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

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform 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.

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

AVIS

La qualité de cette microforme 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.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

**SMALL AREA ESTIMATION UNDER ONE AND TWO-FOLD
NESTED ERROR REGRESSION MODELS**

by

Diana M. Stukel

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

Doctor of Philosophy

Department of Mathematics and Statistics
Carleton University
Ottawa, Canada
April, 1991

© Copyright, 1991
Diana M. Stukel

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

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

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

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

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

ISBN 0-315-68815-7

Canada

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

**SMALL AREA ESTIMATION UNDER ONE AND TWO-FOLD
NESTED ERROR REGRESSION MODELS**

submitted by
Diana M. Stukel
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Chair, Department of Mathematics and Statistics

Thesis Supervisor

External Examiner

Carleton University
April 1991
© Copyright, 1991
Diana M. Stukel

This thesis is dedicated to my Father and Mother

Abstract

The focus of this thesis is the estimation of small area means under two small area models: the one-fold and two-fold nested error regression models with unequal error variances. For both models, generalized least square estimates of the regression parameters are obtained by transforming the models such that the resulting errors are uncorrelated with constant variances. Best linear unbiased predictors (BLUPs) of the small area means are derived, assuming that the variance components are known. Then, unbiased estimates of the variance components are obtained using Henderson's method of fitting constants. This method is also shown to be equivalent to the computationally convenient transformation method of estimating variance components, as suggested by Fuller and Battese (1973). Empirical best linear unbiased predictors (EBLUPs) of the small area means are constructed, by substituting the estimated variances into the original expressions for the BLUPs. Finally, a second order approximation to the MSE of the EBLUP, and an estimator of the MSE are derived.

For a special case of the two-fold nested error regression model, jackknife estimators of the variances of, and the covariances between, unbiased estimators of the variance components are found, and shown to be asymptotically consistent. Jackknife confidence intervals on the ratios of variance components are formed, and shown to be robust, via a Monte Carlo study.

Finally, the two-fold nested error regression model is viewed from a Hierarchical Bayes (HB) standpoint. A unified general mixed linear model approach to HB estimation is applied to the model, and explicit expressions for the posterior expectation and variance of the small area means, conditional on the sampled data, are found.

Acknowledgements

I wish to thank a number of people whose input into this thesis made its completion possible. First and foremost, I would like to express my overwhelming gratitude to my thesis supervisor, Dr. J.N.K. Rao, whose remarkable knowledge of, and intuitive feel for, all areas of statistics, was a source of inspiration to me. He tirelessly spent countless hours guiding me through to the successful completion of this work, supplying invaluable ideas and constructive comments along the way. I also wish to extend my appreciation to Dr. Avi Singh, who provided me many opportunities to help expedite the completion of this thesis. It is my distinct pleasure to acknowledge and thank Diane Berezowski, who devoted long hours and special care in typesetting this thesis, and produced a high quality final document, from a difficult and often indecipherable original manuscript. Thanks are also due to my father, who enthusiastically took on the task of proof-reading a good deal of this thesis. Special thanks go to Dr. C.W.L. Garner, the Chairman of the Department of Mathematics, who was instrumental in assuring me scholarship funding and research grants, throughout the course of my graduate work at Carleton University. I will always fondly remember my friends at the department, who made my stay there a very pleasant one. I am indebted to my friends Riccardo Brun del Re and Lise Saulnier, and my sister, Thérèse, for their continued encouragement at times when my morale was at its lowest. Last, but not least, I would like to express my deep appreciation and love to my parents, to whom this thesis is dedicated. Their own thirst for knowledge instilled in me a respect for education and higher learning, at a young age.

Table of Contents

Abstract	iv
Acknowledgements.....	v
Table of Contents	vi
List of Tables	ix
[CHAPTER 1] - Introduction	1
1.1 An Overview of Small Area Methodology	1
1.1.1 Estimation of Small Area Means	2
1.1.2 A Second Order Approximation to the MSE of the EBLUP	11
1.1.3 Jackknife Estimation of Variances of Variance Components and Robust Confidence Intervals on Functions of Variance Components	14
1.1.4 A Hierarchical Bayes Approach	18
1.2 Thesis Outline	22
[CHAPTER 2] – The One-Fold Nested Error Regression Model with Unequal Error Variances	24
2.1 Description of the Model	24
2.2 Transformed Model with Uncorrelated Errors and Constant Variances	26
2.3 Best Linear Unbiased Prediction Under the Assumption of Known Variance Components.....	28
2.3.1 The Infinite Population Case	29
2.3.2 The Finite Population Case.....	31
2.4 Estimation of Variance Components.....	32

2.4.1 Estimation of σ_{μ}^2 and σ_{ϵ}^2	33
2.5 A Second Order Approximation to the MSE of the EBLUP	36
2.6 Estimation of the MSE of the EBLUP	43
[CHAPTER 3] – The Two-Fold Nested Error Regression Model with Unequal	
Error Variances	46
3.1 Description of the Model	47
3.2 Transformed Model with Uncorrelated Errors and Constant Variances	50
3.3 Best Linear Unbiased Prediction Under the Assumption of Known Variance	
Components	52
3.3.1 The Infinite Population Case	53
3.3.2 The Finite Population Case	55
3.4 Estimation of Variance Components	59
3.4.1 Estimation of σ_{ξ}^2 , σ_{μ}^2 and σ_{ϵ}^2	59
3.5 A Second Order Approximation to the MSE of the EBLUP	65
3.5.1 The Variances and Covariances of the Henderson Estimators Under	
the Assumption of Normality	75
3.6 Estimation of the MSE of the EBLUP	81
3.7 Monte Carlo Study	85
3.7.1 Objectives of the Study	85
3.7.2 Description of the Experiment	86
3.7.3 Results and Discussion	87

[CHAPTER 4] – Jackknife Variance Estimation and Robust Confidence Intervals on Functions of Variance Components	95
4.1 The Jackknife Method	95
4.2 Jackknife Estimation of Variances and Covariances of Variance Component Estimators.....	97
4.3 Robust Confidence Intervals on Nonlinear Functions of Variance Components	117
4.4 Monte Carlo Study.....	119
4.4.1 Objectives of the Study	120
4.4.2 Description of the Experiment	124
4.4.3 Results and Discussion	124
[CHAPTER 5] – A Hierarchical Bayes Approach to the Two-Fold Nested Error Regression Model	132
5.1 Description of the HB Model and Its Application to the Two-Fold Nested Error Regression Model.....	132
5.2 An HB Analysis Assuming λ is Known	135
5.3 An HB Analysis Assuming λ is Unknown	146
[CHAPTER 6] – Summary and Proposed Future Research	150
6.1 Summary	150
6.2 Proposed Future Research	152
Appendix A	154
Appendix B	170
References	179

List of Tables

Table 3.1 Percentage Relative Error of $MSE[t_1(\hat{\sigma}^2, y)]$, the Second Order Approximation to the MSE of the EBLUP, With Respect to the True MSE

Table 3.2 Monte Carlo Probabilities of Negative Variance Component Estimates for $\hat{\sigma}_{\xi}^2$ and $\hat{\sigma}_{\eta}^2$ where Each Pair Represents $(P_M(\hat{\sigma}_{\xi}^2 < 0), P_M(\hat{\sigma}_{\eta}^2 < 0))$

Table 3.3 Percentage Relative Bias of $mse[t_1(\hat{\sigma}^2, y)]$, the Estimator of the MSE of the EBLUP, With Respect to the True MSE

Table 3.4 Percentage Relative Bias of $mse_1[t_1(\hat{\sigma}^2, y)]$, the Naive Estimator, With Respect to the True MSE

Table 3.5 Percentage Relative Bias of $mse_2[t_1(\hat{\sigma}^2, y)]$, the Adaptive Estimator, With Respect to the True MSE

Table 4.1 Coverage Probabilities of Jackknife and Normality Based Confidence Intervals on ρ_1 for Nominal Values of 90% and 95%, and Percentage Relative Biases of $v_J(\tilde{\theta}_1)$ with Respect to the True Variance of $\tilde{\theta}_1$

Table 4.2 Coverage Probabilities of Jackknife Confidence Intervals on ρ_2 for Nominal Values of 90% and 95% (Denoted $CP(90\%)$ and $CP(95\%)$ Respectively), and Percentage Relative Biases of $v_J(\tilde{\theta}_2)$ with Respect to the True Variance of $\tilde{\theta}_2$, Under the Normal Distribution

Table 4.3 Coverage Probabilities of Jackknife Confidence Intervals on ρ_2 for Nominal Values of 90% and 95% (Denoted $CP(90\%)$ and $CP(95\%)$ Respectively), and Percentage Relative Biases of $v_J(\tilde{\theta}_2)$ with Respect to the True Variance of $\tilde{\theta}_2$, Under the Double Exponential Distribution

Table 4.4 Coverage Probabilities of Jackknife Confidence Intervals on ρ_2 for Nominal Values of 90% and 95% (Denoted $CP(90\%)$ and $CP(95\%)$ Respectively), and Percentage Relative Biases of $v_J(\tilde{\theta}_2)$ with Respect to the True Variance of $\tilde{\theta}_2$, Under the Uniform Distribution

CHAPTER 1

Introduction

The demand for reliable small area statistics, both at the Federal and Provincial levels, has received considerable attention in recent years, due to their increased use in formulating policies and programs. Survey data are combined efficiently with census data and administrative records to produce small area estimates. The overwhelming interest in this topic has given rise to two international symposia devoted entirely to small area statistics, the proceedings of which can be found in Platek, Rao, Särndal and Singh (1987), and in NIDA (1979). Section 1.1 of this chapter presents an overview of small area methodology to date, while Section 1.2 provides the thesis outline.

1.1 An Overview of Small Area Methodology

The focus of most surveys is to provide estimates of one or more characteristics of interest, at the population level. At times, it is also of interest to produce similar estimates for subpopulations or domains. Purcell and Kish (1980) classify domains by their size relative to the entire population as follows: (i) "Major domains" consist of $1/10$ or more of the population units; (ii) "Minor domains" consist of between $1/10$ and $1/100$ of the population units; (iii) "Mini domains" consist of between $1/100$ and $1/10000$ of the population units; (iv) "Rare domains" consist of less than $1/10000$ of the population units. For a major domain, the usual survey estimators, based only on the data from the domain, are reliable. However, frequently for minor domains, and usually for mini domains, those same estimates break down in the sense that their standard errors are too large due to insufficient sample size. Alternative estimators can be found by "borrowing strength" from other related small areas through the use of supplementary data, such as census

or administrative data. These new estimators show a marked improvement in efficiency over the traditional survey estimators. Finally, for rare domains, the usual survey estimators are basically useless and entirely new methods are required. We shall consider minor domains and mini domains to be small areas, and focus on them only.

1.1.1 Estimation of Small Area Means

Suppose that a target population consisting of N units is divided into t mutually exclusive and exhaustive small areas. Let m_i ; $i = 1, \dots, t$ be the number of primary sampling units (PSUs) drawn from M_i population PSUs within the i^{th} small area. Further, suppose that n_{ij} secondary sampling units (SSUs) are drawn from N_{ij} population SSUs in the j^{th} PSU within the i^{th} small area. Often, the parameters of interest are the small area means for a specific characteristic y , i.e.,

$$\bar{Y}_{i..} = N_{i..}^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk}$$

where $N_{i..} = \sum_{j=1}^{M_i} N_{ij}$ and y_{ijk} ; $i = 1, \dots, t$; $j = 1, \dots, M_i$; $k = 1, \dots, N_{ij}$ is the value of y for the k^{th} SSU within the j^{th} PSU within the i^{th} small area. This thesis is mainly concerned with the estimation of small area means, under the above sampling design.

Several methods have been proposed as viable ways of estimating small area means. These can be classified into three distinct groups: (i) symptomatic methods, (ii) traditional sample survey methods, and (iii) methods which “borrow strength” from other small areas. The three methods are summarized briefly below in the context of the two-stage sampling design described above.

(i) Symptomatic Methods

Symptomatic methods utilize current administrative data, such as births and deaths, combined with past census data, to produce estimates of population characteristics. The reliability of these methods clearly depends on the quality and coverage of the administrative files being used. Harter (1983) gives a detailed description of some of these methods. They include: the Census Components Methods I and II (U.S. Bureau of the Census (1949), (1966), (1969)), the Vital Rates Method (Bogue (1950)), the Housing Unit Method (Rives (1976)), and the Composite Method (Bogue and Duncan (1959)).

(ii) Traditional Sample Survey Methods

It is possible to obtain reasonable estimates for $\bar{Y}_{i..}$ using any of the estimators that follow, provided the sample size within the small area being estimated is adequately large. None of these estimators "borrow strength" from other small areas except for the almost unbiased regression estimator, which does so in a weak sense only.

Direct Survey Estimators: For the small area in question, assume that simple random sampling (SRS) has been used at each stage, i.e., for drawing both the m_i PSUs and the n_{ij} SSUs in the i^{th} sampled PSU. Then, two direct survey estimators for $\bar{Y}_{i..}$ are given by

$$\hat{\bar{Y}}_i(DIR) = N_i^{-1} m_i^{-1} M_i \sum_{j \in s_i} N_{ij} y_{ij}; \quad i = 1, \dots, t \quad (1.1)$$

or, alternatively,

$$\hat{\bar{Y}}_i(DIR1) = \left(\sum_{j \in s_i} N_{ij} \right)^{-1} \sum_{j \in s_i} N_{ij} y_{ij}; \quad i = 1, \dots, t \quad (1.2)$$

where s_i is the set of m_i sampled PSUs within the i^{th} small area, $y_{ij} = n_{ij}^{-1} \sum_{k \in s_{ij}} y_{ijk}$ and s_{ij} is the set of n_{ij} sampled SSUs within the j^{th} sampled PSU within the i^{th}

small area. The first of these direct estimators is unbiased, although it is likely to yield a larger standard error than the second estimator.

Ratio Estimators: If the y -variable is positively correlated with some auxiliary x -variable, and the auxiliary variable is observed for all units in the population at the small area level, then the ratio estimator is appropriate. It is given by

$$\begin{aligned}\widehat{Y}_i(RATIO) &= \overline{X}_{i..} \widehat{Y}_i(DIR) / \widehat{X}_i(DIR) \\ &= \overline{X}_{i..} (\sum_{j \in s_i} N_{ij} \bar{y}_{ij.}) / (\sum_{j \in s_i} N_{ij} \bar{x}_{ij.}); \quad i = 1, \dots, t.\end{aligned}\quad (1.3)$$

Here $\overline{X}_{i..} = (\sum_{j=1}^{M_i} N_{ij})^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} x_{ijk}$ and x_{ijk} is the value of the auxiliary x -variable for the k^{th} SSU within the j^{th} PSU within the i^{th} small area. Also, $\bar{x}_{ij.}$ is defined analogously to $\bar{y}_{ij.}$ as is $\widehat{X}_i(DIR)$ to $\widehat{Y}_i(DIR)$. Although this estimator is biased, it is an improvement over the direct estimator in the sense of reduced standard error.

Almost Unbiased Regression Estimators: Like the ratio estimator, the almost unbiased regression estimator improves upon the direct survey estimator when the y -variable is positively correlated with the x -variable. We have

$$\widehat{Y}_i(REG) = \widehat{Y}_i(DIR) + (\overline{X}_{i..} - \widehat{X}_i(DIR)) \widehat{Y} / \widehat{X}; \quad i = 1, \dots, t \quad (1.4)$$

where $\widehat{Y}$ is an estimator of the overall population mean $\bar{Y}$ given by

$$\widehat{Y} = N_{..}^{-1} \sum_{i=1}^t m_i^{-1} M_i \sum_{j \in s_i} N_{ij} \bar{y}_{ij.}$$

or, alternatively, by

$$\widehat{Y} = (\sum_{i=1}^t m_i^{-1} M_i \sum_{j \in s_i} N_{ij})^{-1} \sum_{i=1}^t m_i^{-1} M_i \sum_{j \in s_i} N_{ij} \bar{y}_{ij.}$$

where $N_{..} = \sum_{i=1}^t N_i$ and $\widehat{\bar{X}}$ is defined analogously to $\widehat{\bar{Y}}$. Although $\widehat{\bar{Y}}_i(REG)$ is also biased, the bias is small since $\widehat{\bar{Y}}/\widehat{\bar{X}} \doteq \bar{Y}/\bar{X}$ when the overall sample size is large, where $\bar{Y}/\bar{X}$ is the ratio of population means.

(iii) Methods Which Borrow Strength From Other Small Areas

A few of these are outlined below.

Ratio Synthetic Estimators: Unlike direct survey estimators, synthetic estimators “borrow strength” from related small areas. Assuming that $\widehat{\bar{Y}}_i(DIR)/\widehat{\bar{X}}_i(DIR) \doteq \widehat{\bar{Y}}/\widehat{\bar{X}}$, a ratio synthetic estimator is formed by replacing $\widehat{\bar{Y}}_i(DIR)/\widehat{\bar{X}}_i(DIR)$ in (1.3) by $\widehat{\bar{Y}}/\widehat{\bar{X}}$ to obtain

$$\widehat{\bar{Y}}_i(SYN) = \bar{X}_{i..} \widehat{\bar{Y}}/\widehat{\bar{X}}; \quad i = 1, \dots, t. \quad (1.5)$$

This estimator can be heavily biased if the above assumption is violated. A drawback of synthetic estimation is that it does not directly use the sample data for the small area of interest.

Composite Estimators: A composite estimator of $\bar{Y}_{i..}$ is obtained by taking a weighted combination of two or more different estimators, assigning the weights so that they minimize the overall mean square error (MSE) (see Schaible et al. (1977)). For example, we can combine the direct survey estimator, $\widehat{\bar{Y}}_i(DIR)$, with the ratio synthetic estimator to get

$$\widehat{\bar{Y}}_i(COMP) = \alpha_i \widehat{\bar{Y}}_i(DIR) + (1 - \alpha_i) \widehat{\bar{Y}}_i(SYN); \quad i = 1, \dots, t \quad (1.6)$$

where

$$\alpha_i = \text{MSE}[\widehat{\bar{Y}}_i(SYN)] / [\text{var}(\widehat{\bar{Y}}_i(DIR)) + \text{MSE}(\widehat{\bar{Y}}_i(SYN))]$$

is the optimal weight, assuming

$$\text{cov}[\widehat{\bar{Y}}_i(DIR), \widehat{\bar{Y}}_i(SYN)] = 0.$$

The latter assumption is likely to hold in practice since the direct estimator depends only on the sample data in the small area of interest, while the synthetic estimator is based on all the sample data. The optimal weight depends on $\text{var}[\hat{Y}_i(DIR)]$ and $\text{MSE}[\hat{Y}_i(SYN)]$, which may both be estimated, although estimation of the latter is often difficult due to the bias of $\hat{Y}_i(SYN)$.

