantities in (6.9) and solving for $\hat{\sigma}_a^2$, we get MINQUE $(\lambda, 1)$ of σ_a^2 , given by,

$$\begin{aligned}
 \hat{\sigma}_a^2 &= 2(2n-1)(4n-3)^{-1} p^{-1} \sum_{m=1}^p \sum_{i=1}^2 (\bar{y}_{mi..} - \bar{y}_{m...})^2 \\
 &\quad - \{2np(4n-3)\}^{-1} \sum_{m=1}^p \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^n (y_{mijk} - \bar{y}_{m..j})^2. \quad (6.10)
 \end{aligned}$$

Further, if a_{mi} and e_{mijk} are normally distributed, then the variance of $\hat{\sigma}_a^2$ is given by

$$\text{var}(\hat{\sigma}_a^2) = \{2n^2 p(4n-3)\}^{-1} \{1 + (4n-3)(1+2n\alpha^4/\sigma_e^4)\}. \quad (6.11)$$

Remark 6.1. 1. Under normality, the estimator $\hat{\sigma}_a^2$ is the best quadratic unbiased estimator of σ_a^2 since $\hat{\sigma}_a^2$ is independent of $\bar{y}_{m.}$ and MINQUE (0,1) is locally optimum under normality.

2.. Fellegi (1974) obtained an estimator of σ_a^2 using Data Set A only, which is identical to $\hat{\sigma}_a^2$ given by (6.10). But his derivation based on ANOVA Procedure does not establish the optimality of the estimator.

Since, within each interpenetrated pair, two interviewers are assigned as in cross-over design, it follows that (cf. Dodds and Smith (1973))

$$E\{2^{-1}(\bar{y}_{m1j.} - \bar{y}_{m2j.})^2 - \sum_{i=1}^2 \sum_{k=1}^n (y_{mijk} - \bar{y}_{mij.})^2 / (2n(n-2))\} \\ = \sigma_a^2; \quad m=1, \dots, p; \quad j=1, 2,$$

which leads to another estimator of σ_a^2 , given by

$$\hat{\sigma}_a^2(A) = (2p)^{-1} \sum_{m=1}^p \sum_{j=1}^2 \left[2^{-1}(\bar{y}_{m1j.} - \bar{y}_{m2j.})^2 - \sum_{i=1}^2 \sum_{k=1}^n (y_{mijk} - \bar{y}_{mij.})^2 / (2n(n-1)) \right] \quad (6.12)$$

6.4. Optimal Estimator of σ_a^2 : Data Set A and Data Set B combined

In the previous section, the optimal estimator of σ_a^2 was obtained from Data Set A only by conditioning on the p pairs of EAS in C . However, conditionally on C , Data Set B provides no additional information on σ_a^2 in the sense that MINQUE (0,1) for the combined model is identical to MINQUE (0,1) for Data Set A only. This can be demonstrated

along the lines of Bremer and Strokes (1985). We therefore formulate an unconditional model for the combined data and then obtain the MINQUE (0,1) of σ_a^2 . The unconditional model is obtained by making an additional assumption that μ_{mj} ($m=1, \dots, P, j=1, 2$) are independent random variables with means μ_j and variance σ_b^2 , which is expressed as

$$\mu_{mj} = \mu_j + b_{mj}; \quad m=1, \dots, P; \quad j=1, 2, \quad (6.13)$$

where $E(b_{mj}) = 0$ and $V(b_{mj}) = \sigma_b^2$. Thus the unconditional model, obtained by substituting (6.13) in (6.1), is

$$y = X\beta + U_1b + U_2a + e, \quad (6.14)$$

where

$$y = (y^{(1)}, y^{(2)})'$$

with

$$y^{(1)} = \begin{matrix} \text{Col} & \text{Col} & \text{Col} & \text{Col} \\ 1 \leq m \leq p & 1 \leq i \leq 2 & 1 \leq j \leq 2 & 1 \leq k \leq n \end{matrix} (y_{mijk});$$

$$y^{(2)} = \begin{matrix} \text{Col} & \text{Col} & \text{Col} \\ p+1 \leq m \leq P & 1 \leq i=j \leq 2 & 1 \leq k \leq 2n \end{matrix} (y_{mijk});$$

$$X = \begin{pmatrix} X_{11} & 0 \\ 0 & X_{22} \end{pmatrix}$$

$$\text{with } X_{11} = \begin{matrix} \text{Col} & \text{Col} & \text{diag} & \text{Col} \\ 1 \leq m \leq p & 1 \leq i \leq 2 & 1 \leq j \leq 2 & 1 \leq k \leq n \end{matrix} (1); \quad X_{22} = \begin{matrix} \text{Col} & \text{diag} & \text{Col} \\ p+1 \leq m \leq P & 1 \leq i=j \leq 2 & 1 \leq k \leq 2n \end{matrix} (1);$$

$$\beta = \begin{matrix} \text{Col} \\ 1 \leq j \leq 2 \end{matrix} (\beta_j), \quad b = \begin{matrix} \text{Col} & \text{Col} \\ 1 \leq m \leq P & 1 \leq j \leq 2 \end{matrix} (b_{mj});$$

$$U_2 = \begin{pmatrix} U_{21} & 0 \\ 0 & U_{22} \end{pmatrix}$$

With $U_{21} = \text{diag diag Col Col } (1); U_{22} = U_{12};$
 $1 \leq m \leq p \quad 1 \leq i \leq 2 \quad 1 \leq j \leq 2 \quad 1 \leq k \leq n$

$$a = \text{Col Col } (a_{mi}) \text{ and } e = (e^{(1)'}, e^{(2)'})',$$

$$1 \leq m \leq p \quad 1 \leq i \leq 2$$

$$\text{with } e^{(1)} = \text{Col Col Col Col } (e_{mijk}^{(1)});$$

$$1 \leq m \leq p \quad 1 \leq i \leq 2, 1 \leq j \leq 2 \quad 1 \leq k \leq n$$

$$e^{(2)} = \text{Col Col Col } (e_{mijk}^{(2)}).$$

$$p+1 \leq m \leq P \quad 1 \leq i=j \leq 2 \quad 1 \leq k \leq 2n$$

Then the MINQUE $(0,1)$ of a^2 under the combined model (6.14) is

obtained by computing the "MINQUE equations" (6.3) with $k=1, v_1 = \frac{2}{p+1},$

$$U_3 = I, \quad \alpha_2^2 = \frac{2}{a^2} e^2 \quad \text{and}$$

$$V = U_1 U_1' + a_2 U_2 U_2' + I.$$

The expressions for the elements of the coefficient matrix Z and the components of the vector q of the MINQUE-equations are derived using the Kleffe-Seifert algorithm. The details of the derivation are quite involved and we therefore provide below only the final expressions.