Sample Size Dependent Estimators: A sample size dependent estimator is a composite estimator where the weights δ_i ; $i = 1, \dots, t$ depend only on the sample size (see Drew et al. (1982)). An estimator of this type is given by

$$\hat{Y}_i(SSD) = \delta_i \hat{Y}_i(DIR) + (1 - \delta_i) \hat{Y}_i(SYN); \quad i = 1, \dots, t \quad (1.7)$$

where

$$\delta_i = \begin{cases} 1 & \text{if } \hat{N}_{i.} \geq N_{i.} \\ \hat{N}_{i.}/N_{i.} & \text{if } \hat{N}_{i.} < N_{i.} \end{cases}$$

and where $\hat{N}_{i.} = m_i^{-1} M_i \sum_{j \in \sigma_i} N_{ij}$ is the unbiased estimator of $N_{i.}$. A disadvantage of the sample size dependent estimator is that if $\hat{N}_{i.} > N_{i.}$, then the estimator reduces to $\hat{Y}_i(DIR)$, and so fails to borrow strength from other small areas.

Särndal and Hidiroglou (1989) suggest an alternative sample size dependent estimator, which, for our setup, reduces to a weighted combination of the almost unbiased regression estimator and the ratio synthetic estimator. Thus,

$$\hat{Y}_i(SH) = \delta_i \hat{Y}_i(REG) + (1 - \delta_i) \hat{Y}_i(SYN); \quad i = 1, \dots, t. \quad (1.8)$$

This estimator reduces to $\hat{Y}_i(REG)$ if $\hat{N}_{i.} > N_{i.}$, in which case it borrows strength from other small areas in a weak sense only.

Best Linear Unbiased Predictors (BLUPs): These types of estimators are the focus of this thesis. Several models have been proposed, involving random

effects, under which the estimation of the i^{th} small area mean has been investigated. A review of two of these follows. First, however, we note that BLUPs can be viewed as composite estimators, where the weights depend on the ratios of the variances of the random effects involved in the model in question. These variances are also referred to as “variance components”.

The Fay-Herriot Model: The main feature of the model suggested by Fay and Herriot (1979) is that it only uses direct survey estimators and their variances at the small area level. The model can be described as follows: Suppose $\mathbf{y} = (y_1, \dots, y_t)'$ denotes a t -vector of independent estimators of the small area mean vector $\bar{\mathbf{Y}} = (\bar{Y}_1, \dots, \bar{Y}_t)'$. The elements of $\bar{\mathbf{Y}}$ are assumed to be independent $N(\mathbf{x}_i' \boldsymbol{\beta}, A)$ random variables where A is known, $\mathbf{x}_i' = (x_{i1}, \dots, x_{ip})$ is a $1 \times p$ vector of benchmark variables related to $\bar{Y}_i$ and available for each small area, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a p -vector of unknown regression parameters. In addition, conditional on $\bar{\mathbf{Y}}$, the vector $\mathbf{y}$ is assumed to be $N(\mathbf{Y}, \mathbf{D})$ where $\mathbf{D} = \text{Diag}(D_i)_{1 \leq i \leq t}$ and each D_i is known. The Fay-Herriot model can be restated in terms of the following linear model:

$$y_i = \bar{Y}_i + c_i. \quad (1.9)$$

where

$$\bar{Y}_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i; \quad i = 1, \dots, t$$

and where c_i and u_i are stochastically independent with zero means and variances D_i and A respectively. Prasad and Rao (1990) show that the BLUP of $\bar{Y}_i$ is given by

$$\begin{aligned} \tilde{\bar{Y}}_i(FH) &= \mathbf{x}_i' \tilde{\boldsymbol{\beta}} + \gamma_i(\bar{y}_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}) \\ &= \gamma_i y_i + (1 - \gamma_i) \mathbf{x}_i' \tilde{\boldsymbol{\beta}}; \quad i = 1, \dots, t \end{aligned} \quad (1.10)$$

where $\gamma_i = A/(A + D_i)$ and $\tilde{\beta}$ is the generalized least squares estimator (GLSE) of β . Note that $\tilde{Y}_i(FH)$ is actually a composite estimator, and tends to the direct survey estimator $\bar{y}_i$ as $D_i/(A + D_i)$ tends to zero, and to the synthetic estimator $\mathbf{x}'_i\tilde{\beta}$ as $A/(A + D_i)$ tends to zero.

The One-Fold Nested Error Regression Model With Equal Error Variances: To motivate the two-fold nested error regression model studied in this thesis, we first introduce the one-fold nested error regression model. Battese, Harter and Fuller (1988) propose the one-fold nested error regression model as a means of predicting average hectares of corn in a given county, using survey data as well as Landsat satellite data, from north-central Iowa. The model suggested can be represented by

$$y_{ij} = \mathbf{x}'_{ij}\beta + u_i + e_{ij}; \quad i = 1, \dots, t; \quad j = 1, \dots, n_i \quad (1.11)$$

where y_{ij} is the value of the characteristic of interest for the j^{th} sampled unit in the i^{th} small area, $\mathbf{x}'_{ij} = (x_{ij1}, \dots, x_{ijp})$ is a $1 \times p$ vector of corresponding auxiliary variables, and $\beta = (\beta_1, \dots, \beta_p)'$ is a $p \times 1$ vector of unknown parameters. The random errors u_i ; $i = 1, \dots, t$ are assumed to be independent of the errors e_{ij} ; $i = 1, \dots, t$; $j = 1, \dots, n_i$ and each are assumed to be independent and identically distributed (i.i.d) with zero means and known variances σ_u^2 and σ_e^2 respectively.

Prasad and Rao (1990) show that if N_i is large, or equivalently $n_i N_i^{-1}$ is negligible, where N_i is the number of population units in the i^{th} small area, then the BLUP of $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$ is given by

$$\begin{aligned} \tilde{Y}_i(BF) &= \bar{\mathbf{X}}'_i \tilde{\beta} + \gamma_i (\bar{y}_i - \bar{\mathbf{x}}'_i \tilde{\beta}) \\ &= \gamma_i [\bar{y}_i + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_i)' \tilde{\beta}] + (1 - \gamma_i) \bar{\mathbf{X}}'_i \tilde{\beta}; \quad i = 1, \dots, t \end{aligned} \quad (1.12)$$

where

$$\begin{aligned}\gamma_i &= \sigma_u^2(\sigma_u^2 + \sigma_e^2 n_i^{-1})^{-1}, \\ y_i &= n_i^{-1} \sum_{j=1}^{n_i} y_{ij}, \\ \bar{\mathbf{x}}_i &= n_i^{-1} \sum_{j=1}^{n_i} \mathbf{x}_{ij}, \\ \mathbf{X}_i &= N_i^{-1} \sum_{j=1}^{N_i} \mathbf{x}_{ij}\end{aligned}$$

and $\hat{\beta}$ is the GLSE of β . Note that $\tilde{\bar{Y}}_i(BF)$ is actually a composite estimator, and tends to the almost unbiased regression estimator $y_i + (\mathbf{X}_i - \bar{\mathbf{x}}_i)' \hat{\beta}$ as $\sigma_e^2 n_i^{-1}(\sigma_u^2 + \sigma_e^2 n_i^{-1})^{-1}$ tends to zero, and to the synthetic estimator $\mathbf{X}_i' \hat{\beta}$ as $\sigma_u^2(\sigma_u^2 + \sigma_e^2 n_i^{-1})^{-1}$ tends to zero.

If $n_i N_i^{-1}$ is nonnegligible, Prasad and Rao (1990) give the BLUP of $\bar{Y}_i$, as being

$$\tilde{\bar{Y}}_i^F(BF) = n_i N_i^{-1} y_i + (N_i - n_i) N_i^{-1} \tilde{\bar{Y}}_i^*(BF) \quad (1.13)$$

where $\tilde{\bar{Y}}_i^*(BF)$ is given by (1.12) with $\mathbf{X}_i$ replaced by $\mathbf{x}_i^*$, the mean of the $\mathbf{x}_{ij}$ for the $N_i - n_i$ nonsampled units.

In deriving the BLUPs under the last two models, the variance components, A in (1.9) and σ_u^2 and σ_e^2 in (1.11), are assumed known. More often than not they are unknown, and must be estimated from sample data, using some reliable method. Since the expressions for the BLUPs, given by (1.10), (1.12) and (1.13), depend on the unknown variance components, they (the components) must be replaced by their estimators in the expressions for the BLUPs, resulting in what are known as empirical BLUPs or EBLUPs, denoted by $\hat{\bar{Y}}_i$. In this thesis, we confine ourselves to quadratic estimators of variance components obtained by the well-known fitting of constants method, also known as Henderson's method 3 (see, Searle (1971), Chapter 10). This method provides consistent estimators which

are even, translation invariant functions of the y_{ij} 's (see Definition 2.6 in Chapter 2), and which, when substituted into the expressions for the BLUPs, maintain the unbiasedness of the EBLUPs for symmetrically distributed, not necessarily normal, errors.

It is interesting to note that the transformation method of estimating variance components, as suggested by Fuller and Battese (1973), produces estimators identical to those that result from using Henderson's method. For instance, Fuller and Battese (1973) consider the one-fold model, given by (1.11), and apply their method to obtain quadratic estimators of σ_e^2 and σ_u^2 . The details of the method follows.

Estimating σ_e^2 involves a regression of $y_{ij} - y_i$ on $\mathbf{x}_{ij} - \bar{\mathbf{x}}_i$, as well as calculating the corresponding residual sum of squares, given by $\sum_{i=1}^I \sum_{j=1}^{n_i} \tilde{e}_{ij}^2$, where $\tilde{e}_{ij}$ is the appropriate residual for this regression. Then, using the method of moments, one obtains the unbiased estimator $\tilde{\sigma}_e^2$ for σ_e^2 . Thus,

$$\tilde{\sigma}_e^2 = (n - p - t + \lambda)^{-1} \sum_{i=1}^I \sum_{j=1}^{n_i} \tilde{e}_{ij}^2, \quad (1.14)$$

where $n = \sum_{i=1}^I n_i$, and $\lambda = 1$ if the model includes an intercept term and $\lambda = 0$ otherwise.

Estimating σ_u^2 involves a regression of y_{ij} on $\mathbf{x}_{ij}$, where this time, the corresponding residual sum of squares is given by $\sum_{i=1}^I \sum_{j=1}^{n_i} \tilde{u}_{ij}^2$, where $\tilde{u}_{ij}$ is the appropriate residual for this regression. Again, using the method of moments, one obtains the unbiased estimator $\tilde{\sigma}_u^2$ for σ_u^2 . Thus,

$$\tilde{\sigma}_u^2 = n_*^{-1} \left[\sum_{i=1}^I \sum_{j=1}^{n_i} \tilde{u}_{ij}^2 - (n - p) \tilde{\sigma}_e^2 \right] \quad (1.15)$$

where

$$n_* = n - \sum_{i=1}^t n_i^2 \mathbf{x}_i' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_i$$

and

$$\mathbf{X} = \begin{matrix} \text{col} \\ 1 \leq i \leq t \end{matrix} \begin{matrix} \text{col} \\ 1 \leq j \leq n_i \end{matrix} (\mathbf{x}_{ij}').$$

It is possible for $\hat{\sigma}_u^2$ to take on negative values, and thus the estimator is truncated to zero whenever $\hat{\sigma}_u^2$ is less than zero. The truncated estimator is defined as $\hat{\sigma}_u^2 = \max(0, \hat{\sigma}_u^2)$. To maintain consistency of notation, we define $\hat{\sigma}_t^2 = \hat{\sigma}_t^2$.

Under the Fay-Herriot model given by (1.9), Prasad and Rao (1990) find an unbiased quadratic estimator $\tilde{A}$ of A , given by

$$\tilde{A} = (t - p)^{-1} \left[\sum_{i=1}^t \tilde{u}_i^2 - \sum_{i=1}^t D_i (1 - \mathbf{x}_i' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_i) \right], \quad (1.16)$$

where $\tilde{u}_i = y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{\bar{y}}$, and $\mathbf{X} = \begin{matrix} \text{col} \\ 1 \leq i \leq t \end{matrix} (\mathbf{x}_i')$. The corresponding truncated estimator is given by $\hat{A} = \max(0, \tilde{A})$.

1.1.2 A Second Order Approximation to the MSE of the EBLUP

In the previous section, the method for finding EBLUPs of small area means under two different models was outlined. It is of interest to find an expression for the MSE of the EBLUP, denoted $\text{MSE}(\hat{\bar{Y}}_i)$. Unfortunately, an exact expression is difficult to obtain in most cases, and thus an approximation is required. One can write the following decomposition, assuming normality of the random errors:

$$\text{MSE}(\hat{\bar{Y}}_i) = E(\hat{\bar{Y}}_i - \bar{Y}_i)^2 = E(\tilde{\bar{Y}}_i - \bar{Y}_i)^2 + E(\hat{\bar{Y}}_i - \tilde{\bar{Y}}_i)^2$$

since Kackar and Harville (1984) show that the crossproduct term, $2E(\tilde{\bar{Y}}_i - \bar{Y}_i)(\hat{\bar{Y}}_i - \tilde{\bar{Y}}_i)$, is zero under the assumption of normality of the random errors, and translation invariance of the variance components (see Definition 2.6

in Chapter 2). Now, the first term on the right hand side of the above expression is simply $\text{MSE}(\tilde{\bar{Y}}_i)$, for which one can always find an exact expression using the methods of Henderson (1975). The second term on the right hand side can only be approximated in most cases. Kackar and Harville (1984) give an approximation to this term. Prasad and Rao (1990) give a further refinement to the approximation under certain regularity conditions, and justify that any neglected terms are of order $o(t^{-1})$, for large t . Using these results, Prasad and Rao give the explicit expressions for the approximations under models (1.9) and (1.11).

For the Fay-Herriot model, we have

$$\begin{aligned}\text{MSE}[\hat{\bar{Y}}_i(FH)] &\doteq AD_i(A + D_i)^{-1} + D_i^2(A + D_i)^{-2}\mathbf{x}_i'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}_i \\ &\quad + D_i^2(A + D_i)^{-3}\text{var}(\tilde{A}) \\ &= g_{1i}(A) + g_{2i}(A) + g_{3i}(A),\end{aligned}\tag{1.17}$$

say, where $\mathbf{V} = \text{Diag} \left(A + D_i \right)_{1 \leq i \leq t}$ is the variance-covariance matrix of $\mathbf{y}$. Further, assuming normality, the variance of $\tilde{A}$ is given by:

$$\text{var}(\tilde{A}) \doteq 2t^{-1}[A^2 + 2A\sum_{i=1}^t D_i/t + \sum_{i=1}^t D_i^2/t],\tag{1.18}$$

neglecting the terms of order $o(t^{-1})$. Note that $g_{1i}(A) + g_{2i}(A) = \text{MSE}[\tilde{\bar{Y}}_i(FH)]$, where $\text{MSE}[\tilde{\bar{Y}}_i(FH)]$ is the MSE of the BLUP.

For the one-fold nested error regression model, we have

$$\begin{aligned}\text{MSE}[\hat{\bar{Y}}_i(BF)] &\doteq (1 - \gamma_i)\sigma_u^2 + (\mathbf{X}_i - \gamma_i\mathbf{x}_i)'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}_i - \gamma_i\mathbf{x}_i) \\ &\quad + n_i^{-2}(\sigma_u^2 + \sigma_e^2/n_i)^{-3} \text{var}(\tilde{\sigma}_e^2\sigma_u^2 - \tilde{\sigma}_u^2\sigma_e^2) \\ &= g_{1i}(\sigma^2) + g_{2i}(\sigma^2) + g_{3i}(\sigma^2),\end{aligned}\tag{1.19}$$

say, where $\mathbf{V} = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_u^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}'$ is the variance-covariance matrix of $\mathbf{y} = \text{col}_{1 \leq i \leq t} \text{col}_{1 \leq j \leq n_i} (y_{ij})$ and

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) = (\sigma_e^2)^{-1} \left\{ \sum_i \sum_j \mathbf{x}_{ij} \mathbf{x}_{ij}' - \sum_i \gamma_i n_i \mathbf{x}_i \mathbf{x}_i' \right\}.$$

Here $\mathbf{I}_{n_i}$ represents an identity matrix of order n_i and $\mathbf{1}_{n_i}$ is an $n_i \times 1$ column vector consisting entirely of ones. Furthermore, under normality of the u_i 's and the e_{ij} 's, Battese and Fuller (1981) show that

$$\text{var}(\hat{\sigma}_e^2) = 2(n - t - p + \lambda)^{-1} \sigma_e^4, \quad (1.20)$$

$$\text{var}(\hat{\sigma}_u^2) = 2n_*^{-2} [(n - t - p + \lambda)^{-1} (t - \lambda)(n - p) \sigma_e^4 + n_{..} \sigma_u^4 + 2n_* \sigma_e^2 \sigma_u^2] \quad (1.21)$$

and

$$\text{cov}(\hat{\sigma}_e^2, \hat{\sigma}_u^2) = -n_*^{-1} (t - \lambda) \text{var}(\hat{\sigma}_e^2), \quad (1.22)$$

where

$$n_{..} = \text{trace}[\mathbf{M}\mathbf{Z}\mathbf{Z}']^2,$$

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

and

$$\mathbf{Z} = \text{Diag} (\mathbf{1}_{n_i}).$$

Finally, Prasad and Rao (1990) find an estimator of the mean square error of the EBLUP, denoted $\text{mse}(\hat{\bar{Y}}_t)$, having a bias of order $o(t^{-1})$. For the Fay-Herriot model, the estimator turns out to be

$$\text{mse}[\hat{\bar{Y}}_t(FH)] = g_{1t}(\hat{A}) + g_{2t}(\hat{A}) + 2g_{3t}(\hat{A}) \quad (1.23)$$

where

$$g_{\ell t}(\hat{A}) = g_{\ell t}(A)|_{A=\hat{A}}; \quad \ell = 1, 2, 3.$$

Likewise, for the one-fold model, we have

$$\text{mse}[\hat{Y}_i(BF)] = g_{1i}(\hat{\sigma}^2) + g_{2i}(\hat{\sigma}^2) + 2g_{3i}(\hat{\sigma}^2) \quad (1.24)$$

where $\hat{\sigma}^2 = (\hat{\sigma}_e^2, \hat{\sigma}_u^2)'$ and $g_{\ell i}(\hat{\sigma}^2) = g_{\ell i}(\sigma^2)|_{\sigma^2 = \hat{\sigma}^2}$; $\ell = 1, 2, 3$.

1.1.3 Jackknife Estimation of Variances of Variance Components and Robust Confidence Intervals on Functions of Variance Components

In this section, we confine ourselves to the one-fold nested error regression model with one auxiliary variable, i.e., $p = 2$, $x_{ij1} = 1$, $x_{ij2} = x_{ij}$, $\beta_1 = \alpha$ and $\beta_2 = \beta$. In the last section, expressions for the variances of $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$ and the covariance between them are given under the assumption of normality of the random errors. It is useful, however, to obtain estimators of the variances and the covariance, robust to nonnormality. To this end, Prasad (1985) devises alternative estimators, utilizing the jackknife method.

Prasad begins by modifying the estimator $\hat{\sigma}_u^2$ of σ_u^2 , along the lines of Arvesen (1969), in order to facilitate the construction of weighted jackknife estimators. The new estimator is given by

$$\sigma_u^{*2} = n_e^{-1} \left[\sum_{i=1}^i \sum_{j=1}^{n_i} u_{ij}^{*2} - \sum_{i=1}^i \lambda_{1i}^{-1} \lambda_{2i} \sum_{j=1}^{n_i} c_{ij}^{*2} \right]$$

where

$$c_{ij}^* = (y_{ij} - y_i) - \beta^*(x_{ij} - \bar{x}_i),$$

$$u_{ij}^* = y_{ij} - \hat{\alpha} - \hat{\beta}x_{ij},$$

$$\beta^* = \left[\sum_{i=1}^i \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \right]^{-1} \left[\sum_{i=1}^i \sum_{j=1}^{n_i} (y_{ij} - y_i)(x_{ij} - \bar{x}_i) \right],$$

$$\hat{\beta} = \left[\sum_{i=1}^i \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 \right]^{-1} \left[\sum_{i=1}^i \sum_{j=1}^{n_i} (y_{ij} - \bar{y})(x_{ij} - \bar{x}) \right],$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x},$$

$$\begin{aligned}
\lambda_{1i} &= n_i - 1 - [\sum_{j=1}^i \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2]^{-1} [\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2], \\
\lambda_{2i} &= n^{-1} n_i (n - 1) - [\sum_{j=1}^i \sum_{j=1}^{n_i} (x_{ij} - x)^2]^{-1} [\sum_{j=1}^{n_i} (x_{ij} - x)^2], \\
\bar{x}_i &= n_i^{-1} \sum_{j=1}^{n_i} x_{ij}, \\
x &= n^{-1} \sum_{i=1}^t \sum_{j=1}^{n_i} x_{ij}, \\
\bar{y}_i &= n_i^{-1} \sum_{j=1}^{n_i} y_{ij}, \\
\bar{y} &= n^{-1} \sum_{i=1}^t \sum_{j=1}^{n_i} y_{ij}
\end{aligned}$$

and

$$n_* = n - n^{-1} \sum_{i=1}^t n_i^2 - [\sum_{i=1}^t \sum_{j=1}^{n_i} (x_{ij} - x)^2]^{-1} [\sum_{i=1}^t n_i^2 (\bar{x}_i - x)^2].$$

The estimator $\hat{\sigma}_\epsilon^2$ of σ_ϵ^2 remains unchanged, but is denoted by σ_ϵ^{*2} , for notational consistency.

The jackknife procedure is based on deleting the residuals $\{c_{ij}^*, u_{ij}^*; j = 1, \dots, n_i\}$, in turn for $i = 1, \dots, t$, and then computing the modified estimators of σ_ϵ^2 and σ_u^2 , giving

$$\sigma_\epsilon^{*2}(-i) = (\sum_{i=1}^t \lambda_{1i} - \lambda_{1i})^{-1} [\sum_{i=1}^t \sum_{j=1}^{n_i} c_{ij}^{*2} - \sum_{j=1}^{n_i} c_{ij}^{*2}]$$

and

$$\begin{aligned}
\sigma_u^{*2}(-i) &= (n_* - n_i^*)^{-1} [(\sum_{i=1}^t \sum_{j=1}^{n_i} u_{ij}^{*2} - \sum_{i=1}^t \lambda_{1i}^{-1} \lambda_{2i} \sum_{j=1}^{n_i} c_{ij}^{*2}) \\
&\quad - (\sum_{j=1}^{n_i} u_{ij}^{*2} - \lambda_{1i}^{-1} \lambda_{2i} \sum_{j=1}^{n_i} c_{ij}^{*2})]
\end{aligned}$$

where

$$\begin{aligned}
\sum_{i=1}^t \lambda_{1i} &= n - t - 1, \\
n_i^* &= n_i - 2n^{-1} n_i^2 + n^{-2} n_i \sum_{i=1}^t n_i^2
\end{aligned}$$

$$\begin{aligned}
& + [\sum_{i=1}^t \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2]^{-2} [\sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2] [\sum_{i=1}^t n_i^2 (\bar{x}_i - \bar{x})^2] \\
& + 2 [\sum_{i=1}^t \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2]^{-1} n_i (\bar{x}_i - \bar{x}) n^{-1} \sum_{i=1}^t n_i^2 (\bar{x}_i - \bar{x}) \\
& - 2 [\sum_{i=1}^t \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2]^{-1} n_i^2 (\bar{x}_i - \bar{x})^2,
\end{aligned}$$

and $n_* = \sum_{i=1}^t n_i^*$. Note that the estimators $\sigma_c^{*2}(-i)$ and $\sigma_u^{*2}(-i)$ are both unbiased. In addition, σ_c^{*2} and σ_u^{*2} can be expressed as weighted sums of $\sigma_c^{*2}(-i)$ and $\sigma_u^{*2}(-i)$ respectively, i.e.,

$$\sigma_c^{*2} = (t-1)^{-1} \sum_{i=1}^t a_i \sigma_c^{*2}(-i)$$

and

$$\sigma_u^{*2} = (t-1)^{-1} \sum_{i=1}^t b_i \sigma_u^{*2}(-i)$$

where

$$a_i = 1 - (n-t-1)^{-1} \lambda_{1i}$$

and

$$b_i = 1 - n_*^{-1} n_i^*.$$

Now, robust estimators of $\text{var}(\sigma_c^{*2})$, $\text{var}(\sigma_u^{*2})$ and $\text{cov}(\sigma_c^{*2}, \sigma_u^{*2})$ are given by

$$\begin{aligned}
v_J(\sigma_c^{*2}) &= [t(t-1)]^{-1} \sum_{i=1}^t [Q_i(c) - \bar{Q}(c)]^2 \\
&= t(t-1)^{-1} \sum_{i=1}^t a_i^2 [\sigma_c^{*2}(-i) - \sigma_c^{*2}]^2,
\end{aligned} \tag{1.25}$$

$$\begin{aligned}
v_J(\sigma_u^{*2}) &= [t(t-1)]^{-1} \sum_{i=1}^t [Q_i(u) - \bar{Q}(u)]^2 \\
&= t(t-1)^{-1} \sum_{i=1}^t b_i^2 [\sigma_u^{*2}(-i) - \sigma_u^{*2}]^2,
\end{aligned} \tag{1.26}$$

and

$$\begin{aligned}
\text{cov}_J(\sigma_c^{*2}, \sigma_u^{*2}) &= [t(t-1)]^{-1} \sum_{i=1}^t [Q_i(c) - \bar{Q}(c)][Q_i(u) - \bar{Q}(u)] \\
&= t(t-1)^{-1} \sum_{i=1}^t a_i b_i [\sigma_c^{*2}(-i) - \sigma_c^{*2}][\sigma_u^{*2}(-i) - \sigma_u^{*2}]
\end{aligned} \tag{1.27}$$

where

$$Q_i(c) = \sigma_c^{*2} + ta_i[\sigma_c^{*2} - \sigma_c^{*2}(-i)],$$

$$Q_i(u) = \sigma_u^{*2} + tb_i[\sigma_u^{*2} - \sigma_u^{*2}(-i)],$$

are the weighted pseudo-values, and where

$$Q(c) = t^{-1} \sum_{i=1}^t Q_i(c),$$

and

$$\bar{Q}(u) = t^{-1} \sum_{i=1}^t Q_i(u).$$

The estimators (1.25)-(1.27) remain consistent as $t \rightarrow \infty$, and can be used to estimate $g_{3i}(\sigma^2)$ from expression (1.19).

Prasad (1985) then generalizes his results to estimating any nonlinear function $g(\sigma^2)$ of the variance components $\sigma^2 = (\sigma_c^2, \sigma_u^2)'$. He gives the jackknife estimator of $g(\sigma_c^2, \sigma_u^2)$ as being

$$g_J(\sigma_c^{*2}, \sigma_u^{*2}) = t^{-1} \sum_{i=1}^t Q_i(g) \quad (1.28)$$

where

$$Q_i(g) = g(\sigma_c^{*2}, \sigma_u^{*2}) - t[g\{\sigma_c^{*2} + a_i(\sigma_c^{*2}(-i) - \sigma_c^{*2}), \sigma_u^{*2} + b_i(\sigma_u^{*2}(-i) - \sigma_u^{*2})\} - g(\sigma_c^{*2}, \sigma_u^{*2})],$$

and the associated jackknife variance estimator as being

$$v_J(g) = t^{-1}(t-1)^{-1} \sum_{i=1}^t [Q_i(g) - g_J(\sigma_c^{*2}, \sigma_u^{*2})]^2. \quad (1.29)$$

The above results can be used to form robust confidence intervals on $g(\sigma_c^2, \sigma_u^2)$, i.e.,

$$g(\sigma_c^2, \sigma_u^2) \in [g_J(\sigma_c^{*2}, \sigma_u^{*2}) - t_{\frac{\alpha}{2}; t-1}(v_J(g))^{1/2}, g_J(\sigma_c^{*2}, \sigma_u^{*2}) + t_{\frac{\alpha}{2}; t-1}(v_J(g))^{1/2}] \quad (1.30)$$

where $t_{\frac{\alpha}{2}; t-1}$ is the upper $\frac{\alpha}{2}$ -point of a t -distribution with $t - 1$ degrees of freedom. The applications of robust confidence intervals on nonlinear functions of variance components are numerous. For instance, Arvesen and Schmitz (1970) consider them in the context of a balanced, one-fold nested error mean model with equal error variances, for measuring heritabilities and genetic correlations.

In the context of the one-fold model, a typical nonlinear function of σ^2 of interest is the ratio of the variance components, i.e., $g(\sigma^2) = \sigma_u^2/\sigma_e^2$. Prasad and Rao (1988) obtain a weighted jackknife version of the Spjøtvoll (1967) test of hypothesis $H_0 : \rho = \sigma_u^2/\sigma_e^2 \leq \rho_0$ against $H_A : \rho > \rho_0$ for the unbalanced one-fold nested error mean model $y_{ij} = \beta + u_i + e_{ij}$; $i = 1, \dots, t$; $j = 1, \dots, n_i$.

1.1.4 A Hierarchical Bayes Approach

The hierarchical Bayes (HB) approach has been suggested in recent literature as an alternative to the EBLUP or the classical empirical Bayes (EB) approaches.

An EB scenario is one where a vector of parameters, say $\theta = (\theta_1, \dots, \theta_k)'$, is assumed to have probability density function (p.d.f.) $\pi(\theta|\lambda)$, where the prior π is structurally known except possibly for some vector of parameters, λ . A parametric EB procedure is one where λ is then estimated from the marginal distribution of the data using some classical method, such as the method of moments or the method of maximum likelihood. The estimate of λ is substituted into the Bayes estimator of θ , given λ , to arrive at an EB estimator of θ_i , for $i = 1, \dots, k$.

The HB procedure models the prior distribution in stages. The following is an example of two-stage modelling, but the idea can be extended to many stages, without loss of generality: At the first stage, conditional on $\Lambda = \lambda$, θ is assumed to have p.d.f. $\pi(\theta|\lambda)$. At the second stage, a prior distribution, $\pi(\lambda)$, is assigned to Λ . The resulting Bayes estimator of θ_i is the HB estimator.

Interestingly enough, the three methods often lead to comparable results in the context of point estimation. For instance, the empirical Bayes estimators of small area means are essentially equivalent to EBLUPs, assuming either normality or posterior linearity. However, the HB method has a clear edge over the other two methods in terms of measuring the uncertainty associated with the estimators. Whereas there are no exact measures associated with EBLUPs or EB point estimation, and often approximations are needed, the same is not true for HB estimators. Specifically, if the parameter of interest is estimated using a posterior expectation, then the corresponding posterior variance is a natural measure of risk associated with it. However, the evaluation of the posterior expectation and variance often involves multiple numerical integration, which becomes increasingly difficult as the dimension of $\mathbf{A}$ increases.

Ghosh and Meeden (1986) discuss the EB method in the context of the one-fold nested error mean model, i.e., model (1.11) with $p = 1$, $x_{ij1} = 1$ and $\beta_1 = \beta$. They give the Bayes estimator of the small area means under squared error loss, assuming normality of the random errors. Since β and σ_u^2/σ_e^2 are unknown parameters, they find a maximum likelihood estimator (MLE) for β assuming known σ_u^2/σ_e^2 . They then find an unbiased, consistent estimator for σ_u^2/σ_e^2 and substitute it back into the MLE of β . The resulting parameter estimators are themselves substituted back into the expression for the Bayes estimator, resulting in an EB estimator which is essentially equivalent to an EBLUP. Ghosh and Meeden show that the relative increase in the Bayes risk of the EB estimator over the Bayes estimator tends to zero as the number of small areas tends to infinity.

Ghosh and Lahiri (1987) relax the above assumption of normality and assume posterior linearity of the Bayes estimator instead. They then show that the procedure remains robust to nonnormality, via a comparison of Bayes risks.

Finally, Ghosh and Lahiri (1988) derive an EB estimator of small area means under the two-fold nested error mean model, i.e.,

$$y_{ijk} = \beta + \xi_i + u_{ij} + e_{ijk}; \quad i = 1, \dots, t; \quad j = 1, \dots, M_i; \quad k = 1, \dots, N_{ij},$$

again dispensing with the normality assumption and replacing it with posterior linearity. Ghosh and Lahiri show that, once again, the relative increase in the Bayes risk tends to zero.

Malec and Sedransk (1985) consider HB inference for a general linear function of the finite population elements under three-stage sampling, assuming known variance components. This sampling scheme can be modelled as a two-fold nested error mean model, as represented above.

Ghosh and Lahiri (1989) discuss the HB method in the context of the one-fold nested error regression model, given by (1.11), for the special case of $\mathbf{x}_{ij} = \mathbf{x}_i$.

Datta and Ghosh (1990) provide a unified approach to the HB method, for general mixed linear models. They introduce a general Bayesian model, which encompasses as special cases of it, the one-fold nested error regression model, as well as many other models. An outline of the results of Datta and Ghosh are summarized below, in the context of the one-fold model.

At the first stage of modelling the priors under (1.11), one can assume that conditional on $\mathbf{B} = \boldsymbol{\beta}$ and $L = \ell$, the y_{ij} 's are assumed to be independent and distributed normally, each having mean $\mathbf{x}'_{ij}\boldsymbol{\beta}$ and variance $\ell^{-1} + (\lambda\ell)^{-1}$, where $\ell = (\sigma_e^2)^{-1}$ is assumed unknown and $\lambda = (\sigma_u^2)^{-1}\sigma_e^2$ is assumed known, i.e., $y_{ij}|\boldsymbol{\beta}, \ell \sim N(\mathbf{x}'_{ij}\boldsymbol{\beta}, \ell^{-1} + (\lambda\ell)^{-1})$. At the second stage, $\mathbf{B}$ and L are assumed to be independent of each other, and are assigned prior distributions. For instance,

one can assign an independent uniform ($\mathbb{R}^p$) distribution to $\mathbf{B}$, and a gamma ($a_0/2, b_0/2$) distribution to L .

One can partition the y_{ij} 's into sampled and nonsampled data as follows:

$$\mathbf{Y} = \begin{matrix} \text{col} \\ 1 \leq i \leq t \end{matrix} \begin{matrix} \text{col} \\ 1 \leq j \leq N_i \end{matrix} (y_{ij}) = \begin{pmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{pmatrix} = \begin{pmatrix} \begin{matrix} \text{col} & \text{col} \\ 1 \leq i \leq t & 1 \leq j \leq n_i \end{matrix} & (y_{ij}) \\ \begin{matrix} \text{col} & \text{col} \\ 1 \leq i \leq t & n_i+1 \leq j \leq N_i \end{matrix} & (y_{ij}) \end{pmatrix},$$

where n_i is the number of sampled units at the j^{th} level, and N_i is the number of population units at the j^{th} level. Then, the Bayes estimator of the linear combination $\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)}$ where $\mathbf{A}$ and $\mathbf{C}$ are known matrices, is given by the posterior expectation $\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}E[\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}]$. An explicit expression for the posterior expectation is given by Datta and Ghosh, under the general mixed linear model setup, and is shown to be essentially equivalent to the BLUP of $\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)}$. Note that for an appropriate choice of $\mathbf{A}$ and $\mathbf{C}$, the linear combination $\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)}$ reduces to a vector of small area means, which is the parameter of interest in this thesis.

A more general HB analysis assumes λ to be unknown and thus assigns a prior to ΛL (having realization $\lambda\ell$). For instance, ΛL can be assumed to be independent of $\mathbf{B}$ and L , and distributed as a gamma ($a/2, b/2$) random variable. Using the HB method offered by Datta and Ghosh, one can find the conditional distribution of Λ given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$, and hence the posterior expectation,

$$E[\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] = \int E(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \Lambda = \lambda)\pi(\lambda|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)})d\lambda,$$

and the posterior variance,

$$\begin{aligned} \text{var}[\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] &= E_{\lambda|\mathbf{y}^{(1)}}[\text{var}(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \Lambda = \lambda)] \\ &\quad + \text{var}_{\lambda|\mathbf{y}^{(1)}}[E(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \Lambda = \lambda)] \end{aligned}$$

where $\pi(\lambda|Y^{(1)} = y^{(1)})$ is the posterior distribution of Λ conditional on $Y^{(1)} = y^{(1)}$.

1.2 Thesis Outline

Chapter 2 deals exclusively with the one-fold nested error regression model with unequal error variances. The model is described and a transformation is performed on it such that the random errors involved become uncorrelated, and have constant variances. Thus, the estimation of the unknown regression parameters, which depends on these variances, is simplified. Then the BLUPs of the small area means are derived for both the infinite and finite population cases, under the assumption of known variance components. The variance components are also estimated using Henderson's method of fitting constants, to enable the construction of EBLUPs, as discussed earlier. The transformation method of estimating variance components, as suggested by Fuller and Battese (1973), is shown to produce estimators identical to those that result from using Henderson's method. Assuming normality of the random errors, a second order approximation to the MSE of the EBLUP is given, as well as an estimator of the MSE having bias of order $o(t^{-1})$ for large t , the number of small areas.

Chapter 3 basically mimics the methods of Chapter 2, except that the model under consideration is the two-fold nested error regression model, with unequal error variances. This model is appropriate for the two-stage sampling design described earlier. The accuracy of the approximation to the MSE is justified, by showing that any neglected terms are of order $o(t^{-1})$, for large t . The results of a small Monte Carlo study on the accuracy of the second order approximation to the MSE of the EBLUP and the bias of the estimators of the MSE are also reported, for a special case.

Throughout Chapter 4, we assume a special case of the two-fold model described in Chapter 3. A jackknife procedure for estimating the variances of the variance component estimators, as well as the covariances between them, is presented, as well as a proof of the consistency of these estimators. The jackknife procedure is also used to form robust confidence intervals on functions of variance components. Finally, the results of a small Monte Carlo study on the robustness of these confidence intervals are reported.