Without loss of generality, we denote the observations on $y : \{y_{mijk}; m=1, \dots, p; i, j = 1, 2; k = 1, \dots, n\}$ from Data Set A and the observations on $\tilde{y} : \{y_{mijk}; m = p+1, \dots, P; i = j = 1, 2; k=1, \dots, 2n\}$ from Data Set B respectively by $\{y_{mijk}^{(1)}; m = 1, \dots, p; i, j = 1, 2; k = 1, \dots, n\}$ and $\{y_{mijk}^{(2)}; m = p+1, \dots, P; i = j = 1, 2; k = 1, \dots, 2n\}.$

Then, using (6.5)

$$q_1 = \|U_3^*(MVM)^+Y\|^2 = \|U_3^*(MVM)^+Y\|^2$$

$$\sum_{m=1}^P \sum_{i=1}^2 \left\{ \sum_{j=1}^2 \sum_{k=1}^n t_{mijk}^{(1)} \right\}^2 + \sum_{m=p+1}^P \sum_{j=1}^2 \left\{ \sum_{k=1}^{2n} t_{mjk}^{(2)} \right\}^2$$

$$I_2 = \|U_2^*(MVM)^+Y\|^2$$

$$= \sum_{m=1}^P \sum_{i=1}^2 \left\{ \sum_{j=1}^2 \sum_{k=1}^n t_{mijk}^{(1)} \right\}^2 + \sum_{m=p+1}^P \sum_{j=1}^2 \left\{ \sum_{k=1}^{2n} t_{mjk}^{(2)} \right\}^2$$

and

$$q_1 = \|U_1^*(MVM)^+Y\|^2$$

$$= \sum_{m=1}^P \sum_{j=1}^2 \left\{ \sum_{i=1}^2 \sum_{k=1}^n t_{mijk}^{(1)} \right\}^2 + \sum_{m=p+1}^P \sum_{j=1}^2 \left\{ \sum_{k=1}^{2n} t_{mjk}^{(2)} \right\}^2$$

where

$$t_{mijk}^{(1)} = y_{mijk}^{(1)} - b_1 y_{m..j}^{(1)} - b_2 y_{m1..}^{(1)} + b_3 y_{m...}^{(1)}$$

$$- b_4 y_{..j..}^{(2)} - b_5 y_{...j.}^{(1)} + b_6 y_{....}^{(1)} + b_7 y_{.....}^{(2)}$$

and

$$t_{mijk}^{(2)} = y_{mijk}^{(2)} - c_1 y_{m..j}^{(2)} - c_2 y_{..j.}^{(2)} - c_3 y_{...j.}^{(1)} + c_4 y_{....}^{(2)} - c_5 y_{.....}^{(1)}$$

with

$$h = 1/(1+2na_1+2na_2), \quad \bar{h} = (1+2na_1+2na_2)(1+2na_1)^{-1}$$

$$c = (s-1)[2ps+2(P-p)]^{-1}, \quad d = 2c(P-p)[2P^{-1}-(ps)^{-1}]$$

$$b_1 = (1-hs)/2n, \quad b_2 = \frac{h}{2(1+2na_2)}^{-1},$$

$$b_3 = b_2/2, \quad b_4 = h(1+2dp)/2nP,$$

$$b_5 = h[Ps-(P-p)(1+2dp)]/2npP,$$

$$b_6 = [(s-1)h-2d(P-p)]/2nP, \quad b_7 = hdp(P-p)/2npP,$$

$$c_1 = (1-h)/2n, \quad c_2 = h(1+2cp)/2nP,$$

$$c_3 = h[1-2c(P-p)]/2nP; \quad c_4 = hc(P-p)/2nP,$$

$$c_5 = hcp/2nP.$$

Further, using (6.4)

$$\begin{aligned} z_{11} &= ||U_1'(MVM)^+U_1||^2 \\ &= 4n^2h^2P^{-1} [2P(P-1) + P(p-1)(s^2-1) - 4p(P-p)(a-b)], \end{aligned}$$

$$\begin{aligned} z_{12} &= ||U_1'(MVM)^+U_2||^2 \\ &= 4n^2h^2P^{-2} [P(P-1)(2P-p) - 4p(P-p)^2(a-b)], \end{aligned}$$

$$z_{13} = 2nh^2 [2(P-1) + (p-1)s^2 - 1] - 4p(P-p)(a-b)/P,$$

$$z_{22} = 4nh^2P^{-2} [2P^3 - 2P(2P-p) + (P-p)^2 + p^2] - 4n^2P_2(1+2na_2)^{-2} - 4n^2((1-1)a^2)],$$

$$z_{23} = 4nh^2P^{-1} [(2P-p)(P-1) - 4p(P-p)a] + 4np(1+2na_2)^{-2}$$

and

$$z_{33} = 2[(2n-1)P + h^2(sp+P-p-1)] - 4p(P-p)(a-b)/P + (s^2-1)h^2(p-1).$$

where

$$a = c - c^2(P-p) \quad \text{and} \quad b = d + d^2p.$$

The MINQUE $(\alpha_1, \alpha_2, 1)$ of σ_a^2 is now obtained by solving the following three linear equations

$$\begin{aligned} z_{11}\hat{\sigma}_b^2 + z_{12}\hat{\sigma}_a^2 + z_{13}\hat{\sigma}_e^2 &= q_1 \\ z_{21}\hat{\sigma}_b^2 + z_{22}\hat{\sigma}_a^2 + z_{23}\hat{\sigma}_e^2 &= q_2 \\ z_{31}\hat{\sigma}_b^2 + z_{32}\hat{\sigma}_a^2 + z_{33}\hat{\sigma}_e^2 &= q_3. \end{aligned} \quad (6.15)$$

Under the assumption of normality for the errors b , a and e , the MINQUE $(\alpha_1, \alpha_2, 1)$ of σ_a^2 is locally optimum in the class of quadratic unbiased estimators of σ_a^2 . The variance of $\hat{\sigma}_a^2$ at $\sigma_b^2/\sigma_e^2 = \alpha_1$ and $\sigma_e^2/\sigma_a^2 = \alpha_2$ is given by (see eq. 6.7)

$$\text{Var}(\hat{\sigma}_a^2) = 2\sigma_a^4(2)'Z^{-1}(2)\sigma_e^{-4} \quad (6.16)$$

where $(2)' = (0, 1, 0)$ and Z^{-1} is the inverse of the coefficient matrix in (6.15).

Remark 6.2. 1. If the prior guesses of σ_b^2/σ_e^2 and σ_a^2/σ_e^2 are not available, then we can take $\alpha_1 = 0$ or 1 , $\alpha_2 = 0$ or 1 . We consider only the cases $\alpha_1 = 0, \alpha_2 = 0$; $\alpha_1 = 1, \alpha_2 = 0$ and $\alpha_1 = 1, \alpha_2 = 1$ and accordingly denote the estimates by MINQUE(0,0,1), MINQUE(1,0,1) and MINQUE(1,1,1). The efficiency of these estimators is studied through a simulation study in Section 6.5.