In Chapter 5, the two-fold nested error regression model, with equal error variances, is viewed from an HB standpoint. The unified approach suggested by Datta and Ghosh (1990) is applied to the above model and explicit expressions are derived for the posterior expectation and variance of the small area means, conditional on the sampled data. An alternative, simpler derivation of the Datta and Ghosh general result is also given.

Finally, Chapter 6 summarizes the results of the thesis and offers suggestions for further research.

CHAPTER 2

The One-Fold Nested Error Regression Model With Unequal Error Variances

In this chapter the one-fold nested error regression model with unequal error variances is considered. A detailed description of the model is given in Section 2.1. In Section 2.2, a transformation is derived, such that the above model is reduced to a model with constant, uncorrelated error variances. The regression parameters are then estimated from the transformed model using ordinary least squares (OLS). Best linear unbiased predictors (BLUPs), as suggested by Henderson (1975), are discussed in Section 2.3, for both the infinite and finite population cases, under the assumption of known variance components. Section 2.4 deals with the estimation of the variance components via Henderson's method of fitting constants, and the ensuing empirical best linear unbiased predictors (EBLUPs) that result from substituting the estimated components back into the original expressions for the BLUPs. Finally, Sections 2.5 and 2.6 respectively deal with second order approximations to the mean square error (MSE) of the EBLUPs, and estimates of the MSE having bias of order $o(t^{-1})$ for large t , the number of small areas.

2.1 Description of the Model

Prasad and Rao (1990) investigate in great detail the unbalanced one-fold nested error regression model with constant error variances in the context of small area estimation. The model proposed here is more general. The assumption of constant error variances is relaxed, and the error variances are assumed to

be proportional to some known function of the auxiliary regression variables. A representation of the model is given by

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i + f_{ij}e_{ij}; \quad i = 1, \dots, t; \quad j = 1, \dots, n_i \quad (2.1)$$

where y_{ij} is the value of the response variable for the j^{th} sample unit in the i^{th} small area (n_i sample units are drawn from N_i population units in the i^{th} small area), $\mathbf{x}'_{ij} = (x_{ij1}, \dots, x_{ijp})$ is a $1 \times p$ vector of known auxiliary variables while $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a $p \times 1$ vector of unknown regression parameters, to be determined. The random errors u_i ; $i = 1, \dots, t$ are assumed to be independent of the random errors e_{ij} ; $i = 1, \dots, t$; $j = 1, \dots, n_i$, and each in turn is assumed independent with mean zero and respective variances σ_u^2 and σ_e^2 . These variances are henceforth referred to as "variance components". Finally, the f_{ij} 's are known, positive constants. For instance, f_{ij} can be a known, bounded function of the auxiliary variables $\mathbf{x}_{ij}$, i.e., $f_{ij} = g(\mathbf{x}_{ij})$. Choudhry and Rao (1988) find the BLUPs of the small area means for the special cases $f_{ij} = x_{ij}$ and $f_{ij} = x_{ij}^{1/2}$ assuming $p = 1$.

Remark 2.1 For the model, as it stands, the variance of the error term $f_{ij}e_{ij}$ is given by $\text{var}(f_{ij}e_{ij}) = \sigma_e^2 f_{ij}^2$. It is straightforward, however, to transform the model to one having constant error variances by dividing the existing model through by f_{ij} and obtaining

$$z_{ij} = \mathbf{w}'_{ij}\boldsymbol{\beta} + f_{ij}^{-1}u_i + e_{ij}; \quad i = 1, \dots, t; \quad j = 1, \dots, n_i \quad (2.2)$$

where $z_{ij} = f_{ij}^{-1}y_{ij}$ and $\mathbf{w}'_{ij} = (w_{ij1}, \dots, w_{ijp}) = (f_{ij}^{-1}x_{ij1}, \dots, f_{ij}^{-1}x_{ijp})$. For the model (2.2), the covariance structure can be written as

$$\text{cov}(z_{ij}, z_{i'j'}) = \begin{cases} f_{ij}^{-2}\sigma_u^2 + \sigma_e^2 & \text{if } i = i' \text{ and } j = j' \\ f_{ij}^{-1}f_{i'j'}\sigma_u^2 & \text{if } i = i' \text{ and } j \neq j' \\ 0 & \text{if } i \neq i'. \end{cases}$$

Remark 2.2 It is convenient to express the model (2.2) in matrix form, as a general mixed linear model (see Definition 2.4 ahead) as follows:

$$\mathbf{z} = \mathbf{W}\boldsymbol{\beta} + \mathbf{D}\mathbf{u} + \mathbf{e} \quad (2.3)$$

where $\mathbf{z} = \text{col}_{1 \leq i \leq t} \text{col}_{1 \leq j \leq n_i} (z_{ij}) = (z_{11}, \dots, z_{1n_1}, \dots, z_{t1}, \dots, z_{tn_t})'$, $\mathbf{W} = \text{col}_{1 \leq i \leq t} \text{col}_{1 \leq j \leq n_i} (\mathbf{w}'_{ij})$ and $\boldsymbol{\beta}$ is as before. In addition, $\mathbf{D} = \text{Diag} (\mathbf{f}_i^{-1})$ where $\mathbf{f}_i^{-1} = \text{col}_{1 \leq j \leq n_i} (f_{ij}^{-1})$. Finally, $\mathbf{u} = \text{col}_{1 \leq i \leq t} (\mathbf{u}_i)$ and $\mathbf{e} = \text{col}_{1 \leq i \leq t} \text{col}_{1 \leq j \leq n_i} (e_{ij})$. The vectors $\mathbf{u}$ and $\mathbf{e}$ are independent with mean 0 and variance-covariance matrix

$$\text{var} \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \sigma_u^2 \mathbf{I}_t & \mathbf{0} \\ \mathbf{0} & \sigma_e^2 \mathbf{I}_n \end{pmatrix} = \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix},$$

say, where $\mathbf{I}_t$ and $\mathbf{I}_n$ are identity matrices of order t and $n = \sum_{i=1}^t n_i$ respectively. We denote the variance-covariance matrix of $\mathbf{z}$ as $\mathbf{V}$ where

$$\mathbf{V} = \text{Diag} (\sigma_e^2 \mathbf{I}_{n_i} + \sigma_u^2 (\mathbf{f}_i^{-1})(\mathbf{f}_i^{-1})') = \text{Diag} (\mathbf{V}_i) \quad (2.4)$$

and $\mathbf{I}_{n_i}$ represents the identity matrix of order n_i .

2.2 Transformed Model with Uncorrelated Errors and Constant Variances

In the next section, we will be dealing, among other things, with the estimation of $\boldsymbol{\beta}$, the vector of unknown regression parameters. The focus of this section is to transform the model (2.3) so that the estimation of $\boldsymbol{\beta}$ is simplified considerably.

Fuller and Battese (1973) show that, given the model $\mathbf{z} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\psi}$ where $\boldsymbol{\psi}$ is an $n \times 1$ vector with mean 0 and variance-covariance matrix $\mathbf{V}$, it is possible to find a transformation matrix $\mathbf{T}$ such that the transformed model $\mathbf{Tz} = \mathbf{TW}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\psi}$ has uncorrelated errors $\boldsymbol{\psi}^* = \mathbf{T}\boldsymbol{\psi}$ with constant variances.

The above transformation is useful in that under the original model, forming the generalized least squares estimator of $\boldsymbol{\beta}$, given by $\tilde{\boldsymbol{\beta}} = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}^{-1}\mathbf{z}$,

involves the inversion of V , an $n \times n$ matrix, which may be prohibitively large, whereas the identical estimator for β is obtained by forming the OLS of β under the transformed model. This is given by $\tilde{\beta} = (W'T'TW)^{-1}W'T'Tz$. In addition, it is possible to test for normality of the errors, using, for example, the Shapiro-Wilk statistic on the transformed residuals.

Fuller and Battese show that the appropriate transformation matrix is given by $T = \sum_{k=1}^r \lambda_k^{-1/2} A_k$ where $\lambda_1, \dots, \lambda_r$ are the distinct eigenvalues of V , having respective multiplicities $m_1, \dots, m_r$ where $\sum_{k=1}^r m_k = r$, and where $A_k = B_k B_k'$ such that B_k is an $n \times m_k$ matrix consisting of the m_k columns of mutually orthogonal, normalized eigenvectors associated with λ_k .

For the model (2.3), $V = \text{Diag} (V_i)_{1 \leq i \leq t}$ is as stated in (2.4). Using the following lemma, obtaining the eigenvalues of V_i , and hence of V , is immediate.

Lemma 2.1 (Graybill (1969), Theorem 8.5.3, pp. 185) *Let the $k \times k$ matrix C be given by $C = D + \lambda ab'$ where $D = dI_k$, a and b are each $k \times 1$ vectors and α is a scalar such that $\alpha \neq -(\sum_{i=1}^k a_i b_i d_{ii}^{-1})^{-1}$. Then, $k-1$ characteristic roots are equal to d and one is equal to $d + \alpha \sum_{i=1}^k a_i b_i$.*

Letting $k = n_i$, $d = \sigma_e^2$, $\alpha = \sigma_u^2$ and $a = b = f_i^{-1}$ in Lemma 2.1 we get $n_i - 1$ eigenvalues of V_i equal to σ_e^2 and one equal to $\sigma_e^2 + \eta_i \sigma_u^2$ where

$$\eta_i = \sum_{j=1}^{n_i} f_{ij}^{-2} = (f_i^{-1})'(f_i^{-1}).$$

Matching the eigenvalue $\lambda_{1i} = \sigma_e^2$ with its corresponding eigenvector matrix $A_{1i} = I_{n_i} - \eta_i^{-1}(f_i^{-1})(f_i^{-1})'$ and the eigenvalue $\lambda_{2i} = \sigma_e^2 + \eta_i \sigma_u^2$ with $A_{2i} = \eta_i^{-1}(f_i^{-1})(f_i^{-1})'$, we form the required transformed model as follows: Set

$$T_i^* = \sum_{k=1}^2 \lambda_{ki}^{-1/2} A_{ki} = \sigma_e^{-1}(I_{n_i} - \eta_i^{-1}(f_i^{-1})(f_i^{-1})') + (\sigma_e^2 + \eta_i \sigma_u^2)^{-1/2} \eta_i^{-1}(f_i^{-1})(f_i^{-1})'.$$

Then, set $\mathbf{T}_i = \sigma_e \mathbf{T}_i^*$, so that the transformed model has constant variances equal to σ_e^2 . Thus $\mathbf{T}_i = \mathbf{I}_{n_i} - \rho_i \eta_i^{-1} (\mathbf{f}_i^{-1}) (\mathbf{f}_i^{-1})'$ where $\rho_i = 1 - (1 - \gamma_i)^{1/2}$ and

$$\gamma_i = \sigma_u^2 (\sigma_u^2 + \sigma_e^2 \eta_i^{-1})^{-1}.$$

Finally, premultiply the model by the transformation matrix $\mathbf{T} = \text{Diag} (\mathbf{T}_i)_{1 \leq i \leq t}$ to obtain the transformed model. Upon doing so, we find the $(i, j)^{\text{th}}$ element of the resultant model to be

$$f_{ij}^{-1} (y_{ij} - \rho_i \bar{y}_{i\cdot}) = f_{ij}^{-1} (\mathbf{x}_{ij} - \rho_i \bar{\mathbf{x}}_{i\cdot})' \boldsymbol{\beta} + \psi_{ij}^* \quad (2.5)$$

where

$$\bar{y}_{i\cdot} = \eta_i^{-1} \sum_{j=1}^{n_i} f_{ij}^{-1} z_{ij},$$

$$\bar{\mathbf{x}}_{i\cdot} = \eta_i^{-1} \sum_{j=1}^{n_i} f_{ij}^{-1} \mathbf{w}_{ij}$$

and ψ_{ij}^* is the $(i, j)^{\text{th}}$ element of $\boldsymbol{\psi}^*$. Note that the ψ_{ij}^* 's are uncorrelated with constant variances. Thus, the GLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ under (2.3) is identical to the OLS estimator under (2.5). Fuller and Battese (1973) derive the transformed model for the special case of equal error variances, i.e., $f_{ij} = 1$ for all (i, j) .

2.3 Best Linear Unbiased Prediction Under the Assumption of Known Variance Components

Assuming a superpopulation model of the form (2.1) for the N_i population units in the i^{th} small area ($i = 1, \dots, t$), we wish to predict the population mean for the i^{th} area, i.e., predict $\bar{Y}_i = \bar{\mathbf{X}}_i' \boldsymbol{\beta} + u_i + \bar{E}_i$ where $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$ and $\bar{\mathbf{X}}_i = N_i^{-1} \sum_{j=1}^{N_i} \mathbf{x}_{ij}$ are the means of the y_{ij} 's and $\mathbf{x}_{ij}$'s respectively, for the N_i population units in the i^{th} small area, and $\bar{E}_i = N_i^{-1} \sum_{j=1}^{N_i} f_{ij} e_{ij}$.

2.3.1 The Infinite Population Case

We assume here that N_i is large or that the sampling fraction n_i/N_i^{-1} is negligible, and that f_{ij} is bounded for $i = 1, \dots, t$; $j = 1, \dots, N_i$. In this case, $\bar{E}_i$ can be taken to be approximately zero by applying the Weak Law of Large Numbers for independent random variables with bounded variances to $\bar{E}_i$. (see Rao (1965), pp. 112). Therefore, $\bar{Y}_i \doteq \bar{X}_i' \beta + u_i$.

Henderson's results (1975) concerning the prediction of the linear combination $\mathbf{k}'\beta + \mathbf{m}'\mathbf{u}$ in the context of a general mixed linear model, where $\mathbf{k}$ and $\mathbf{m}$ are $p \times 1$ and $t \times 1$ known vectors respectively, are summarized in Theorem 2.1 below. First, some fundamental definitions are given, which are borrowed from Kackar (1979). The prediction of $\bar{Y}_i$ is then a direct application of Theorem 2.1. Here the term "prediction" is used rather than the term "estimation" since a random quantity is being considered instead of a parameter.

Definition 2.1 A predictor $t(\mathbf{y})$ is linear iff it can be written as $t(\mathbf{y}) = c + \ell'\mathbf{y}$ where c is a constant, ℓ is an $n \times 1$ vector of constants, and $\mathbf{y}$ is the $n \times 1$ vector of observations.

Definition 2.2 A predictor $t(\mathbf{y})$ is said to be unbiased for $\mathbf{k}'\beta + \mathbf{m}'\mathbf{u}$ iff $E(t(\mathbf{y}) - \mathbf{k}'\beta - \mathbf{m}'\mathbf{u}) = 0$.

Definition 2.3 A linear, unbiased predictor $t(\mathbf{y})$ of $\mathbf{k}'\beta + \mathbf{m}'\mathbf{u}$ is said to be the best linear unbiased predictor (BLUP) of $\mathbf{k}'\beta + \mathbf{m}'\mathbf{u}$ iff the MSE of $t(\mathbf{y})$, given by $\text{MSE}(t(\mathbf{y})) = E[t(\mathbf{y}) - \mathbf{k}'\beta - \mathbf{m}'\mathbf{u}]^2$ is less than or equal to $\text{MSE}(t^*(\mathbf{y}))$, where $t^*(\mathbf{y})$ is any other linear unbiased predictor of $\mathbf{k}'\beta + \mathbf{m}'\mathbf{u}$. Equality occurs iff $t(\mathbf{y}) = t^*(\mathbf{y})$ almost everywhere.

Definition 2.4 A general mixed linear model is of the form $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$, where $\mathbf{y}$ is an $n \times 1$ vector of observations, $\mathbf{X}$ and $\mathbf{Z}$ are

known $n \times p$ and $n \times b$ matrices respectively, and β is a $p \times 1$ vector of unknown parameters. Further, u and e are respectively $b \times 1$ and $n \times 1$ random vectors with mean 0 and variance-covariance matrix

$$\text{var} \begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix},$$

where G and R are nonsingular matrices, depending on σ^2 , the vector consisting of variance components. Finally, $\Sigma = ZGZ' + R$ is the variance-covariance matrix of y .

Theorem 2.1 Assume a general mixed linear model with known σ^2 . If X is of full rank, then $t(y) = k'\tilde{\beta} + m'GZ'\Sigma^{-1}(y - X\tilde{\beta})$ is the unique BLUP of $k'\beta + m'u$ where $\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$ is the GLSE of β and $\Sigma = ZGZ' + R$.

Proof See Henderson (1975).

The estimation of $\bar{Y}_i$ is a direct application of Theorem 2.1 with $k = \bar{X}_i$ and $m = \text{col}_{1 \leq \ell \leq t} (\delta_{i\ell})$ where $\delta_{i\ell} = 1$ if $\ell = i$ and $\delta_{i\ell} = 0$ otherwise. One obtains for the BLUP of $\bar{Y}_i$, the following explicit expression:

$$\begin{aligned} t_i(\sigma^2, y) &= \gamma_i[y_{i\cdot} + (\bar{X}_i - \bar{x}_{i\cdot})'\tilde{\beta}] + (1 - \gamma_i)\bar{X}_i'\tilde{\beta} \\ &= \bar{X}_i'\tilde{\beta} + \gamma_i(y_{i\cdot} - \bar{x}_{i\cdot}'\tilde{\beta}) \\ &= \bar{X}_i'\tilde{\beta} + \tilde{u}_i \end{aligned} \tag{2.6}$$

say, where $\sigma^2 = (\sigma_e^2, \sigma_u^2)'$.

Remark 2.3 The proof of (2.6) involves the inversion of the variance-covariance matrix V , given by (2.4). Using Lemma 2.2 below, this inverse is given by

$$V^{-1} = \text{Diag} (V_i^{-1})_{1 \leq i \leq t} \tag{2.7}$$

where

$$\mathbf{V}_i^{-1} = \sigma_e^{-2}(\mathbf{I}_{n_i} - \eta_i^{-1}\gamma_i(\mathbf{f}_i^{-1})(\mathbf{f}_i^{-1})').$$

Also, $\tilde{\beta}$ is the OLS estimator under the transformed model (2.5), as noted earlier.

Lemma 2.2 (Graybill (1969), Theorem 8.3.3, pp. 169) *Let the $k \times k$ matrix $\mathbf{C}$ be given by $\mathbf{C} = \mathbf{D} + \alpha \mathbf{a} \mathbf{b}'$ where $\mathbf{D}$ is a nonsingular diagonal matrix, $\mathbf{a}$ and $\mathbf{b}$ are each $k \times 1$ vectors, and α is a scalar such that $\alpha \neq -(\sum_{i=1}^k a_i b_i d_{ii}^{-1})^{-1}$. The inverse of $\mathbf{C}$ is then given by $\mathbf{C}^{-1} = \mathbf{D}^{-1} + \gamma \mathbf{a}^* \mathbf{b}^{*'} where $\gamma = -\alpha(1 + \alpha \sum_{i=1}^k a_i b_i d_{ii}^{-1})^{-1}$, $a_i^* = a_i d_{ii}^{-1}$, $b_i^* = b_i d_{ii}^{-1}$ and d_{ii} is the i^{th} diagonal element of $\mathbf{D}$.$*

2.3.2 The Finite Population Case

We now turn to the finite N_i case and find the BLUP of $\bar{Y}_i$ under the superpopulation model.