2. It may be noted that the estimators $\hat{\sigma}_a^2$ and $\hat{\sigma}_a^2(A)$ based on the interpenetration data set only, are also (unconditionally) unbiased under the model (6.14).

Fellegi (1974) provided a different estimator of σ_a^2 from both data sets, based on the ANOVA technique. His estimator, in terms of our notation used in this section, is given by

$$\hat{\sigma}_a^2 = 2^{-1} (P-p)^{-1} \sum_{m=p+1}^P (\bar{y}_{m1.} - \bar{y}_{m2.})^2$$

$$= 2^{-1} p^{-1} \sum_{m=1}^p (\bar{y}_{m.1.} - \bar{y}_{m.2.})^2, \quad (6.17)$$

where $\bar{y}_{m1.}$ and $\bar{y}_{m.2.}$ are the average values of y over the missing subscripts. The estimator $\hat{\sigma}_a^2$ is also unbiased under the model (6.14).

Fellegi (1974) conjectured that the simple arithmetic mean of $\hat{\sigma}_a^2$ and $\hat{\sigma}_a^2$ might be more efficient than either of them taken separately.

At the time of writing this thesis, Biemer and Stokes (1985) reported some results on estimating σ_a^2 from both the data sets (Data Set A and Data Set B). They established the independence of the estimators $\hat{\sigma}_a^2$ and $\hat{\sigma}_a^2$ when the random error components are normally distributed.

Further, they obtained the optimum linear combination of $\hat{\sigma}_a^2$ and $\hat{\sigma}_a^2$:

$$\hat{\sigma}_a^2(B) = \lambda \hat{\sigma}_a^2 + (1-\lambda) \hat{\sigma}_a^2, \quad (6.18)$$

where $\lambda = V(\hat{\sigma}_a^2) / [V(\hat{\sigma}_a^2) + V(\hat{\sigma}_a^2)]$ which depends on the ratios σ_b^2/σ_c^2 and σ_a^2/σ_e^2 . However, they did not establish the optimality of $\hat{\sigma}_a^2$ defined by (6.10).

6.5. A Monte Carlo Study

A Monte Carlo experiment was undertaken to compare the efficiency of the various estimators of σ_a^2 , proposed in Sections 6.3 and 6.4,

and to evaluate the increase in efficiency brought about by using the non-interpenetrated data set, Data Set B. In this Monte Carlo experiment, the parameters P , p and n were set respectively equal to 30, 5 and 10, which leads to the design with 25 non-interpenetrated and 5 interpenetrated pairs of EAs, each EA consisting of 10 sampled units.

To obtain Monte Carlo estimates of the variances of the estimators of θ , 500 sets of observations on y were obtained as follows. First,

500 sets of independent normal pseudo random deviates: $\{e_{ijmk} : i, j = 1, 2; k = 1, \dots, 10; \text{ for } m = 1, \dots, 5 \text{ and } i = j = 1, 2; k = 1, \dots, 20 \text{ for } m = 6, \dots, 30\}$; $\{a_{mi} : i = 1, 2; m = 1, \dots, 30\}$ and $\{b_{mj} : j = 1, 2; m = 1, \dots, 30\}$ with zero means and variances σ_e^2 , σ_a^2 and σ_b^2 respectively were obtained. The value of σ_e^2 was set equal to one, while

for the pair (σ_a^2, σ_b^2) five pairs of values were considered, namely, (0.25, 1.5), (0.25, 1.0), (0.49, 1.5), (0.49, 1.0) and (0.81, 0.25).

Using these pseudo random deviates, the observations on y : $\{y_{ijmk} :$

$i, j = 1, 2; k = 1, \dots, 10; \text{ for } m = 1, \dots, 5 \text{ and } i = j = 1, 2; k = 1, \dots, 20 \text{ for } m = 6, \dots, 30\}$ were then obtained from the model defined by (6.1) and (6.13)

with $\gamma_j = 0$. For convenience, the nine estimators considered in Sections

6.2 and 6.4 viz., $\hat{\sigma}_a^2$, $\hat{\sigma}_a^2(A)$, $\hat{\sigma}_a^2$, $(\hat{\sigma}_a^2 + \hat{\sigma}_a^2)/2$, MINQUE (0,0,1), MINQUE (1,0,1), MINQUE (1,1,1), $\hat{\sigma}_a^2(B)$ and the optimal estimator $\hat{\sigma}_a^2$, were denoted

respectively by t_i ($i = 1, \dots, 9$). The relative efficiency of the estimators t_i ($i=1, \dots, 8$) with respect to the optimal estimator

$t_9 = \hat{\sigma}_a^2$ was computed using the formula

$$RE(i) = RE(t_i : \hat{\sigma}_a^2) = V_s(\hat{\sigma}_a^2) / V_s(t_i); i = 1, \dots, 8. \quad (6.19)$$

Here V_s denotes the Monte Carlo variance. It should be recalled here that the estimators $t_1 = \hat{\sigma}_a^2$ and $t_2 = \hat{\sigma}_a^2(A)$ are based only on the interpenetrated data set, Data Set A, while all other estimators utilize both portions of the complete data set i.e., Data Set A and Data Set B. Further, the estimators $t_8 = \hat{\sigma}_a^2(B)$ and $t_9 = \hat{\sigma}_a^2$ assume the knowledge of the ratios σ_b^2/σ_e^2 and σ_a^2/σ_e^2 .

From RE(1), the percentage increase in efficiency (IE) due to using the non-interpenetrated data set can be computed, and is given by

$$IE = 100(1/RE(1) - 1). \quad (6.20)$$

The values of RE(i) ($i = 1, \dots, 8$) and IE are reported in Table 6.1 (see page 121).

The following are the major conclusions that can be drawn from Table 6.1:

Estimators based on Data Set A: The values corresponding to RE(1) and RE(2) indicate that estimator $t_1 = \hat{\sigma}_a^2$ is superior to $t_2 = \hat{\sigma}_a^2(A)$, as expected, since $\hat{\sigma}_a^2$ is optimal in the class of quadratic unbiased estimators based on Data Set A. However, the difference in their relative efficiencies is small, the maximum difference being 0.02 for the case $\sigma_b^2 = 0.25$ and $\sigma_a^2 = 0.81$. It can be seen that the relative efficiency of t_1 and t_2 decrease monotonically as σ_b^2/σ_a^2 decreases. When $\sigma_b^2/\sigma_a^2 = 3$, the relative efficiency of t_1 equals 0.91 and it increases to 0.97 as σ_b^2/σ_a^2 increases to 6. However, when $\sigma_b^2/\sigma_a^2 < 1$, both the estimators t_1 and t_2 i.e., $\hat{\sigma}_a^2$ and $\hat{\sigma}_a^2(A)$, have poor relative efficiency for the case $\sigma_b^2/\sigma_a^2 = 0.32$, the efficiency of t_1 relative to the optimal MINQU estimator is only 34%. However, many

practical situations correspond to values $\sigma_b^2/\sigma_a^2 \geq 1$.