Theorem 2.2 *Assume a superpopulation model of the form (2.1) with known σ^2 . Then the BLUP of $\bar{Y}_i$ is given by*

$$t_i^F(\sigma^2, \mathbf{y}) = N_i^{-1} \left[\sum_{j \in s_i} y_{ij} + \sum_{j \in s_i^c} y_{ij}^* \right]$$

where s_i is the set of all sampled units in the i^{th} small area and y_{ij}^* is a predictor of y_{ij} for the j^{th} nonsampled unit within the i^{th} small area. Specifically,

$$t_i^F(\sigma^2, \mathbf{y}) = N_i^{-1} n_i y_i + N_i^{-1} (N_i - n_i) (\bar{\mathbf{x}}_i^{*'} \tilde{\beta} + \tilde{u}_i) \quad (2.8)$$

where

$$y_i = n_i^{-1} \sum_{j \in s_i} y_{ij}$$

is the mean of the n_i sampled y_{ij} 's in the i^{th} small area,

$$\bar{\mathbf{x}}_i^* = (N_i - n_i)^{-1} \sum_{j \in s_i^c} \mathbf{x}_{ij}$$

is the mean vector of the $N_i - n_i$ nonsampled $\mathbf{x}_{ij}$'s in the i^{th} small area, and

$$\hat{u}_i = \gamma_i(\bar{y}_{iw} - \bar{\mathbf{x}}'_{iw}\tilde{\boldsymbol{\beta}}).$$

Proof The proof is omitted since it is similar to the proof presented in Chapter 3, for the more complex two-fold nested error regression model.

Remark 2.4 If $n_i N_i^{-1} \rightarrow 0$, then the BLUP for the finite population case tends to the BLUP for the infinite population case, noting that $\bar{\mathbf{x}}_i^* \rightarrow \bar{\mathbf{X}}_i$.

2.4 Estimation of Variance Components

The BLUPs (2.6) and (2.8) in the previous section are obtained under the assumption that the vector of variance components $\boldsymbol{\sigma}^2 = (\sigma_e^2, \sigma_u^2)'$ is known. When the variance components are unknown, they must be estimated using some appropriate method. The estimators are then substituted for the unknown variance components in (2.6) and (2.8), resulting in what are known as empirical BLUPs, or EBLUPs. The EBLUPs remain unbiased if the estimators of σ_e^2 and σ_u^2 are both even, translation invariant functions of $\mathbf{y}$ (see Definition 2.6 ahead), and the distributions of u_i and e_{ij} are both symmetric (see Kackar and Harville (1984)).

The method of estimation used here is Henderson's method of fitting constants (see Searle (1971)) which produces consistent estimators. Interestingly enough, under the one-fold model described in this chapter and the two-fold model described in the next chapter, Henderson's method produces estimators which are identical to those formed by using the transformation method described by Fuller and Battese (1973). The latter method has the advantage that it involves the inversion of matrices of size $p \times p$ whereas the former method involves the inversion of matrices of size $(t + p) \times (t + p)$, which increases with t , the number of small areas. The specifics of the derivations follow.

2.4.1 Estimation of σ_u^2 and σ_e^2

The method of fitting constants procedure involves performing two OLS regressions on the model (2.3), since there are two unknown variance components. The method of moments is then used to form estimators, one from each regression.

Estimating σ_e^2 involves a regression of $\mathbf{z}$ on $\mathbf{X}_1$ where $\mathbf{X}_1$ is the matrix $\mathbf{W}$ augmented by the matrix $\mathbf{D}$, i.e., $\mathbf{X}_1 = (\mathbf{W}|\mathbf{D})$. This is tantamount to treating both β and $\mathbf{u}$ as fixed effects. We calculate the residual sum of squares for this model, given by $\text{SSE}(\beta, \mathbf{u}) = \mathbf{z}'\mathbf{z} - \mathbf{z}'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^-\mathbf{X}_1'\mathbf{z}$ where $(\mathbf{X}_1'\mathbf{X}_1)^-$ is the generalized inverse of $\mathbf{X}_1'\mathbf{X}_1$, as follows: Note that $\text{SSE}(\beta, \mathbf{u})$ is unique for any choice of basis for $\mathbf{X}_1$. In particular, we can choose $\tilde{\mathbf{X}}_1 = (\mathbf{W}|\mathbf{D})$ if the model (2.3) does not include an intercept term and $\tilde{\mathbf{X}}_1 = (\mathbf{W}_1|\mathbf{D})$ if it does include one, where $\mathbf{W} = (\text{col}_{1 \leq i \leq t}(\mathbf{f}_i^{-1})|\mathbf{W}_1)$. Thus $\text{SSE}(\beta, \mathbf{u}) = \mathbf{z}'\mathbf{z} - \mathbf{z}'\tilde{\mathbf{X}}_1(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1)^{-1}\tilde{\mathbf{X}}_1'\mathbf{z}$, noting that $\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1$ is nonsingular.

If an intercept is not included in the model, then applying the formula for the inverse of a partitioned matrix (see Rao (1965), pp. 33) to

$$\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1 = \begin{pmatrix} \mathbf{W}'\mathbf{W} & \mathbf{W}'\mathbf{D} \\ \mathbf{D}'\mathbf{W} & \mathbf{D}'\mathbf{D} \end{pmatrix}$$

we obtain

$$(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1)^{-1} = \begin{pmatrix} \mathbf{H}^{-1} & -\mathbf{H}^{-1}\mathbf{W}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1} \\ -(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}\mathbf{H}^{-1} & (\mathbf{D}'\mathbf{D})^{-1} + (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}\mathbf{H}^{-1}\mathbf{W}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1} \end{pmatrix}$$

where $\mathbf{H} = \mathbf{W}'\mathbf{W} - \mathbf{W}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}$. Further manipulations give

$$\begin{aligned} \text{SSE}(\beta, \mathbf{u}) &= \mathbf{z}'\mathbf{z} - \mathbf{z}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{z} \\ &\quad - (\mathbf{W}'\mathbf{z} - \mathbf{W}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{z})'\mathbf{H}^{-1}(\mathbf{W}'\mathbf{z} - \mathbf{W}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{z}). \end{aligned}$$

Now, drawing on the fact that

$$\mathbf{z}'\mathbf{z} - \mathbf{z}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{z} = (\mathbf{z} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{z})'(\mathbf{z} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{z}),$$

$$\mathbf{z}'\mathbf{W} - \mathbf{z}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{W} = (\mathbf{z} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{z})'(\mathbf{W} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}),$$

and

$$H = (\mathbf{W} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{W})'(\mathbf{W} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}),$$

we can finally write

$$\text{SSE}(\boldsymbol{\beta}, \mathbf{u}) = \mathbf{z}^{*'}\mathbf{z}^* - \mathbf{z}^{*'}\mathbf{W}^*(\mathbf{W}^{*'}\mathbf{W}^*)^{-1}\mathbf{W}^{*'}\mathbf{z}^*$$

where

$$\mathbf{z}^* = \mathbf{z} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{z}$$

and

$$\mathbf{W}^* = \mathbf{W} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}.$$

If the model includes an intercept term, then we simply replace $\mathbf{W}$ by $\mathbf{W}_1$ in $\mathbf{W}^*$.

Therefore, $\text{SSE}(\boldsymbol{\beta}, \mathbf{u})$ can also be thought of as the residual sum of squares due to the regression of $z_{ij}^* = f_{ij}^{-1}(\cdot - \bar{y}_{iw})$ on $w_{ij}^* = f_{ij}^{-1}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{iw})$; $i = 1, \dots, t$; $j = 1, \dots, n_i$. This is indeed the regression to be employed as a logical extension of the transformation method of estimation of variance components as suggested by Fuller and Battese (1973), for the case $f_{ij} = 1$. Therefore, the two methods produce identical estimators for σ_e^2 .

The unbiased estimator $\tilde{\sigma}_e^2$ for σ_e^2 is given by

$$\tilde{\sigma}_e^2 = (n - p - t + \lambda)^{-1} \text{SSE}(\boldsymbol{\beta}, \mathbf{u}) \quad (2.9)$$

where $\lambda = 1$ for the intercept model and $\lambda = 0$ for the no-intercept model. This follows by noting that

$$E[SSE(\beta, \mathbf{u})] = (n - p - t + \lambda)\sigma_\epsilon^2.$$

Estimating σ_u^2 involves a regression of $\mathbf{z}$ on $\mathbf{W}$. Therefore $SSE(\beta) = \mathbf{z}'\mathbf{z} - \mathbf{z}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{z}$ is the residual sum of squares due to the regression of z_{ij} on $\mathbf{w}_{ij}$.

The unbiased estimator $\hat{\sigma}_u^2$ for σ_u^2 is given by

$$\hat{\sigma}_u^2 = \eta_*^{-1}[SSE(\beta) - (n - p)\hat{\sigma}_\epsilon^2] = \eta_*^{-1}[SSE(\beta) - SSE(\beta, \mathbf{u}) - (t - \lambda)\hat{\sigma}_\epsilon^2] \quad (2.10)$$

where

$$\eta_* = \sum_{i=1}^I \eta_i - \sum_{i=1}^I \eta_i^2 \bar{\mathbf{x}}'_{im} (\mathbf{W}'\mathbf{W})^{-1} \bar{\mathbf{x}}_{im}.$$

This follows by using Lemma 3.5 in Chapter 3, which offers an expression for the expectation of a quadratic form, and by noting that

$$\begin{aligned} E[SSE(\beta)] &= E[\mathbf{z}'(\mathbf{I}_n - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{z}] \\ &= \text{trace}[(\mathbf{I}_n - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{V}] \\ &= \eta_*\sigma_u^2 + (n - p)\sigma_\epsilon^2 \end{aligned}$$

where $\mathbf{V}$ is given by (2.4).

Any unbiased quadratic estimator for σ_u^2 can take on negative values (see LaMotte (1973)). We therefore truncate the estimator by setting it equal to zero whenever $\hat{\sigma}_u^2 < 0$. We define the truncated estimator as follows:

$$\hat{\sigma}_u^2 = \max(0, \hat{\sigma}_u^2).$$

Unfortunately, the truncated estimator is no longer unbiased, but remains consistent as $t \rightarrow \infty$. We define $\hat{\sigma}_\epsilon^2 = \hat{\sigma}_\epsilon^2$ for consistency of notation.

Remark 2.5 The estimators $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$ can be expressed as quadratic forms as follows:

$$\hat{\sigma}_e^2 = \mathbf{z}'\mathbf{C}_1\mathbf{z}$$

where

$$\mathbf{C}_1 = (n - p - t + \lambda)^{-1}[\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'],$$

and

$$\hat{\sigma}_u^2 = \mathbf{z}'\mathbf{C}_2\mathbf{z}$$

where

$$\mathbf{C}_2 = \eta_s^{-1}[\mathbf{I}_n - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' - (n - p)\mathbf{C}_1].$$

These forms are needed in Section 2.5.

2.5 A Second Order Approximation to the MSE of the EBLUP

We begin with a theorem for finding the MSE of the BLUP when σ^2 is known.

Theorem 2.3 Assume a general mixed linear model given by Definition 2.4, where σ^2 is known. Let $t(\mathbf{y})$ be the BLUP of $\mathbf{k}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u}$ found in Theorem 2.1. Then the MSE of $t(\mathbf{y})$ is given by

$$\begin{aligned} \text{MSE}[t(\mathbf{y})] &= E[t(\mathbf{y}) - \mathbf{k}'\boldsymbol{\beta} - \mathbf{m}'\mathbf{u}]^2 \\ &= \mathbf{k}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{k} + \mathbf{m}'\mathbf{G}\mathbf{m} - \mathbf{m}'\mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{G}\mathbf{m} \\ &\quad + \mathbf{m}'\mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{G}\mathbf{m} \\ &\quad - 2\mathbf{k}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{G}\mathbf{m}. \end{aligned}$$

Proof Now

$$\begin{aligned} \text{MSE}[t(\mathbf{y})] &= E[\mathbf{k}'\tilde{\boldsymbol{\beta}} + \mathbf{m}'\mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) - \mathbf{k}'\boldsymbol{\beta} - \mathbf{m}'\mathbf{u}]^2 \\ &= E[\mathbf{k}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z}\mathbf{u} + \mathbf{e}) + \mathbf{m}'\mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{Z}\mathbf{u} + \mathbf{e}) - \mathbf{m}'\mathbf{u}]^2 \end{aligned}$$

where

$$\mathbf{A} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}.$$

This reduces to

$$\begin{aligned} &MSE[t(\mathbf{y})] \\ &= \mathbf{k}'(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{k} \\ &\quad + \mathbf{m}'\mathbf{GZ}'\Sigma^{-1}\mathbf{A}\Sigma\mathbf{A}'\Sigma^{-1}\mathbf{ZGm} + \mathbf{m}'\mathbf{Gm} - 2\mathbf{k}'(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}'\Sigma^{-1}\mathbf{ZGm} \\ &\quad - 2\mathbf{k}'(\mathbf{X}\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{ZGm} - 2\mathbf{m}'\mathbf{GZ}'\Sigma^{-1}\mathbf{AZGm} \\ &= \mathbf{k}'(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{k} - \mathbf{m}'\mathbf{GZ}'\Sigma^{-1}\mathbf{AZGm} + \mathbf{m}'\mathbf{Gm} \\ &\quad - 2\mathbf{k}'(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{ZGm} \end{aligned}$$

by noting that $\mathbf{A}\Sigma\mathbf{A}'\Sigma^{-1} = \mathbf{A}$ and that $\mathbf{AX} = \mathbf{0}$. Substituting the expression for $\mathbf{A}$ gives the desired result. ■

Now applying Theorem 2.3 to model (2.3), we obtain an exact expression for the MSE of $t_i(\sigma^2, \mathbf{y})$ using the explicit expression for $\mathbf{V}^{-1}$ given by (2.7):

$$\begin{aligned} MSE[t_i(\sigma^2, \mathbf{y})] &= \sigma_u^2(1 - \gamma_i) + (\bar{\mathbf{X}}_i - \gamma_i\bar{\mathbf{x}}_{i,w})'(\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}(\bar{\mathbf{X}}_i - \gamma_i\bar{\mathbf{x}}_{i,w}) \\ &= g_{1i}(\sigma^2) + g_{2i}(\sigma^2), \text{ say.} \end{aligned} \quad (2.11)$$

Note that $g_{1i}(\sigma^2)$ can be thought of as the MSE when β is known, and $g_{2i}(\sigma^2)$ as the inflation due to estimating β .

Before proceeding with an approximation to the MSE of the EBLUP of $\bar{Y}_i$, we require the following definitions, borrowed from Bishop, Fienberg and Holland (1975):

Definition 2.5 (a) If $\{a_n\}$ and $\{b_n\}$ are non-stochastic sequences of real numbers, then $a_n = O(b_n)$ if there exists a real number K and an integer $m(K)$ such that if $n > m(K)$ then $|a_n/b_n| < K$. If, instead, $\{a_n\}$ is a sequence of random variables, then we write $a_n = O_p(b_n)$ if the ratio $|a_n/b_n|$ is bounded for large n .

(b) If $\{a_n\}$ and $\{b_n\}$ are non-stochastic sequences of real numbers, then $a_n = o(b_n)$ if for any $\epsilon > 0$ there exists an integer $m(\epsilon)$ such that if $n > m(\epsilon)$ then $|a_n/b_n| < \epsilon$. If, instead, $\{a_n\}$ is a sequence of random variables, then we write $a_n = o_p(b_n)$ if the ratio $|a_n/b_n|$ converges in probability to zero.

Now, when σ^2 is unknown, finding an exact expression for the MSE of $t_i(\hat{\sigma}^2, \mathbf{y})$, the EBLUP of $\bar{Y}_i$, is more difficult, and in most cases intractable. One can naively use $g_{1i}(\sigma^2)|_{\sigma^2=\hat{\sigma}^2} + g_{2i}(\sigma^2)|_{\sigma^2=\hat{\sigma}^2}$ as an estimator of this quantity, but doing so results in serious underestimation of the MSE of the EBLUP, which practitioners are unwilling to accept in general. The underestimation arises due to the fact that this estimator ignores the uncertainty in estimating σ^2 . Now, approximations to the MSE of the EBLUP which do not ignore this uncertainty are possible, correct to order $O(t^{-1})$ for large t , under the assumption of normality of the random errors. We can write

$$\begin{aligned} &MSE[t_i(\hat{\sigma}^2, \mathbf{y})] \\ &= E[t_i(\sigma^2, \mathbf{y}) - \mathbf{k}'\beta - \mathbf{m}'\mathbf{u} + t_i(\hat{\sigma}^2, \mathbf{y}) - t_i(\sigma^2, \mathbf{y})]^2 \\ &= E[t_i(\sigma^2, \mathbf{y}) - \mathbf{k}'\beta - \mathbf{m}'\mathbf{u}]^2 + E[t_i(\hat{\sigma}^2, \mathbf{y}) - t_i(\sigma^2, \mathbf{y})]^2 \end{aligned} \quad (2.12)$$

since Kackar and Harville (1984) show that the crossproduct term is zero under the assumption of normality of the random errors and translation invariance of $\hat{\sigma}^2$ (see definition below).

Definition 2.6 We say that $\hat{\sigma}^2(\mathbf{y})$, a quadratic estimator of σ^2 under the general mixed linear model, is translation invariant iff $\hat{\sigma}^2(\mathbf{y}) = \hat{\sigma}^2(\mathbf{y} - \mathbf{X}\mathbf{a})$ for all constant vectors $\mathbf{a}' = (a_1, \dots, a_p)$.

The first term in (2.12) is simply $\text{MSE}[t_i(\sigma^2, \mathbf{y})]$ which, for model (2.3), is given by (2.11). Kackar and Harville (1984) give an approximation to the second term of (2.12), namely

$$E[t_i(\hat{\sigma}^2, \mathbf{y}) - t_i(\sigma^2, \mathbf{y})]^2 \doteq E[(\hat{\sigma}^2 - \sigma^2)' \mathbf{d}(i)]^2$$

where $\mathbf{d}(i)$ is the vector of first order derivatives of $t_i(\hat{\sigma}^2, \mathbf{y})$ with respect to $\hat{\sigma}^2$, evaluated at $\hat{\sigma}^2 = \sigma^2$. Prasad and Rao (1990) give a further approximation to this term, assuming the following regularity conditions on the general mixed linear model, $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, of which model (2.3) is a special case:

(C1) The elements of $\mathbf{X}$ and $\mathbf{Z}$ are uniformly bounded such that $\mathbf{X}'\Sigma^{-1}\mathbf{X} = O(t)$.

(C2) The elements of $\mathbf{R}_i$ and $\mathbf{G}_i$ are uniformly bounded and differentiable with respect to σ^2 , where $\mathbf{R} = \text{Diag}(\mathbf{R}_i)_{1 \leq i \leq t}$ and $\mathbf{G} = \text{Diag}(\mathbf{G}_i)_{1 \leq i \leq t}$, and $\mathbf{R}_i$ and $\mathbf{G}_i$ are respectively $n_i \times n_i$ and $b_i \times b_i$ matrices. Here $n = \sum_{i=1}^t n_i$ and $b = \sum_{i=1}^t b_i$.