Estimators based on both the data sets: Among all the estimators of

σ_a^2 , Fellegi's second estimator based on the combined data set, viz

$t_3 = \frac{-2}{a}$ has the poorest relative efficiency, reaching a maximum of

only 0.22 for the case $\sigma_b^2 = 0.25$ and $\sigma_a^2 = 0.81$. As a result, the

estimator t_4 which is the mean of t_1 and t_3 , also performs poorly,

in particular when $\sigma_b^2/\sigma_a^2 = 2$. This does not support Fellegi's conjecture

that the arithmetic mean of the two estimators t_1 and t_3 should

outperform the estimator $t_1 = \sigma_a^2$, based on data set A alone. However,

when $\sigma_b^2/\sigma_a^2 = 0.32$, his conjecture is valid. Turning to the comparison of

the estimators which do not assume knowledge of the ratios σ_b^2/σ_e^2 and

σ_a^2/σ_e^2 , the MINQU estimators t_5, t_6 and t_7 , namely MINQUE (0,0,1)

MINQUE (1,0,1) and MINQUE (1,1,1), perform well for all values of

σ_b^2/σ_a^2 ; even for the case $\sigma_b^2/\sigma_a^2 = 0.32$, their relative efficiencies

are over 79%.

The estimator $t_8 = \hat{\sigma}_a^2(B)$ which assumes a knowledge of both the

ratios σ_b^2/σ_e^2 and σ_a^2/σ_e^2 , has relative efficiency greater than

0.92 for the cases $\sigma_b^2/\sigma_a^2 \geq 2$. However, for the case $\sigma_b^2/\sigma_a^2 = 0.32$,

its relative efficiency drops to 0.79. From the relative efficiency

values of the estimators t_6, t_7 and t_8 , namely MINQUE (1,0,1),

MINQUE (1,1,1) and the composite estimator $\hat{\sigma}_a^2(B)$, it can be seen that the

differences in their relative efficiencies are small. When $\sigma_b^2/\sigma_a^2 \geq 2$, the

estimator $t_8 = \hat{\sigma}_a^2(B)$ has greater efficiency, at the cost of assuming a

knowledge of σ_b^2/σ_a^2 and σ_a^2/σ_e^2 . The maximum increase in the relative

efficiency is only about 0.03 over both MINQUE (1,0,1) and

MINQUE (1,1,1). When $\sigma_b^2/\sigma_a^2 = 0.32$, the relative efficiency of MINQUE (1,1,1) exceeds that of $\hat{\sigma}_a^2(B)$ by about 6%.

Increase in efficiency due to using Data Set B: The percentage increase in efficiency (IE) due to the use of non-interpenetrated data set, Data Set B, increases as σ_b^2/σ_a^2 decreases. When the ratio $\sigma_b^2/\sigma_a^2 = 2$, the percentage gain in efficiency (IE) ranges from 3 to 19. On the other hand, when $\sigma_b^2/\sigma_a^2 < 2$, large gains in efficiencies are possible, e.g., when $\sigma_b^2 = 0.25$ and $\sigma_a^2 = 0.81$, then the increase in efficiency is 194%.

In summary, the non-interpenetrated data makes a worthwhile contribution to the efficiency of the estimator of σ_a^2 when interviewer variance σ_a^2 is large relative to the between EA variance σ_b^2 . Such a relationship implies homogeneous enumeration areas since interviewers' variance is usually small. Biemer and Stokes (1985), for the data from U.S. 1980 Census of population and housing, do report examples of homogeneity for some characteristics e.g., sex and race. These authors also concluded that use of non-interpenetrated data was important only for the case of homogeneity. However, they based their conclusions on a definition of IE similar to us except that they used $V(\hat{\sigma}_a^2(B))$ instead of $\text{Var}(\hat{\sigma}_a^2)$. With respect to the choice of an estimator of σ_a^2 , MINQUE(1,0,1) and MINQUE(1,1,1) are the best candidates. When the computational complexity is taken into account, MINQUE(1,0,1) is the recommended choice. Of course, if the ratios σ_b^2/σ_e^2 and σ_a^2/σ_e^2 are known, $\hat{\sigma}_a^2$ should be used.

Remark 6.3. A crude estimate of Monte Carlo error was obtained by computing $\text{Var}(\hat{\sigma}_a^2)$ and $\text{Var}(\tilde{\sigma}_a^2)$ from the analytic expressions given by (6.11) and (6.16), and then comparing these with the Monte Carlo variance estimators of $\text{Var}(\hat{\sigma}_a^2)$ and $\text{V}(\tilde{\sigma}_a^2)$. The discrepancy was found to be about 6%.

Table 6.1: Relative Efficiency of CRV estimators with respect to the optimal estimators σ_a^2 and the Percentage Increase in Efficiency (IE) by using Data Set B

($p = 5$, $p = 30$, $n = 10$, $\sigma_e^2 = 1.0$).

σ_b^2	σ_a^2	σ_b^2/σ_a^2	Relative efficiency of CRV estimators								IE.
			t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	
1.50	0.25	6	0.97	0.96	0.03	0.12	0.97	0.99	0.97	1.00	3
1.00	0.25	4	0.92	0.92	0.03	0.12	0.93	0.94	0.94	0.95	9
1.50	0.49	3	0.91	0.90	0.08	0.29	0.93	0.96	0.98	0.99	10
1.00	0.49	2	0.84	0.82	0.08	0.32	0.86	0.90	0.91	0.92	19
0.25	0.81	0.32	0.34	0.32	0.22	0.55	0.79	0.80	0.85	0.79	194

CHAPTER 7

SUMMARY AND CONCLUSIONS

This thesis deals with two problems viz, (i) the prediction of small area means and (ii) the measurement of response error variance in surveys. The main results of this thesis may be summarized as follows:

1. Three small area models, proposed by Battese and Fuller (1982), Dempster *et al.* (1981) and Fay and Herriot (1979), are expressed as a general linear model with components of variance error structure. For the known variance components case, best linear unbiased predictors (BLUPs) of small area means and their variances are derived in Chapter 2.
2. The results of Chapter 3 can be used to obtain confidence intervals for a non-linear function of the variance components of a regression model with components of variance error structure. Confidence intervals proposed in this chapter are based on a new weighted jackknife procedure. Through a Monte Carlo study, it has been demonstrated that these intervals are more robust to deviations from normality assumption, as compared to normality-based intervals.
3. When the estimates of the variance components are used to compute the BLUP, the explicit expression for the mean square error (MSE) of the estimated BLUP is not easily tractable. An approximate expression, MSE^* , for the true MSE is obtained under normality assumption in Chapters 4 and 5. A normality-based estimator and a jackknife estimator

of MSE^* are also derived. Through a Monte Carlo study, it has been observed that (i) the estimated BLUP outperforms other estimators in terms of mean square error; (ii) MSE^* values are closer to the true MSE values for data generated from a normal or a uniform distribution. In the case of data generated from exponential and double exponential distributions, MSE^* yields values exceeding the true MSE values significantly; (iii) for estimating the true MSE, both the normality-based estimator and the jackknife estimator of MSE^* perform well only under the assumption of a normal or a uniform distribution. In other cases viz., exponential or double exponential distributions, the biases of both the estimators are unacceptably high in estimating the true MSE. However, the jackknife estimator estimates the approximation to MSE with a low bias.

From the Monte Carlo experiment, it has been found that the Monte Carlo estimate of the covariance term viz., $E[t_1(\sigma^2, y) - t_1(\sigma^2, y)] \cdot [t_1(\sigma^2, y) - \mu_1]$ defined in Chapter 4, which is zero under normality, is negative for exponential and double exponential data. Thus it would be profitable to obtain a robust estimator of the true MSE directly. One possible approach is to obtain jackknife estimators of the covariance term and the first component on the right side of (4.5), viz., $E[t_1(\sigma^2, y) - t_1(\sigma^2, y)]^2$. Here, it should be recalled that the second term of (4.5) viz., the variance of the BLUP, can be evaluated without the normality assumption (see Theorem 2.2).

(iv) The customary estimator of MSE, which is obtained from the variance of the BLUP by replacing the unknown variance components by

their asymptotically consistent estimators, seriously underestimates the true MSE.

4. The estimation of the correlated response variance (CRV) from survey data is considered in Chapter 6. A variance component model is proposed for obtaining the optimal estimator of CRV, utilizing both interpenetrated and non-interpenetrated data. The optimal estimator, which is derived using the method of minimum norm quadratic unbiased estimation (MINQUE), requires a knowledge of variance component ratios. Through a Monte Carlo study it has been demonstrated that in the absence of a knowledge of variance ratios, the estimators MINQUE (0,0,1), MINQUE (1,0,1) and MINQUE (1,1,1) defined in Chapter 6, are good choices for estimating the CRV. Further, it has been shown that the non-interpenetrated data makes a worthwhile contribution to the efficiency of the estimator of CRV only when the between interviewer variability is large relative to the variability between the means of enumeration areas. Biemer and Stokes (1985), suggested a different estimator (also based on normality assumption and on a knowledge of the variance component ratios) of CRV. The gain in efficiency of optimal estimator of CRV over the Biemer-Stokes estimator is found to be substantial only under the cases when the non-interpenetrated data contribute significantly to the increase in efficiency of the estimator of CRV. The above Monte Carlo results are obtained under normality assumption. It would be useful to assess the performance of the Biemer-Stokes estimator over the optimal estimator under non-normal data sets.

APPENDIX

A.1 Proof of Lemma 3.2.

The proof of Lemma 3.2 needs the following lemma.

Lemma A: (Bishop, Fienberg and Holland (1977), Theorem 14.4.1, p.476).

Let $\{\xi_n\}$ be a stochastic sequence such that $\mu_n = E\xi_n$ and $\sigma_n^2 = \text{var}(\xi_n) < \infty$. Then $\xi_n - \mu_n = O_p(\sigma_n)$.

(a) To establish the result (a) of Lemma 3.2, viz.,

$$\sigma_e^2 = \sigma_e^2 + O_p(t^{-1}),$$

we consider

$$\hat{\sigma}_e^2 = (n-t-1)^{-1} \sum_{i=1}^t \sum_{j=1}^{n_i} e_{ij}^2,$$

where e_{ij} is defined by (3.22). It follows that

$$\hat{\sigma}_e^2 = \sigma_e^2 - (\beta^* - \beta)^2 (n-t-1)^{-1} \left[\sum_{i=1}^t \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \right], \quad (\text{A.1})$$

where β^* is defined by (3.44) and

$$\beta^* = \left[\sum_{i=1}^t \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \right]^{-1} \left[\sum_{i=1}^t \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i) \right]. \quad (\text{A.2})$$

In view of the regularity condition C4, the desired result, viz.,

$$\hat{\sigma}_e^2 = \sigma_e^2 + O_p(t^{-1}), \text{ follows from (A.1) provided } (\beta^* - \beta) = O_p(t^{-1}).$$

The latter result is obtained from Lemma A by noting that

$$\text{Var}(\beta^*) = \sigma_e^2 \left[\sum_{i=1}^t \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \right]^{-1} = O(t^{-1}). \quad \square$$

(b) The result (b), viz.

$$\text{Var}(\bar{e}^2) = \text{Var}(\bar{e}^2) + O(t^{-2}),$$

follows from (A.1) by establishing that $\text{Var}\{(b^* - \bar{e})^2\}$ and $\text{Cov}\{\bar{e}^2, (b^* - \bar{e})^2\}$ both are of order $O(t^{-2})$. The proof for the result

$$\text{Cov}(\bar{e}^2, (b^* - \bar{e})^2) = O(t^{-2}) \quad (\text{A.3})$$

is provided below. Along the same lines the result

$$\text{Var}\{(b^* - \bar{e})^2\} = O(t^{-2})$$

follows.

To establish (A.3), we first define

$$d_1 = \sum_j (x_{1j} - \bar{x}_{1.}) \bar{e}_{1j} = \sum_j (x_{1j} - \bar{x}_{1.}) e_{1j} \quad (\text{A.4})$$

and

$$z_1 = \sum_j \bar{e}_{1j}^2 - E \sum_j \bar{e}_{1j}^2$$

Note that the random vectors $(d_1, z_1)'$; $i=1, \dots, t$ are independent with $E(d_1, z_1)' = (0; 0)'$; $i=1, \dots, t$. From the definition of $\bar{e}^2$ and b^* , it follows that

$$\text{Cov}(\bar{e}^2, (b^* - \bar{e})^2) = (n-t-1)^{-1} \left(\sum_{i,j} (x_{ij} - \bar{x}_{i.})^2 \right)^{-2} \text{Cov} \left(\sum_i z_i, \left(\sum_i d_i \right)^2 \right).$$