(C3) (i) $\sup_{1 \leq i \leq t} n_i < \infty$ and
(ii) $\sup_{1 \leq i \leq t} b_i < \infty$.

(C4) $\hat{\sigma}_j^2 = \mathbf{y}'\mathbf{C}_j\mathbf{y}$ is a translation invariant unbiased estimator of σ_j^2 for $j = 1, \dots, r$ where r is the number of variance components in the model and $\mathbf{C}_j$ is of the form $\mathbf{C}_j = \text{Diag}[O(t^{-1})]_{n_i \times n_i} + [O(t^{-2})]_{n \times n}$. Here $[O(t^{-s})]_{m \times m}$ denotes an $m \times m$ matrix with element uniformly of order $O(t^{-s})$.

The Prasad-Rao approximation is then given in terms of the general mixed linear model as being

$$E[t_i(\hat{\sigma}^2, \mathbf{y}) - t_i(\sigma^2, \mathbf{y})]^2 = \text{trace}[(\nabla \mathbf{b})\Sigma(\nabla \mathbf{b}')E[(\hat{\sigma}^2 - \sigma^2)(\hat{\sigma}^2 - \sigma^2)']] + o(t^{-1})$$

where $\nabla \mathbf{b}' = \text{col}_{1 \leq j \leq r} (\partial \mathbf{b}' / \partial \sigma_j^2)$ and $\mathbf{b}' = \mathbf{m}'\mathbf{G}\mathbf{Z}'\Sigma^{-1}$.

Now we consider regularity conditions equivalent to (C1) – (C4) on the model (2.3). Conditions (C2) and (C3) (ii) hold since under this model, $b_i = 1$, $R_i = \sigma_e^2 \mathbf{1}_n$, and G_i is the 1×1 matrix consisting of the single element σ_u^2 . Further, a condition equivalent to (C4) holds via Remark 2.5 since for model (2.3), $r = 2$, $\hat{\sigma}_1^2 = \hat{\sigma}_e^2$ and $\hat{\sigma}_2^2 = \hat{\sigma}_u^2$. Hence, we need only assume regularity condition (C3)(i), and a condition equivalent to (C1) given by the following:

(C1)' The elements of W and D are uniformly bounded such that $W'V^{-1}W = O(t)$.

Hence, assuming regularity conditions (C1)' and (C3)(i) on model (2.3), the Prasad-Rao approximation simplifies to

$$\begin{aligned} E[t_i(\hat{\sigma}^2, y) - t_i(\sigma^2, y)]^2 &= \eta_i^{-2}(\sigma_u^2 + \eta_i^{-1}\sigma_e^2)^{-3} \text{var}(\hat{\sigma}_e^2\sigma_u^2 - \hat{\sigma}_u^2\sigma_e^2) \\ &\quad + o(t^{-1}) \\ &= g_{3i}(\sigma^2) + o(t^{-1}). \end{aligned} \quad (2.13)$$

Expressions for $\text{var}(\hat{\sigma}_e^2)$, $\text{var}(\hat{\sigma}_u^2)$ and $\text{cov}(\hat{\sigma}_e^2, \hat{\sigma}_u^2)$ are found assuming normality, by using the fact that the variance component estimators $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$ can be written as quadratic forms (see Remark 2.5), and by using Lemma 2.3 below.

Lemma 2.3 (Searle (1971), Corollary 12, pp. 57) *If y is distributed normally with mean μ and variance-covariance matrix Σ , then $\text{var}(y'Ay) = 2\text{trace}(A\Sigma A\Sigma) + 4\mu'A\Sigma A\mu$, where A is an $n \times n$ matrix.*

The resultant expressions are

$$\text{var}(\hat{\sigma}_e^2) = 2(n - p - t + \lambda)^{-1}\sigma_e^4, \quad (2.14)$$

$$\begin{aligned} \text{var}(\hat{\sigma}_u^2) &= \eta_*^{-2}[2(t - \lambda)(n - p)(n - p - t + \lambda)^{-1}\sigma_e^4 \\ &\quad + 2\eta_{**}\sigma_u^4 + 4\eta_*\sigma_e^2\sigma_u^2] \end{aligned} \quad (2.15)$$

and

$$\text{cov}(\hat{\sigma}_e^2, \hat{\sigma}_u^2) = -\eta_*^{-1}(t - \lambda) \text{var}(\hat{\sigma}_e^2) \quad (2.16)$$

where η_* is as in (2.10) and

$$\begin{aligned} \eta_{**} = & \sum_{i=1}^I \eta_i^2 - 2 \sum_{i=1}^I \eta_i^3 \bar{\mathbf{x}}'_{iwr} (\mathbf{W}'\mathbf{W})^{-1} \mathbf{x}_{iwr} \\ & + \text{trace} [(\mathbf{W}'\mathbf{W})^{-1} \sum_{i=1}^I \eta_i^2 \bar{\mathbf{x}}_{iwr} \mathbf{x}'_{iwr}]^2. \end{aligned}$$

Assembling the results from (2.11) to (2.16), we finally obtain an expression for the second order approximation.

Theorem 2.4 *For the nested error regression model defined by (2.3), assume normality of the random errors as well as regularity conditions (C1)' and (C3)(i). Then, the MSE of the EBLUP, $t_i(\hat{\sigma}^2, \mathbf{y})$, can be approximated by*

$$\begin{aligned} \text{MSE}[t_i(\hat{\sigma}^2, \mathbf{y})] &= \sigma_u^2(1 - \gamma_i) + (\bar{\mathbf{X}}_i - \gamma_i \bar{\mathbf{x}}_{iwr})' (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1} (\bar{\mathbf{X}}_i - \gamma_i \mathbf{x}_{iwr}) \\ &\quad + \eta_i^{-2} (\sigma_u^2 + \eta_i^{-1} \sigma_e^2)^{-3} \text{var}(\hat{\sigma}_e^2 \sigma_u^2 - \hat{\sigma}_u^2 \sigma_e^2) + o(t^{-1}) \\ &= g_{1i}(\sigma^2) + g_{2i}(\sigma^2) + g_{3i}(\sigma^2) + o(t^{-1}) \end{aligned} \quad (2.17)$$

where $\text{var}(\hat{\sigma}_e^2)$, $\text{var}(\hat{\sigma}_u^2)$ and $\text{cov}(\hat{\sigma}_e^2, \hat{\sigma}_u^2)$ are defined by (2.14)–(2.16).

Proof The proof that the neglected terms are of order $o(t^{-1})$ is omitted here since a similar proof is offered for the more complex two-fold nested error model, in Chapter 3.

Having established a second order approximation for the case of large N_i , it is easy to do likewise for the case of finite N_i by making some simple observations. First, the BLUP $t_i^F(\sigma^2, \mathbf{y})$, given in (2.8), can be written as

$$t_i^F(\sigma^2, \mathbf{y}) = N_i^{-1} n_i y_i + N_i^{-1} (N_i - n_i) t_i^*(\sigma^2, \mathbf{y})$$

where

$$t_i^*(\sigma^2, \mathbf{y}) = \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}} + \gamma_i (y_{iwr} - \bar{\mathbf{x}}'_{iwr} \hat{\boldsymbol{\beta}}).$$

Also $\bar{Y}_i = N_i^{-1}n_i\bar{y}_i + N_i^{-1}(N_i - n_i)\bar{y}_i^*$ where $\bar{y}_i^* = (N_i - n_i)^{-1} \sum_{j \in I_i} y_{ij}$. Combining these results, we have

$$t_i^F(\hat{\sigma}^2, \mathbf{y}) - \bar{Y}_i = N_i^{-1}(N_i - n_i)[t_i^*(\hat{\sigma}^2, \mathbf{y}) - (N_i - n_i)^{-1} \sum_{j \in I_i} \mathbf{x}_{ij}'\boldsymbol{\beta} \\ u_i - (N_i - n_i)^{-1} \sum_{j \in I_i} f_{ij}e_{ij}]$$

and finally

$$\begin{aligned} MSE[t_i^F(\hat{\sigma}^2, \mathbf{y})] &= E[t_i^F(\hat{\sigma}^2, \mathbf{y}) - \bar{Y}_i]^2 \\ &= N_i^{-2}(N_i - n_i)^2 MSE[t_i^*(\hat{\sigma}^2, \mathbf{y})] + \sigma_e^2 N_i^{-2} \sum_{j \in I_i} f_{ij}^2. \end{aligned} \quad (2.18)$$

Therefore, no new theory is required here. A second order approximation to $MSE[t_i^F(\hat{\sigma}^2, \mathbf{y})]$ can be found from (2.18) by replacing $MSE[t_i^*(\hat{\sigma}^2, \mathbf{y})]$ with the second order approximation to $MSE[t_i(\hat{\sigma}^2, \mathbf{y})]$ given in Theorem 2.4, and substituting $\bar{\mathbf{X}}_i$ with $\mathbf{x}_i^*$. Thus

$$\begin{aligned} MSE[t_i^F(\hat{\sigma}^2, \mathbf{y})] &= N_i^{-2}(N_i - n_i)^2 [\sigma_u^2(1 - \gamma_i) \\ &\quad + (\bar{\mathbf{x}}_i^* - \gamma_i \bar{\mathbf{x}}_{i\mathbf{w}})'(\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}(\bar{\mathbf{x}}_i^* - \gamma_i \bar{\mathbf{x}}_{i\mathbf{w}}) \\ &\quad + \eta_i^{-2}(\sigma_u^2 + \eta_i^{-1}\sigma_e^2)^{-3} \text{var}(\hat{\sigma}_e^2\sigma_u^2 - \hat{\sigma}_u^2\sigma_e^2)] \\ &\quad + \sigma_e^2 N_i^{-2} \sum_{j \in I_i} f_{ij}^2 + o(t^{-1}) \\ &= N_i^{-2}(N_i - n_i)^2 [g_{1i}(\sigma^2) + g_{2i}^*(\sigma^2) + g_{3i}(\sigma^2)] \\ &\quad + g_{4i}(\sigma^2) + o(t^{-1}) \end{aligned} \quad (2.19)$$

where

$$\begin{aligned} g_{2i}^*(\sigma^2) &= (\bar{\mathbf{x}}_i^* - \gamma_i \bar{\mathbf{x}}_{i\mathbf{w}})'(\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}(\bar{\mathbf{x}}_i^* - \gamma_i \bar{\mathbf{x}}_{i\mathbf{w}}), \\ g_{4i}(\sigma^2) &= \sigma_e^2 N_i^{-2} \sum_{j \in I_i} f_{ij}^2 \end{aligned}$$

and $g_{1i}(\sigma^2)$ and $g_{3i}(\sigma^2)$ are as before.

For the special case $f_{ij} = 1$, (2.17) and (2.19) reduce to the formulae derived by Prasad and Rao (1990).

2.6 Estimation of the MSE of the EBLUP

Having obtained the second order approximation to the MSE of the EBLUP in Theorem 2.4, it is also useful to obtain an estimator of the approximation that depends on $\hat{\sigma}^2$, yet has a bias of order $o(t^{-1})$, i.e.,

$$E[\text{mse}(t_i(\hat{\sigma}^2, y)) - \text{MSE}(t_i(\hat{\sigma}^2, y))] = o(t^{-1})$$

where $\text{mse}[t_i(\hat{\sigma}^2, y)]$ is the estimator in question.

Since $g_{2i}(\sigma^2)$ and $g_{3i}(\sigma^2)$ are of order $O(t^{-1})$, it follows that

$$E[g_{2i}(\hat{\sigma}^2)] = g_{2i}(\sigma^2) + o(t^{-1})$$

and

$$E[g_{3i}(\hat{\sigma}^2)] = g_{3i}(\sigma^2) + o(t^{-1})$$

where

$$g_{2i}(\hat{\sigma}^2) = g_{2i}(\sigma^2)|_{\sigma^2=\hat{\sigma}^2}$$

and

$$g_{3i}(\hat{\sigma}^2) = g_{3i}(\sigma^2)|_{\sigma^2=\hat{\sigma}^2}.$$

Therefore $g_{2i}(\hat{\sigma}^2)$ and $g_{3i}(\hat{\sigma}^2)$ are the correct estimators of $g_{2i}(\sigma^2)$ and $g_{3i}(\sigma^2)$ respectively, each with bias $o(t^{-1})$. Now $g_{1i}(\sigma^2)$ is of order $O(1)$, so $g_{1i}(\hat{\sigma}^2) = g_{1i}(\sigma^2)|_{\sigma^2=\hat{\sigma}^2}$ is not the correct estimator of $g_{1i}(\sigma^2)$, since the bias is of order $O(t^{-1})$. Thus, the bias in $g_{1i}(\hat{\sigma}^2)$ must be estimated to the correct order. We will now show that $E[g_{1i}(\hat{\sigma}^2) + g_{3i}(\hat{\sigma}^2)] = g_{1i}(\sigma^2) + o(t^{-1})$.

The Taylor expansion of $g_{1i}(\hat{\sigma}^2)$ about the point σ^2 yields

$$g_{1i}(\hat{\sigma}^2) = g_{1i}(\sigma^2) + (\hat{\sigma}^2 - \sigma^2)' \nabla g_{1i}(\sigma^2) + 2^{-1}(\hat{\sigma}^2 - \sigma^2)' \nabla^2 g_{1i}(\sigma^2)(\hat{\sigma}^2 - \sigma^2) + O_p(t^{-3/2})$$

where $\nabla g_{1i}(\sigma^2)$ and $\nabla^2 g_{1i}(\sigma^2)$ are respectively the 2×1 vector of first order derivatives and the 2×2 matrix of second order derivatives of $g_{1i}(\sigma^2)$ with respect to σ^2 . Now applying the expectation operator to the above expression yields

$$E[g_{1i}(\hat{\sigma}^2)] = g_{1i}(\sigma^2) + E[(\hat{\sigma}^2 - \sigma^2)' \nabla g_{1i}(\sigma^2)] + 2^{-1} E[(\hat{\sigma}^2 - \sigma^2)' \nabla^2 g_{1i}(\sigma^2)(\hat{\sigma}^2 - \sigma^2)] + O(t^{-2}),$$

noting that $E[O_p(t^{-3/2})] = O(t^{-2})$. A rigorous proof of the latter assertion is omitted here, but follows along the lines of the derivations in Chapter 3. Now, in Chapter 3, for the more complex two-fold nested error regression model, we show that

$$E[(\hat{\sigma}^2 - \sigma^2)' \nabla g_{1i}(\sigma^2)] = o(t^{-1})$$

and that

$$E[(\hat{\sigma}^2 - \sigma^2)' \nabla^2 g_{1i}(\sigma^2)(\hat{\sigma}^2 - \sigma^2)] = E[(\hat{\sigma}^2 - \sigma^2)' \nabla^2 g_{1i}(\sigma^2)(\hat{\sigma}^2 - \sigma^2)] + o(t^{-1}).$$

Similar proofs apply to the two assertions, for the simpler model at hand. Thus,

$$E[g_{1i}(\hat{\sigma}^2)] = g_{1i}(\sigma^2) + 2^{-1} E[(\hat{\sigma}^2 - \sigma^2)' \nabla^2 g_{1i}(\sigma^2)(\hat{\sigma}^2 - \sigma^2)] + o(t^{-1}).$$

Applying Lemma 3.5 from Chapter 3, which offers an expression for the expectation of a quadratic form, yields

$$\begin{aligned} E[g_{1i}(\hat{\sigma}^2)] &= g_{1i}(\sigma^2) + 2^{-1} \text{trace}[\text{var}(\hat{\sigma}^2) \nabla^2 g_{1i}(\sigma^2)] + o(t^{-1}) \\ &= g_{1i}(\sigma^2) - \eta_i^{-2}(\sigma_u^2 + \eta_i^{-1} \sigma_e^2)^{-3} [\sigma_e^4 \text{var}(\hat{\sigma}_u^2) \\ &\quad + \sigma_u^4 \text{var}(\hat{\sigma}_e^2) - 2\sigma_e^2 \sigma_u^2 \text{cov}(\hat{\sigma}_e^2, \hat{\sigma}_u^2)] + o(t^{-1}) \\ &= g_{1i}(\sigma^2) - g_{3i}(\sigma^2) + o(t^{-1}) \end{aligned}$$

since

$$\nabla^2 g_{1i}(\sigma^2) = -2\eta_i^{-2}(\sigma_u^2 + \eta_i^{-1}\sigma_e^2)^{-3} \begin{pmatrix} \sigma_u^4 & -\sigma_e^2\sigma_u^2 \\ -\sigma_e^2\sigma_u^2 & \sigma_e^4 \end{pmatrix}$$

and

$$\text{var}(\tilde{\sigma}^2) = \begin{pmatrix} \text{var}(\tilde{\sigma}_e^2) & \text{cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_u^2) \\ \text{cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_u^2) & \text{var}(\tilde{\sigma}_u^2) \end{pmatrix}$$

where the elements of the matrix $\text{var}(\tilde{\sigma}^2)$ are defined in (2.14)–(2.16) assuming normality of the random errors. Thus $g_{1i}(\hat{\sigma}^2) + g_{3i}(\hat{\sigma}^2)$ is the correct estimator of $g_{1i}(\sigma^2)$, having bias $o(t^{-1})$.

Combining the above results, we have that

$$\text{mse}[t_i(\hat{\sigma}^2, \mathbf{y})] = g_{1i}(\hat{\sigma}^2) + g_{2i}(\hat{\sigma}^2) + 2g_{3i}(\hat{\sigma}^2) \quad (2.20)$$

is an estimator of $\text{MSE}[t_i(\hat{\sigma}^2, \mathbf{y})]$ with bias of order $o(t^{-1})$.

In addition, since $g_{4i}(\hat{\sigma}^2) = g_{4i}(\sigma^2)|_{\sigma^2=\hat{\sigma}^2}$, is unbiased for $g_{4i}(\sigma^2)$, finding a similar estimator of (2.19) is immediate. It is denoted by $\text{mse}[t_i^F(\hat{\sigma}^2, \mathbf{y})]$ and can be seen to be

$$\begin{aligned} \text{mse}[t_i^F(\hat{\sigma}^2, \mathbf{y})] &= N_i^{-2}(N_i - n_i)^2[g_{1i}(\hat{\sigma}^2) + g_{2i}^*(\hat{\sigma}^2) + 2g_{3i}(\hat{\sigma}^2)] \\ &\quad + g_{4i}(\hat{\sigma}^2) \end{aligned} \quad (2.21)$$

where

$$g_{2i}^*(\hat{\sigma}^2) = g_{2i}^*(\sigma^2)|_{\sigma^2=\hat{\sigma}^2}.$$

For the special case $f_{ij} = 1$, (2.20) and (2.21) reduce to the expressions obtained by Prasad and Rao (1990).

CHAPTER 3

The Two-Fold Nested Error Regression Model With Unequal Error Variances

In this chapter, the two-fold nested error regression model with unequal error variances is considered. A detailed description of the model is given in Section 3.1. In Section 3.2, a transformation is found such that a special case of the above model is reduced to one with constant, uncorrelated error variances, to enable the estimation of the regression parameters involved, using OLS, rather than GLS. Best linear unbiased predictors of the small area means for the infinite and finite population cases, under the assumption of known variance components, are found in Section 3.3. For the finite case, a proof that the expression given is in fact the BLUP is offered. Section 3.4 deals with the estimation of the variance components using Henderson's method. Just as before, these estimators are substituted back into the expressions for the BLUPs found in Section 3.3, to get the EBLUPs. Section 3.5 gives a second order approximation to the MSE of the EBLUP, as well as a rigorous proof, for a special case, that the remainder terms are of order $o(t^{-1})$ for large t , the number of small areas. Included in this section are expressions for the variances of the variance component estimates as well as the covariances between them, under the assumption of normality of the random errors. Section 3.6 gives the derivation of an estimator of the MSE of the EBLUP, having bias of order $o(t^{-1})$. Finally, Section 3.7 reports the results of a small Monte Carlo study using the balanced, two-fold nested error mean model with equal error variances. The accuracy of the second order approximation to the MSE of the EBLUP and the bias of the estimators of the MSE are investigated in this study.

3.1 Description of the Model

Chapter 2 dealt with the one-fold nested error regression model, with unequal error variances. A natural extension of this model is one where more levels of error nesting are introduced. Such is the case for the two-fold nested error regression model. Fuller and Battese (1973) deal with this model, specializing to the partially balanced case, with constant error variances. This chapter focuses on the Fuller and Battese model, generalizing to the unbalanced case with unequal error variances. A representation of the generalized model is given by

$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + \xi_i + u_{ij} + f_{ijk}e_{ijk}; \quad i = 1, \dots, t; \quad j = 1, \dots, m_i; \quad k = 1, \dots, n_{ij}. \quad (3.1)$$

This model can be interpreted in terms of a two-stage sampling design, having t small areas. First one samples m_i primary sampling units (PSUs) from M_i population PSUs within the i^{th} small area. One then subsamples n_{ij} secondary sampling units (SSUs) from N_{ij} population SSUs in the j^{th} PSU within the i^{th} small area. We have that y_{ijk} is the value of the response variable for the k^{th} SSU in the j^{th} PSU within the i^{th} small area, and $\mathbf{x}'_{ijk} = (x_{ijk1}, \dots, x_{ijkp})$ is a $1 \times p$ vector of known auxiliary variables, while $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a $p \times 1$ vector of unknown regression parameters to be determined. The additive random errors ξ_i ; $i = 1, \dots, t$, u_{ij} ; $i = 1, \dots, t$; $j = 1, \dots, m_i$ and e_{ijk} ; $i = 1, \dots, t$; $j = 1, \dots, m_i$; $k = 1, \dots, n_{ij}$ are assumed to be independent of each other, and each in turn is assumed independent with mean zero and respective variances σ_ξ^2 , σ_u^2 and σ_e^2 . Finally, the f_{ijk} 's are known, positive constants. For instance, f_{ijk} can be a known, bounded function of the auxiliary variables $\mathbf{x}_{ijk}$, i.e., $f_{ijk} = g(\mathbf{x}_{ijk})$. Analysis of diagnostics may help to make an appropriate choice for f_{ijk} , given the data at hand.

The question of diagnostics for random effects models, such as (3.1), is only currently being investigated, however. Lange and Ryan (1989) consider a goodness-of-fit test for the one-fold model. Calvin and Sedransk (1991) used standardized BLUP residual plots to check for constancy of error variances under the two-fold nested error mean model. They also use weighted normal probability plots, as proposed by Dempster and Ryan (1985), to study the assumption of normality of the random errors.

As an application of the two-fold nested error regression model, consider the recent problem of predicting the average crop yield in a specified tehal, or group of villages, in India. For this application, tehails can be considered as small areas, villages within tehails can be considered as PSUs, and fields within villages within tehails can be considered as SSUs. Auxiliary information is also available at the PSU level in the form of total village acreage.

Remark 3.1 We can take a similar approach to Chapter 2 in that we transform model (3.1) to one having constant error variances by dividing the existing model through by f_{ijk} and obtaining

$$\begin{aligned} z_{ijk} &= \mathbf{w}'_{ijk}\beta + f_{ijk}^{-1}\xi_i + f_{ijk}^{-1}u_{ij} + e_{ijk}; \quad i = 1, \dots, t; \\ j &= 1, \dots, m_i; \quad k = 1, \dots, n_{ij} \end{aligned} \quad (3.2)$$

where $z_{ijk} = f_{ijk}^{-1}y_{ijk}$ and $\mathbf{w}'_{ijk} = (w_{ijk1}, \dots, w_{ijkp}) = (f_{ijk}^{-1}x_{ijk1}, \dots, f_{ijk}^{-1}x_{ijkp})$.

The covariance structure for this model can be written as

$$\text{cov}(z_{ijk}, z_{i'j'k'}) = \begin{cases} f_{ijk}^{-2}\sigma_{\xi}^2 + f_{ijk}^{-2}\sigma_u^2 + \sigma_e^2 & \text{if } i = i', j = j' \text{ and } k = k' \\ f_{ijk}^{-1}f_{ijk'}^{-1}\sigma_{\xi}^2 + f_{ijk}^{-1}f_{ijk'}^{-1}\sigma_u^2 & \text{if } i = i', j = j' \text{ and } k \neq k' \\ f_{ijk}^{-1}f_{ij'k'}^{-1}\sigma_{\xi}^2 & \text{if } i = i' \text{ and } j \neq j' \\ 0 & \text{if } i \neq i'. \end{cases}$$

Remark 3.2 It is possible to express model (3.2) as a mixed linear model in matrix form, just as was done for the one-fold model. Thus,

$$\mathbf{z} = \mathbf{W}\boldsymbol{\beta} + \mathbf{D}\mathbf{u} + \mathbf{e} \quad (3.3)$$

where $\mathbf{z} = \begin{smallmatrix} \text{col} \\ 1 \leq i \leq t \end{smallmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq j \leq m_i \end{smallmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq k \leq n_{ij} \end{smallmatrix} (z_{ijk})$, and $\mathbf{W} = \begin{smallmatrix} \text{col} \\ 1 \leq i \leq t \end{smallmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq j \leq m_i \end{smallmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq k \leq n_{ij} \end{smallmatrix} (\mathbf{w}'_{ijk})$ are obtained by arranging the z_{ijk} 's and $\mathbf{w}_{ijk}$'s in lexicographical order. Further,

$$\mathbf{D} = (\text{Diag} \begin{smallmatrix} (\mathbf{f}_i^{-1}) \\ 1 \leq i \leq t \end{smallmatrix} | \text{Diag} \begin{smallmatrix} \text{Diag} \begin{smallmatrix} (\mathbf{f}_{ij}^{-1}) \\ 1 \leq j \leq m_i \end{smallmatrix} \end{smallmatrix}) = (\mathbf{D}_1 | \mathbf{D}_2)$$

say, where

$$\mathbf{f}_i^{-1} = \begin{smallmatrix} \text{col} \\ 1 \leq j \leq m_i \end{smallmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq k \leq n_{ij} \end{smallmatrix} (f_{ijk}^{-1})$$

and

$$\mathbf{f}_{ij}^{-1} = \begin{smallmatrix} \text{col} \\ 1 \leq k \leq n_{ij} \end{smallmatrix} (f_{ijk}^{-1}).$$

Finally

$$\mathbf{u} = \begin{pmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq i \leq t \end{smallmatrix} (\xi_i) \\ \begin{smallmatrix} \text{col} \\ 1 \leq i \leq t \end{smallmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq j \leq m_i \end{smallmatrix} (u_{ij}) \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix},$$

say, and $\mathbf{e} = \begin{smallmatrix} \text{col} \\ 1 \leq i \leq t \end{smallmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq j \leq m_i \end{smallmatrix} \begin{smallmatrix} \text{col} \\ 1 \leq k \leq n_{ij} \end{smallmatrix} (e_{ijk})$. The vectors $\mathbf{u}$ and $\mathbf{e}$ are independent with mean $\mathbf{0}$ and joint variance-covariance matrix

$$\text{var} \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \sigma_\xi^2 \mathbf{I}_t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \mathbf{I}_{\Sigma m_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_e^2 \mathbf{I}_n \end{pmatrix} = \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix},$$

say, where $\mathbf{I}_t$, $\mathbf{I}_{\Sigma m_i}$ and $\mathbf{I}_n$ are identity matrices of order t , $\sum_{i=1}^t m_i$ and

$n = \sum_{i=1}^t \sum_{j=1}^{m_i} n_{ij}$ respectively, and

$$\mathbf{G} = \begin{pmatrix} \sigma_\xi^2 \mathbf{I}_t & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \mathbf{I}_{\Sigma m_i} \end{pmatrix}$$

while $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$. The variance-covariance matrix of $\mathbf{z}$ is given by

$$\mathbf{V} = \text{Diag} (\mathbf{V}_i)_{1 \leq i \leq t} \quad (3.4)$$

where

$$\mathbf{V}_i = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_u^2 \text{Diag} (\mathbf{f}_{ij}^{-1})(\mathbf{f}_{ij}^{-1})' + \sigma_\xi^2 (\mathbf{f}_i^{-1})(\mathbf{f}_i^{-1})'$$

and $\mathbf{I}_{n_i}$ represents an identity matrix of order $n_i = \sum_{j=1}^{m_i} n_{ij}$.

3.2 Transformed Model with Uncorrelated Errors and Constant Variances

Once again, we are interested in transforming the model so that the estimation of β is simplified in the sense that no unwieldy matrix inversions are required in the process. To achieve this end, we mimic the method described in Section 2.2. Unfortunately, application of this method to model (3.1), or equivalently to model (3.2), results in eigenvalues and eigenvectors which are extremely cumbersome. Therefore, we confine ourselves to the following specialized version of the model:

$$y_{ijk} = \mathbf{x}_{ijk}' \beta + \xi_i + u_{ij} + f_i c_{ijk}; \quad i = 1, \dots, t; \quad j = 1, \dots, m_i; \quad k = 1, \dots, n_i$$

or equivalently

$$z_{ijk} = \mathbf{w}_{ijk}' \beta + f_i^{-1} \xi_i + f_i^{-1} u_{ij} + c_{ijk}; \quad i = 1, \dots, t; \quad j = 1, \dots, m_i; \quad k = 1, \dots, n_i.$$

Note that this model is restricted to the case where f_{ijk} and n_{ij} depend only on i so that $f_{ijk} = f_i$ and $n_{ij} = n_i$. Fuller and Battese (1973) consider a special case of this model with $f_i = 1$ for all i .

The variance-covariance matrix, $\mathbf{V}$, for the restricted model, stated in terms of the z_{ijk} 's, is given by $\mathbf{V} = \text{Diag} (\mathbf{V}_i)_{1 \leq i \leq t}$ with

$$\mathbf{V}_i = \sigma_e^2 \mathbf{I}_{n_i m_i} + \sigma_u^2 \eta_i (\mathbf{1}_{n_i} \mathbf{1}_{n_i}' \otimes \mathbf{I}_{m_i}) + \sigma_\xi^2 \eta_i \mathbf{1}_{n_i m_i} \mathbf{1}_{n_i m_i}'$$

where $\mathbf{1}_n$, and $\mathbf{1}_{n,m_i}$ are respectively $n_i \times 1$ and $n_i m_i \times 1$ vectors consisting entirely of ones, $\mathbf{I}_{n_i m_i}$ and $\mathbf{I}_{m_i}$ are identity matrices with orders $n_i m_i$ and m_i respectively and $\eta_i = f_i^{-2}$. Note that $\otimes$ represents the Kronecker product operation on matrices.

One can easily verify that the eigenvalues of $\mathbf{V}_i$ are $\lambda_{1i} = \sigma_e^2$, $\lambda_{2i} = \sigma_e^2 + \eta_i n_i \sigma_u^2$ and $\lambda_{3i} = \sigma_e^2 + \eta_i n_i \sigma_u^2 + \eta_i n_i m_i \sigma_\xi^2$ with corresponding mutually orthogonal, symmetric, and idempotent eigenvector matrices

$$A_{1i} = \mathbf{I}_{n_i m_i} - n_i^{-1} (\mathbf{1}_n \mathbf{1}'_{n_i} \otimes \mathbf{I}_{m_i}),$$

$$A_{2i} = n_i^{-1} (\mathbf{1}_n \mathbf{1}'_{n_i} \otimes \mathbf{I}_{m_i}) - (n_i m_i)^{-1} \mathbf{1}_{n,m_i} \mathbf{1}'_{n,m_i},$$

and

$$A_{3i} = (n_i m_i)^{-1} \mathbf{1}_{n,m_i} \mathbf{1}'_{n,m_i}.$$

We form the required transformed model as follows: Set

$$\begin{aligned} \mathbf{T}_i^* &= \sum_{k=1}^3 \lambda_{ki}^{-1/2} A_{ki} \\ &= \sigma_e^{-1} (\mathbf{I}_{n_i m_i} - n_i^{-1} (\mathbf{1}_n \mathbf{1}'_{n_i} \otimes \mathbf{I}_{m_i})) \\ &\quad + (\sigma_e^2 + \eta_i n_i \sigma_u^2)^{-1/2} (n_i^{-1} (\mathbf{1}_n \mathbf{1}'_{n_i} \otimes \mathbf{I}_{m_i}) - (n_i m_i)^{-1} \mathbf{1}_{n,m_i} \mathbf{1}'_{n,m_i}) \\ &\quad + (\sigma_e^2 + \eta_i n_i \sigma_u^2 + \eta_i n_i m_i \sigma_\xi^2)^{-1/2} (n_i m_i)^{-1} \mathbf{1}_{n,m_i} \mathbf{1}'_{n,m_i}. \end{aligned}$$

Next, set $\mathbf{T}_i = \sigma_e \mathbf{T}_i^*$ so that the transformed model has constant variances equal to σ_e^2 . Thus

$$\mathbf{T}_i = \mathbf{I}_{n_i m_i} - \rho_{1i} n_i^{-1} (\mathbf{1}_n \mathbf{1}'_{n_i} \otimes \mathbf{I}_{m_i}) - \rho_{2i} (n_i m_i)^{-1} \mathbf{1}_{n,m_i} \mathbf{1}'_{n,m_i},$$

where

$$\rho_{1i} = 1 - (\sigma_e^2 (\sigma_e^2 + \eta_i n_i \sigma_u^2)^{-1})^{1/2}$$

and

$$\rho_{2i} = 1 - \rho_{1i} - (\sigma_e^2(\sigma_e^2 + \eta_i n_i \sigma_u^2 + \eta_i n_i m_i \sigma_\xi^2)^{-1})^{1/2}.$$

Finally, premultiply the model by the transformation matrix $\mathbf{T} = \text{Diag}(\mathbf{T}_i)$ to obtain the transformed model. Upon doing so, we find the $(i, j, k)^{\text{th}}$ element of the resultant model to be

$$z_{ijk} - \rho_{1i} z_{ij.} - \rho_{2i} z_{i..} = (\mathbf{w}_{ijk} - \rho_{1i} \mathbf{w}_{ij.} - \rho_{2i} \mathbf{w}_{i..})' \boldsymbol{\beta} + \psi_{ijk}^* \quad (3.5)$$

where

$$\begin{aligned} z_{ij.} &= n_i^{-1} \sum_{k=1}^{n_i} z_{ijk}, \\ z_{i..} &= (n_i m_i)^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_i} z_{ijk}, \\ \mathbf{w}_{ij.} &= n_i^{-1} \sum_{k=1}^{n_i} \mathbf{w}_{ijk}, \\ \mathbf{w}_{i..} &= (n_i m_i)^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_i} \mathbf{w}_{ijk} \end{aligned}$$

and ψ_{ijk}^* is the $(i, j, k)^{\text{th}}$ element of the $n \times 1$ vector $\boldsymbol{\psi}$ defined in Section 2.2, where for this model $n = \sum_{i=1}^I n_i m_i$. Note that the ψ_{ijk}^* 's are uncorrelated with constant variances. Thus, the OLS estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ under (3.5) is identical to the GLS estimator of $\boldsymbol{\beta}$ under the untransformed model.

3.3 Best Linear Unbiased Prediction Under the Assumption of Known Variance Components

We wish to predict the population mean of the i^{th} small area under model (3.1). Equivalently, we wish to predict $\bar{Y}_{i..} = \bar{\mathbf{X}}_{i..}' \boldsymbol{\beta} + \xi_i + \bar{U}_{i.} + \bar{E}_{i..}$ where $\bar{Y}_{i.} = N_i^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk}$ and $\bar{\mathbf{X}}_{i..} = N_i^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} \mathbf{x}_{ijk}$ are the means of the y_{ijk} 's and the $\mathbf{x}_{ijk}$'s respectively, for the $N_i = \sum_{j=1}^{M_i} N_{ij}$ population units in the i^{th} small area, and $\bar{U}_{i.} = N_i^{-1} \sum_{j=1}^{M_i} N_{ij} u_{ij}$ while $\bar{E}_{i..} = N_i^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} f_{ijk} e_{ijk}$.