Consider

$$\begin{aligned} |\text{Cov}(\sum_i z_i, (\sum_i d_i)^2)| &= |E(\sum_i z_i) (\sum_i d_i)^2| = |E \sum_i z_i d_i^2| \\ &\leq \sum_i E |z_i d_i^2| \\ &\leq \left[\sum_i E z_i^2 \sum_i E d_i^4 \right]^{1/2} \end{aligned}$$

by Cauchy-Schwarz inequality. From the regularity conditions C1 and (C3-C5), it follows that

$$\sup_{1 \leq i \leq t} E z_1^2 < \infty; \text{ and } \sup_{1 \leq i \leq t} E d_1^4 < \infty.$$

Hence $\text{Cov}(\sum_1^t z, (\sum_1^t d_1)^2) = O(t)$, from which the result (A.3) follows by

noting that $(n-t-1)^{-1} \{ \sum_{i,j} (x_{ij} - \bar{x}_{i.})^2 \}^{-2} = O(t^{-3})$. $\square$

(c) To prove (c) i.e., $v_J(\hat{\sigma}_e^2) = v_J(\bar{\sigma}_e^2) + O_p(t^{-3/2})$, we write $v_J(\hat{\sigma}_e^2)$ and $v_J(\bar{\sigma}_e^2)$ as follows: From (3.31), (3.35) and (3.38) it follows that

$$v_J(\hat{\sigma}_e^2) = t(t-1)^{-1} (n-t-1)^{-2} \left\{ \sum_i \left(\sum_j \hat{e}_{ij}^2 - w_i \hat{\sigma}_e^2 \right)^2 \right\} \quad (\text{A.5})$$

and similarly from (3.43) - (3.45)

$$v_J(\bar{\sigma}_e^2) = t(t-1)^{-1} (n-t-1)^{-2} \left\{ \sum_i \left(\sum_j \bar{e}_{ij}^2 - w_i \bar{\sigma}_e^2 \right)^2 \right\}. \quad (\text{A.6})$$

It can be shown that $\sum_j \hat{e}_{ij}^2$ and $\sum_j \bar{e}_{ij}^2$ are related by

$$\sum_j \hat{e}_{ij}^2 = \sum_j \bar{e}_{ij}^2 - 2(\beta^* - \beta) d_{i1} + (\beta^* - \beta)^2 \sum_j (x_{ij} - \bar{x}_{i.})^2, \quad (\text{A.7})$$

where β^* and d_{i1} are defined by (A.2) and (A.4) respectively. Since

$\beta^* - \beta = O_p(t^{-1/2})$ and $\hat{\sigma}_e^2 - \bar{\sigma}_e^2 = O_p(t^{-1})$, the substitution of (A.7) in

(A.5) yields (after simplification)

$$\begin{aligned} v_J(\hat{\sigma}_e^2) &= v_J(\bar{\sigma}_e^2) - 2(\beta^* - \beta) t(t-1)^{-1} (n-t-1)^2 \left\{ \sum_i d_{i1} w_i \sum_j \hat{e}_{ij}^2 - \sum_i d_{i1} w_i \bar{\sigma}_e^2 \right\} \\ &\quad + (\beta^* - \beta)^2 t(t-1)^{-1} (n-t-1)^{-2} \sum_i d_{i1}^2 w_i^2 + O_p(t^{-2}). \end{aligned} \quad (\text{A.8})$$

Turning to the stochastic order of the terms $\sum_1 d_{11} w_1 \sum_j e_{ij}^2$, $\sum_1 d_{11} w_1 e_{11}^{2-2}$

and $\sum_1 d_{11} w_1^2$ we claim the following:

$$(i) \quad \sum_1 d_{11} w_1 e_{11}^2 = O_p(t^{-1})$$

$$(ii) \quad t^{-1} \sum_1 d_{11} w_1^2 = O_p(t^{-1})$$

$$(iii) \quad \sum_1 d_{11} w_1 (d_{11} - E d_{11}) = o_p(1)$$

and

$$(iv) \quad t^{-1} \sum_1 d_{11} w_1 \sum_j e_{ij}^2 - E d_{11} \sum_j e_{ij}^2 = o_p(1)$$

The results (i) and (ii) follow easily from Lemma A by noting that e_{11}^2 and w_1^2 both are of order $O(t^{-1})$ while the results (iii) and (iv) follow from Lemma 3.2 provided

$$\sup_{1 \leq k \leq t} E(d_{11}^2 | \mathcal{F}_{k-1}) < \infty \quad (A.9)$$

$$\left(\sup_{1 \leq k \leq t} E \left| d_{11} \sum_j e_{ij}^2 - E d_{11} \sum_j e_{ij}^2 \right|^{1+\epsilon} \right)^{1/(1+\epsilon)} < \infty \quad (A.10)$$

For some $\epsilon > 0$.

We establish (A.10) below, while (A.9) follows along the same lines in a simpler way.

Since $\sup_{1 \leq k \leq t} E(d_{11}^2 | \mathcal{F}_{k-1}) < \infty$ (by Cauchy-Schwarz inequality), to show (A.10) it is sufficient to prove that $\sup_{1 \leq k \leq t} E \left| d_{11} \sum_j e_{ij}^2 - E d_{11} \sum_j e_{ij}^2 \right|^{1+\epsilon} < \infty$. Hence we consider

$$E \left| d_{11} \sum_j e_{ij}^2 \right|^{1+\epsilon} \leq E |d_{11}|^{2(1+\epsilon)} E \left| \sum_j e_{ij}^2 \right|^{2(1+\epsilon)/2}$$

(by Cauchy-Schwarz inequality). Taking $\delta' = 1/2$, where δ' is as specified in C1, it follows that $\sup_{1 \leq i \leq t} E |d_i|^2 |2(1+\delta')| < \infty$ and

$$\sup_{1 \leq i \leq t} E \left| \sum_j \hat{e}_{ij}^2 \right|^{2(1+\delta')} < \infty; \text{ and hence } \sup_{1 \leq i \leq t} E |d_i| \left| \sum_j \hat{e}_{ij}^2 \right|^{1+\delta'} < \infty.$$

Now, the desired result (c) of Lemma 4.2, viz. $v_J(\hat{c}_e^2) = v_J(\bar{c}_e^2) + O_p(t^{-3/2})$,

follows from (A.8) using the above results (i) - (iv) and by noting

$$\text{that } \hat{c}^* - \bar{c} = O_p(t^{-1}). \quad \square$$

(d) Turning to proof of (d) i.e.,

$$Ev_J(\bar{c}_e^2) = V(\bar{c}_e^2) + O(t^{-2}),$$

in view of the result (b) of Lemma 3.2 it is sufficient to establish that

$$Ev_J(\bar{c}_e^2) = V(\bar{c}_e^2) + O(t^{-2}).$$

From (A.6), it follows that

$$Ev_J(\bar{c}_e^2) = t(t-1)^{-1}(n-t-1)^{-2} \left\{ \sum_i E \left(\sum_j \hat{e}_{ij}^2 \right)^2 \right.$$

$$+ 2(n-t-1)^{-1} E \left(\sum_i w_i \sum_j \hat{e}_{ij}^2 \right) \left(\sum_i \sum_j \hat{e}_{ij}^2 \right) +$$

$$\left(\sum_i w_i^2 \right) E \bar{c}_e^4 \Big\}.$$

$$= t(t-1)^{-1}(n-t-1)^{-2} \left\{ (n-t-1)^2 V(\bar{c}_e^2) + \left(\sum_i w_i^2 \right) \bar{c}_e^4 \right.$$

$$+ 2(n-t-1)^{-1} \left(\sum_i w_i V \left(\sum_j \hat{e}_{ij}^2 \right) + \sum_i w_i^3 \bar{c}_e^4 + \sum_{i \neq i'} \sum_j w_i^2 w_{i'}^2 \bar{c}_e^4 \right)$$

$$+ \left(\sum_i w_i^2 \right) V(\bar{c}_e^2) + \sum_i w_i^2 \bar{c}_e^4 + O(1) \Big\}.$$

$$\begin{aligned}
V(\bar{e}_{ij}^2) &= O(t^{-1}) - 2t(t-1)(n-t-1)^{-3} \sum_1 w_1 V(\sum_j e_{1j}^2) \\
&+ 4t(t-1)^{-1}(n-t-1)^{-2} \{ 2 \sum_1 w_1^2 - 2(n-t-1)^{-1} (\sum_1 w_1^2) \sum_1 w_1 \} \\
&+ O(t^{-2}) \\
V(\bar{e}_{ij}^2) &= O(t^{-1}) ,
\end{aligned}$$

which follow by noting that $(n-t-1)^{-1} \sum_1 w_1 V(\sum_j e_{1j}^2)$ and

$(n-t-1)^{-1} \sum_1 V(\sum_j e_{1j}^2)$ are order $O(1)$ (from condition (C1)). $\square$

A.1. Values of n_1 , x_{1j} , and $\bar{x}_1$ ($i=1, \dots, t$; $j=1, \dots, n_1$) used in the
Monte Carlo Studies.

The values of n_1 , x_{1j} and $\bar{x}_1$ ($i=1, \dots, t$; $j=1, \dots, n_1$) used
in the Monte Carlo experiments (see Section 3.5, and 4.3) are provided
in Table A below:

Table A: Values of n_1 , x_{1j} and $\bar{x}_1$ used in Monte Carlo studies

i	n_1	x_{1j}	$\bar{x}_1$
1	2	374, 209	300
2	3	253, 432, 367	290
3	3	361, 288, 369	318.21
4	3	206, 316, 145	257.17
5	3	355, 295, 223	291.77
6	3	459, 240, 307	301.26
7	4	252, 393, 206, 302	262.17
8	5	313, 246, 353, 271, 237	314.28
9	5	221, 369, 343, 342, 294	298.65
10	6	220, 340, 355, 261, 187, 350	325.99

REFERENCES

- ARVESEN, J. (1969), Jackknifing U-statistics. *Ann. Math. Statist.* 40, 2076-2100.
- ARVESEN, J., and SCHMITZ, T. (1970), Robust procedures for variance component problems using the jackknife. *Biometrics*, 26, 677-686.
- BATTESE, G.E., and FULLER, W.A. (1982). An error components model for prediction of county crop areas using survey and satellite data. Working papers in econometrics and applied statistics, Department of Economic Statistics, University of New England, Armidale, N.S.W., Australia.
- BATTESE, G.E., FULLER, W.A., and HICKMAN, R.D. (1976), Estimation of response variance from interview-reinterview surveys. *J. Indian Soc. Agricultural Statist.*, 28, 1-14.
- BIEMER, P.P. (1978), The estimation of non-sampling variance components in sample surveys. Unpublished Ph.D. Thesis, Institute of Statistics, Texas A & M.
- BIEMER, P.P. and STOKES, S.L. (1985). Optimal design of interviewer variance experiments in complex surveys. *J. Amer. Statist. Assoc.* 80, 158-166.
- BISHOP, Y.M.M., FIENBERG, S.E. and HOLLAND, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Mass.
- BOGUE, D.J. (1950). A technique for making extensive population estimates. *J. Amer. Statist. Assoc.* 45, 149-163.

CARTER, G. and ROLPH, J. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *J. Amer. Statist. Assoc.* 69, 880-885.

COCHRAN, W.G. (1968). Errors of measurements in statistics. *Technometrics*, 10, 637-666.

COCHRAN, W.G. (1977). *Sampling Techniques*. 3rd ed. John Wiley & Sons, New York.

DALENIUS, T. (1977a). Strain at a gnat and swallow a camel: on the Problem of measuring sampling and non-sampling errors. *Proc. Amer. Statist. Assoc., Social Statistics*, 21-25.

DALENIUS, T. (1977b). Bibliography on non-sampling errors in surveys: I (A to G); II (H to Q); III (R to Z). *Int. Statist. Rev.* 45, 71-89; 181-197; 303-317.

DEMING, W.E., and STEPHAN, R.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* 11, 427-444.

DEMPSTER, A.P., and RAGUHNATHAN, T.E. (1985). Using a covariate for Small Area Estimation: A common sense Bayesian approach. Paper presented at the small area symposium, May 22-24, Ottawa, Canada.

DEMPSTER, A.P., RUBIN, D.B. and TSUTAKAWA, R.K. (1981). "Estimation in covariance components models". *J. Amer. Statist. Assoc.* 76, 341-353.

DODDS, D.J. and SMITH, T.M. (1973). "Estimation of correlated response variance under a linear additive model". Paper presented at the International Association of Survey Statisticians Meetings, Vienna.

DREW, J.D., SINGH, M.P. and CHOUDHRY, G.H. (1982). Evaluation of small area estimation techniques for the Canadian labour force survey. *Survey Methodology*, 8, 17-47.

DUMAIS, J., PATON, D., SRINATH, K.P., and VERMA, R. (1985). Overview of small area estimation techniques. Unpublished Report, Statistics Canada, Canada.

EFRON, B., and MORRIS, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach, *J. Amer. Statist. Assoc.* 68, 117-130.

FAY, R.E. and HERRIOT, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.* 74, 269-277.

FELLEGI, I.P. (1974). An improved method of estimating the correlated response variance. *J. Amer. Statist. Assoc.*, 69, 496-500.

FULLER, W.A. 1976. *Introduction to Statistical Time Series*. John Wiley & Sons, New York.

FULLER, W.A. and Battese, G.E. (1973). Transformations for estimation of linear models with nested-error structure. *J. Amer. Statist. Assoc.* 68, 626-632.

GONZALEZ, M.E. (1973). Use and evaluation of synthetic estimates, *Proc. Amer. Statist. Assoc., Social Statistics Section*, 33-36 Washington, D.C.

GRAYBILL, F.A. (1969). *Introduction to Matrices with Applications in Statistics*. Wadsworth, California.

HANSEN, M.H., HURWITZ, W.N. and BERSHAD, M.A. (1959). Measurement errors in censuses and surveys, *Proc. Amer. Statist. Assoc., Social Statistics Section*, 2-5.

HANSEN, M.H., HURWITZ, W.N. and BERSHAD, M.A. (1961). Measurement errors in censuses and surveys. *Bulletin of the International Statistical Institute*, 38, 359-374.

HARTER, R.M. (1983). Small area estimation using nested-error models and auxiliary data. Unpublished Ph.D. thesis, Iowa State University, Ames, Iowa.

HARTLEY, H.O., and RAO, J.N.K. (1978). The estimation of non-sampling variance components in sample surveys. *Survey Sampling and Measurement*, ed. N.K. Namboodiri, New York: Academic Press, 35-43.

HARTLEY, H.O., RAO, J.N.K., and LAMOTTE, L. (1976). A simple synthesis-based method of variance component estimation. *Biometrics*, 34, 233-242.

HARVILLE, D.A. (1977). Maximum likelihood approach to variance component estimation and related problems. *J. Amer. Statist. Assoc.* 72, 320-340.

HARVILLE, D.A. (1976). Extension of the Gauss-Markov Theorem to include the estimation of random effects. *Ann. Statist.* 4, 2. 384-395.

HENDERSON, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423-447.

HIDIROGLOU, M.A., MORRIS, M., DAGUM, E.G., RAO, J.N.K., SARNDAL, C.E. (1984). Evaluation of small area estimators using administrative records. *Proc. Amer. Statist. Assoc., Social Statistics Section*.

HINKLEY, D.V. (1977). Jackknifing in unbalanced situations. *Technometrics* 19, 285-292.

HOLT, D., SMITH, T.M.F. and TOMBERLIN, T.J. (1979). A model-based approach to estimation for small subgroups of a population. *J. Amer. Statist. Assoc.* 74, 405-410.

JAMES, W., and STEIN, C. (1961). Estimation with quadratic loss.

Proc. Fourth Berkley Symposium. University of California Press, Berkeley, 361-379.

KACKAR, R.N. (1979). Variance approximations for estimators of fixed and random effects in mixed linear models. Unpublished Ph.D. thesis, Iowa State University, Dept. of Statistics, Ames, Iowa.

KACKAR, R.N. and HARVILLE, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.* 79, 853-862.

KLEFFE, J. and SEIFERT, B. (1983). Computation of variance components by the MINQUE-method. Technical Report, Academy of Sciences of G.D.R., Institute of Mathematics.

LAAKE, P. (1979). A predictive approach to subdomain estimation in finite populations. *J. Amer. Statist. Assoc.* 74, 355-358.

LEVY, P.S. (1971). The use of mortality data in evaluation of synthetic estimates. *Proc. Amer. Statist. Assoc., Social Statistics Section.* 328-331.

LOEVE, M. (1963). *Probability Theory.* D. van Nostrand Company, Princeton.

MAHALANOBIS, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *J. Roy. Statist. Soc.*, 109, 325-370.

MILLER, R.G. (1974). An unbalanced jackknife. *Ann. Statist.* 2, 880-891.

PEIXOTO, J.L. (1982). Estimation of random effects in the balanced one-way classification. Ph.D. thesis, Iowa State University, Ames, Iowa.

PURCELL, N.J., and KISH, L. (1979). Estimation for small domains. *Biometrics* 35, 365-384.

PURCELL, N.J., and KISH, L. (1980). Postcensal estimates for local areas (or domains). *Int. Statist. Rev.* 48, 3-18.

QUENOUILLE, M.H. (1949). Notes on bias in estimation. *Biometrika*, 43, 353-360.

QUENOUILLE, M.H. (1956). Approximate tests of correlation in time-series, *J. Roy. Statist. Soc., Ser. B*, 11, 68-74.

RAO, C.R. and KLEFFE, J. (1980). Estimation of variance components, P.R. Krishnaiah, ed. *Handbook of Statistics*, Vol. 1, 1-40, North Holland Publishing Company.

RAO, C.R. and MITRA, S.K. (1971). *Generalized Inverse of Matrices and Its Applications*. John Wiley.

RAO, J.N.K. (1984). Some thoughts on small area estimation. Unpublished report, Statistics Canada, Canada.

RIVES, N.W. (1976). A modified housing-unit method for small-area population estimation. *Proc. Amer. Statistic Assoc., Social Statistics Section*, 717-720.

ROYALL, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57, 377-387.

SARNDAL, C.E. (1984). Design consistent versus model dependent estimation for small domains. *J. Amer. Statist. Assoc.*, 79, 624-631.

SCHMITT, R.C. and CROSETTI, A.H. (1954). Accuracy of the ratio-correlation method for estimating post censal population, *Lond. Econ.* 30, 279-281.

SEARLE, S.R. (1971). *Linear Models*, New York: John Wiley.

SRIVASTAVA, V.K. and TIWARI, R. (1976). Evaluation of expectation of products of stochastic matrices. *Scandinavian J. Statist.* 3, 135-138.

SRINATH, K.P. and HIDIROGLOU, M.A. (1985). Sample size-dependent estimators for small areas with applications to business data, paper presented at the Small Area Symposium, May 22-24, Ottawa, Canada.

STROUD, T.M.F. (1984). Bayesian shrinkage estimates for regression coefficients in m populations. *Commun. Statist.-Theory Meth.*, 13(17), 2085-2109.

SUKHATME, P.V. and SETH, G.R. (1952). Non-sampling errors in surveys, *J. Indian Soc. Agricultural Statist.* 4, 5-41.

TUKEY, J.W. (1958). Bias and confidence in not-quite large samples. *Ann. Math. Stat.*, 29, 614.

U.S. Bureau of the Census (1949). Illustrative examples of two methods of estimating the current population of small areas. *Curr. Popul. Rep.*, Ser P-25, No. 20. U.S. Government Printing Office, Washington, D.C.

U.S. Bureau of the Census (1966). Methods of population estimation: part 1. *Curr. Popul. Rep.*, Ser P-25, No. 339. U.S. Government Printing Office, Washington, D.C.

WALKER, G. and SIGMAN, R. (1984). The use of Landsat for county estimates of crop areas: evaluation of the Huddleston-Ray and the Battese and Fuller estimators for the case of stratified sampling. *Commun. Statist.-Theor. Meth.*, 13(23), 2975-2996.

WU, C.F.J. (1984). Jackknife and bootstrap inference in regression and a class of representations for the LSE. MRC Technical Summary report #2675.

END

FIN

7 04 87