3.3.1 The Infinite Population Case

We assume here that M_i is large or that the sampling fraction $m_i M_i^{-1}$ is negligible, and that f_{ijk} is bounded for $i = 1, \dots, t$; $j = 1, \dots, M_i$; $k = 1, \dots, N_{ij}$. In this case, $\bar{U}_{i..}$ and $\bar{E}_{i..}$ can be taken to be approximately zero by applying the Weak Law of Large Numbers for independent random variables with bounded variances (see Rao (1965), pp. 112). Therefore, $\bar{Y}_{i..} \doteq \bar{\mathbf{X}}_{i..}' \boldsymbol{\beta} + \xi_i$. The BLUP of $\bar{Y}_{i..}$ is found by invoking Theorem 2.1 with $\mathbf{k} = \bar{\mathbf{X}}_{i..}$ and

$$\mathbf{m} = \begin{pmatrix} \text{col}_{1 \leq t \leq t} (\delta_{it}) \\ \mathbf{0}_{\Sigma m_i} \end{pmatrix}$$

where $\mathbf{0}_{\Sigma m_i}$ is a $\sum_{i=1}^t m_i \times 1$ vector consisting entirely of zeros and δ_{it} is the indicator defined in Chapter 2. One then obtains for the BLUP of $\bar{Y}_{i..}$

$$\begin{aligned} t_i(\sigma^2, \mathbf{y}) &= \gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \sum_{j=1}^{m_i} \gamma_{ij} [\bar{y}_{ijw} + (\bar{\mathbf{X}}_{i..} - \bar{\mathbf{x}}_{ijw})' \tilde{\boldsymbol{\beta}}] + (1 - \gamma_i) \bar{\mathbf{X}}_{i..}' \tilde{\boldsymbol{\beta}} \\ &= \bar{\mathbf{X}}_{i..}' \tilde{\boldsymbol{\beta}} + \gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \sum_{j=1}^{m_i} \gamma_{ij} (y_{ijw} - \mathbf{x}_{ijw}' \tilde{\boldsymbol{\beta}}) \\ &= \bar{\mathbf{X}}_{i..}' \tilde{\boldsymbol{\beta}} + \tilde{\xi}_i \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} \gamma_{ij} &= \sigma_u^2 (\sigma_u^2 + \sigma_e^2 \eta_{ij}^{-1})^{-1}, \\ \gamma_i &= \sigma_\xi^2 (\sigma_\xi^2 + \sigma_u^2 (\sum_{j=1}^{m_i} \gamma_{ij})^{-1})^{-1}, \\ \eta_{ij} &= \sum_{k=1}^{n_{ij}} f_{ijk}^{-2} = (\mathbf{f}_{ij}^{-1})' (\mathbf{f}_{ij}^{-1}), \\ \bar{y}_{ijw} &= \eta_{ij}^{-1} \sum_{k=1}^{n_{ij}} f_{ijk}^{-1} z_{ijk}, \\ \bar{\mathbf{x}}_{ijw} &= \eta_{ij}^{-1} \sum_{k=1}^{n_{ij}} f_{ijk}^{-1} \mathbf{w}_{ijk}, \\ \tilde{\xi}_i &= \gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \sum_{j=1}^{m_i} \gamma_{ij} (y_{ijw} - \mathbf{x}_{ijw}' \tilde{\boldsymbol{\beta}}) \end{aligned}$$

and

$$\hat{\beta} = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}^{-1}\mathbf{z}.$$

Remark 3.3 The proof of (3.6) involves the inversion of the variance-covariance matrix $\mathbf{V} = \text{Diag}(\mathbf{V}_i)$, given by (3.4). We employ Lemma 3.1, which follows, to facilitate us with the inversion of $\mathbf{V}_i$, and hence $\mathbf{V}$. Now one can write $\mathbf{V}_i = \boldsymbol{\Sigma}_i + \sigma_\epsilon^2(\mathbf{f}_i^{-1})(\mathbf{f}_i^{-1})'$ where $\boldsymbol{\Sigma}_i = \text{Diag}_{1 \leq j \leq m_i}(\sigma_\epsilon^2 \mathbf{I}_{n_{ij}} + \sigma_u^2(\mathbf{f}_{ij}^{-1})(\mathbf{f}_{ij}^{-1})')$ $= \text{Diag}_{1 \leq j \leq m_i}(\boldsymbol{\Sigma}_{ij})$, say. First, we invert $\boldsymbol{\Sigma}_i$ by noting that $\boldsymbol{\Sigma}_i^{-1} = \text{Diag}_{1 \leq j \leq m_i}(\boldsymbol{\Sigma}_{ij}^{-1})$ and by applying Lemma 3.1 to $\boldsymbol{\Sigma}_{ij}$ with $\mathbf{C} = \sigma_\epsilon^2 \mathbf{I}_{n_{ij}}$, $\mathbf{a} = \sigma_u^2(\mathbf{f}_{ij}^{-1})$ and $\mathbf{b} = (\mathbf{f}_{ij}^{-1})$. We obtain $\boldsymbol{\Sigma}_{ij}^{-1} = \sigma_\epsilon^{-2} \text{Diag}_{1 \leq j \leq m_i}(\mathbf{I}_{n_{ij}} - \eta_{ij}^{-1} \gamma_{ij}(\mathbf{f}_{ij}^{-1})(\mathbf{f}_{ij}^{-1})')$. Finally, we arrive at an expression for $\mathbf{V}_i^{-1}$ using the expression for $\boldsymbol{\Sigma}_i^{-1}$ and applying Lemma 3.1 again, this time to $\mathbf{V}_i$. Now take $\mathbf{C} = \boldsymbol{\Sigma}_i$, $\mathbf{a} = \sigma_\epsilon^2(\mathbf{f}_i^{-1})$ and $\mathbf{b} = (\mathbf{f}_i^{-1})$ and obtain

$$\mathbf{V}^{-1} = \text{Diag}_{1 \leq i \leq t}(\mathbf{V}_i^{-1}) \quad (3.7)$$

where

$$\begin{aligned} \mathbf{V}_i^{-1} = & \sigma_\epsilon^{-2} \{ (\mathbf{I}_{n_i} - \text{Diag}_{1 \leq j \leq m_i}(\eta_{ij}^{-1} \gamma_{ij}(\mathbf{f}_{ij}^{-1})(\mathbf{f}_{ij}^{-1})')) \\ & - \sigma_\epsilon^2 \sigma_u^{-2} \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} [\text{col}_{1 \leq j \leq m_i}(\eta_{ij}^{-1} \gamma_{ij}(\mathbf{f}_{ij}^{-1}))][\text{col}_{1 \leq j \leq m_i}(\eta_{ij}^{-1} \gamma_{ij}(\mathbf{f}_{ij}^{-1}))]' \}. \end{aligned}$$

Lemma 3.1 (Rao (1965), Chapter 1, pp. 33, #2.8) *Let $\mathbf{C}$ be a $k \times k$ nonsingular matrix and $\mathbf{a}$ and $\mathbf{b}$ be two $k \times 1$ vectors. Then, the inverse of $\mathbf{C} + \mathbf{a}\mathbf{b}'$ is given by $(\mathbf{C} + \mathbf{a}\mathbf{b}')^{-1} = \mathbf{C}^{-1} - (1 + \mathbf{b}'\mathbf{C}^{-1}\mathbf{a})^{-1}\mathbf{C}^{-1}\mathbf{a}\mathbf{b}'\mathbf{C}^{-1}$.*

Remark 3.4 The calculation of $\hat{\beta}$ requires expressions for $\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}$ and $\mathbf{W}'\mathbf{V}^{-1}\mathbf{z}$. Using (3.7), we have

$$\begin{aligned} \mathbf{W}'\mathbf{V}^{-1}\mathbf{W} = & \sigma_\epsilon^{-2} \{ \sum_{i=1}^t \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} \mathbf{w}_{ijk} \mathbf{w}_{ijk}' - \sum_{i=1}^t \sum_{j=1}^{m_i} \eta_{ij} \gamma_{ij} \bar{\mathbf{x}}_{ijw} \bar{\mathbf{x}}_{ijw}' \\ & - \sigma_u^{-2} \sigma_\epsilon^2 \sum_{i=1}^t [\gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} (\sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ijw}) (\sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ijw}')] \} \end{aligned}$$

and

$$\begin{aligned} W'V^{-1}z &= \sigma_e^{-2} \left\{ \sum_{i=1}^i \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} w_{ijk} z_{ijk} - \sum_{i=1}^i \sum_{j=1}^{m_i} \eta_{ij} \gamma_{ij} x_{ijw} y_{ijw} \right. \\ &\quad \left. - \sigma_u^{-2} \sigma_e^2 \sum_{i=1}^i \left[\gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \left(\sum_{j=1}^{m_i} \gamma_{ij} x_{ijw} \right) \left(\sum_{j=1}^{m_i} \gamma_{ij} y_{ijw} \right) \right] \right\}. \end{aligned}$$

Note that the transformation method of Section 3.2 for calculating $\tilde{\beta}$ works only in the special case of $f_{ijk} = f_i$ and $n_{ij} = n_i$.

3.3.2 The Finite Population Case

We now direct our attention to the finite population case where we assume that M_i is large or that the sampling fraction $m_i M_i^{-1}$ is nonnegligible. Once again, we are interested in finding an estimate of the i^{th} small area population mean. First, we need to find the BLUP of u_{ij} . This is found through direct application of Theorem 2.1 with $\mathbf{k} = \mathbf{0}_p$ and

$$\mathbf{m} = \begin{pmatrix} \mathbf{0}_t \\ \text{col}_{1 \leq \ell \leq t} \quad \text{col}_{1 \leq s \leq m_i} (\delta_{\ell js}) \end{pmatrix}$$

where $\delta_{\ell js} = 1$ if $\ell = i$ and $s = j$ and $\delta_{\ell js} = 0$ otherwise, and $\mathbf{0}_p$ and $\mathbf{0}_t$ are $p \times 1$ and $t \times 1$ vectors, respectively, consisting entirely of zeros. The resultant BLUP of u_{ij} is given by

$$\tilde{u}_{ij} = \gamma_{ij}(y_{ijw} - \mathbf{x}_{ijw}'\tilde{\beta}) - \gamma_i \gamma_{ij} \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \sum_{j=1}^{m_i} \gamma_{ij} (y_{ijw} - \mathbf{x}_{ijw}'\tilde{\beta}). \quad (3.8)$$

Theorem 3.1, below, gives an expression for the BLUP of $\bar{Y}_i$, assuming the finite population case. The proof that it is indeed the BLUP follows.

Theorem 3.1 *Assuming a superpopulation model of the form (3.1) where $\sigma^2 = (\sigma_\xi^2, \sigma_u^2, \sigma_e^2)'$ is known, the BLUP of the i^{th} small area mean, $\bar{Y}_i$, for the finite population case, is given by*

$$t_i^F(\sigma^2, \mathbf{y}) = N_i^{-1} \left\{ \sum_{j \in s_i} \left(\sum_{k \in s_{ij}} y_{ijk} \right) + \sum_{j \in s_i} \left(\sum_{k \in s_{ij}} y_{ijk}^* \right) + \sum_{j \in s_i} \left(\sum_{k=1}^{N_{ij}} y_{ijk}^{**} \right) \right\}$$

where s_i is the set of all sampled PSUs in the i^{th} small area, demarcated by $j = 1, \dots, m_i$ without loss of generality, and s_{ij} is the set of all sampled SSUs in the j^{th} sampled PSU and the i^{th} small area, demarcated by $k = 1, \dots, n_{ij}$ without loss of generality. Also y_{ijk}^* is a predictor of the nonsampled y_{ijk} in the j^{th} sampled PSU within the i^{th} small area, while $y_{ij,k}^{**}$ is a predictor of the y_{ijk} in the j^{th} nonsampled PSU within the i^{th} small area. Specifically,

$$y_{ijk}^* = \mathbf{x}_{ijk}' \tilde{\boldsymbol{\beta}} + \tilde{\xi}_i + \tilde{u}_{ij}$$

and

$$y_{ij,k}^{**} = \mathbf{x}_{ijk}' \tilde{\boldsymbol{\beta}} + \tilde{\xi}_i$$

so that

$$\begin{aligned} t_i^F(\sigma^2, \mathbf{y}) = & N_i^{-1} \left\{ \sum_{j=1}^{m_i} n_{ij} \bar{y}_{ij.} + \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (\bar{\mathbf{x}}_{ij.}' \tilde{\boldsymbol{\beta}} + \tilde{\xi}_i + \tilde{u}_{ij}) \right. \\ & \left. + \left(\sum_{j=1}^{M_i} N_{ij} \bar{\mathbf{X}}_{ij.} - \sum_{j=1}^{m_i} N_{ij} \bar{\mathbf{X}}_{ij.} \right)' \tilde{\boldsymbol{\beta}} + \left(\sum_{j=1}^{M_i} N_{ij} - \sum_{j=1}^{m_i} N_{ij} \right) \tilde{\xi}_i \right\} \end{aligned} \quad (3.9)$$

where $\bar{\mathbf{x}}_{ij.} = (N_{ij} - n_{ij})^{-1} \sum_{k \in J_{ij}} \mathbf{x}_{ijk}$ is the mean vector of the $N_{ij} - n_{ij}$ nonsampled $\mathbf{x}_{ijk}$'s in the j^{th} sampled PSU and the i^{th} small area. Also, $\bar{\mathbf{X}}_{ij.} = N_{ij}^{-1} \sum_{k=1}^{N_{ij}} \mathbf{x}_{ijk}$ and $\bar{y}_{ij.} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} y_{ijk}$. Finally, $\tilde{u}_{ij}$ is defined by (3.8) and $\tilde{\xi}_i$ is as in (3.6).

Proof We can write

$$\bar{Y}_{i..} = N_i^{-1} \left[\sum_{j=1}^{m_i} n_{ij} \bar{y}_{ij.} + \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \bar{y}_{ij.}^* + \left(\sum_{j=1}^{M_i} N_{ij} \bar{Y}_{ij.} - \sum_{j=1}^{m_i} N_{ij} \bar{Y}_{ij.} \right) \right]$$

where

$$\bar{y}_{ij.}^* = (N_{ij} - n_{ij})^{-1} \sum_{k \in J_{ij}} y_{ijk} \text{ and } \bar{Y}_{ij.} = N_{ij}^{-1} \sum_{k=1}^{N_{ij}} y_{ijk}.$$

It follows that

$$\begin{aligned} E(t_i^F(\sigma^2, \mathbf{y})) &= N_i^{-1} \left(\sum_{j=1}^{m_i} n_{ij} \bar{\mathbf{x}}'_{ij} \beta + \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \mathbf{x}'_{ij} \beta \right) \\ &\quad + \left(\sum_{j=1}^{m_i} N_{ij} \bar{\mathbf{X}}'_{ij} \beta - \sum_{j=1}^{m_i} N_{ij} \bar{\mathbf{X}}'_{ij} \beta \right) = E(\bar{Y}_{i..}) \end{aligned}$$

where $\bar{\mathbf{x}}_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \mathbf{x}_{ijk}$. Therefore $t_i^F(\sigma^2, \mathbf{y})$ is an unbiased predictor of $\bar{Y}_{i..}$, which is linear in the y_{ijk} 's.

To prove the "best" property, we must show that $\text{MSE}(t_i^F(\sigma^2, \mathbf{y}))$ is smaller than the MSE of any other linear unbiased predictor. Assume then that $t_i^L(\sigma^2, \mathbf{y})$ is another linear unbiased predictor of $\bar{Y}_{i..}$ and can be written as $t_i^L(\sigma^2, \mathbf{y}) = t_i^F(\sigma^2, \mathbf{y}) + \mathbf{a}'\mathbf{y}$, where $\mathbf{y} = \begin{matrix} \text{col} \\ 1 \leq i \leq t \end{matrix} \begin{matrix} \text{col} \\ 1 \leq j \leq m_i \end{matrix} \begin{matrix} \text{col} \\ 1 \leq k \leq n_{ij} \end{matrix} (y_{ijk})$ and $\mathbf{a}$ is an $n \times 1$ vector of known constants, i.e., $\mathbf{a} = \begin{matrix} \text{col} \\ 1 \leq i \leq t \end{matrix} \begin{matrix} \text{col} \\ 1 \leq j \leq m_i \end{matrix} \begin{matrix} \text{col} \\ 1 \leq k \leq n_{ij} \end{matrix} (a_{ijk})$. Since $t_i^L(\sigma^2, \mathbf{y})$ is unbiased then $E(\mathbf{a}'\mathbf{y}) = 0$, which in turn implies that $\mathbf{a}'\mathbf{X} = 0 \forall \beta$ where

$\mathbf{X} = \begin{matrix} \text{col} \\ 1 \leq i \leq t \end{matrix} \begin{matrix} \text{col} \\ 1 \leq j \leq m_i \end{matrix} \begin{matrix} \text{col} \\ 1 \leq k \leq n_{ij} \end{matrix} (\mathbf{x}'_{ijk})$. Now, we can write

$$\begin{aligned} \text{MSE}(t_i^L(\sigma^2, \mathbf{y})) &= E(t_i^F(\sigma^2, \mathbf{y}) + \mathbf{a}'\mathbf{y} - \bar{Y}_{i..})^2 \\ &= \text{var}(\mathbf{a}'\mathbf{y} + t_i^F(\sigma^2, \mathbf{y}) - \bar{Y}_{i..}) \\ &= \text{var}(\mathbf{a}'\mathbf{y}) + \text{MSE}(t_i^F(\sigma^2, \mathbf{y})) + 2\text{cov}(\mathbf{a}'\mathbf{y}, t_i^F(\sigma^2, \mathbf{y}) - \bar{Y}_{i..}). \end{aligned}$$

It suffices to show that the covariance term is zero because then $\text{MSE}(t_i^L(\sigma^2, \mathbf{y})) \geq \text{MSE}(t_i^F(\sigma^2, \mathbf{y})) \forall t_i^L(\sigma^2, \mathbf{y})$ since $\text{var}(\mathbf{a}'\mathbf{y}) \geq 0$.

To show that the covariance term is in fact zero we write

$$\text{cov}(\mathbf{a}'\mathbf{y}, t_i^F(\sigma^2, \mathbf{y}) - \bar{Y}_{i..}) = \text{cov}(\mathbf{a}'\mathbf{y}, t_i^F(\sigma^2, \mathbf{y})) - \text{cov}(\mathbf{a}'\mathbf{y}, \bar{Y}_{i..})$$

and substitute in the expressions for $t_i^F(\sigma^2, \mathbf{y})$ and $\bar{Y}_{i..}$. This process results in a rather cumbersome expression which can be simplified considerably upon noting the following facts:

(i) $\text{cov}(\tilde{\beta}, \mathbf{a}'\mathbf{y}) = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\Sigma\mathbf{a} = 0$ since $\mathbf{X}'\mathbf{a} = 0$ where Σ is the variance-covariance matrix of $\mathbf{y}$.

(ii) $\text{cov}(y_{ijk}, \mathbf{a}'\mathbf{y}) = \sigma_u^2 \gamma_{ij}^{-1} \sum_{k=1}^{n_{ij}} a_{ijk} + \sigma_\xi^2 \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} a_{ijk}$ since the covariance structure for model (3.1) can be written as

$$\text{cov}(y_{ijk}, y_{i'j'k'}) = \begin{cases} \sigma_\xi^2 + \sigma_u^2 + \sigma_e^2 f_{ijk}^2 & \text{if } i = i', j = j', k = k' \\ \sigma_\xi^2 + \sigma_u^2 & \text{if } i = i', j = j', k \neq k' \\ \sigma_\xi^2 & \text{if } i = i', j \neq j' \\ 0 & \text{if } i \neq i'. \end{cases}$$

(iii) $\text{cov}(y_{ij.}^*, \mathbf{a}'\mathbf{y}) = \sigma_u^2 \sum_{k=1}^{n_{ij}} a_{ijk} + \sigma_\xi^2 \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} a_{ijk}$.

(iv) $\text{cov}(\sum_{j=1}^{M_i} N_{ij} \bar{Y}_{ij.} - \sum_{j=1}^{m_i} N_{ij} \bar{Y}_{ij.}, \mathbf{a}'\mathbf{y}) = \sigma_\xi^2 (\sum_{j=1}^{M_i} N_{ij} - \sum_{j=1}^{m_i} N_{ij}) \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} a_{ijk}$.

The simplified expression is

$$\begin{aligned} & \sigma_u^2 \gamma_i (\sum_{j=1}^{M_i} N_{ij})^{-1} (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} a_{ijk} \\ & \cdot [\sum_{j=1}^{m_i} (N_{ij} - n_{ij})(1 - \gamma_{ij}) + \sum_{j=1}^{M_i} N_{ij} - \sum_{j=1}^{m_i} N_{ij}] \\ & + \sigma_\xi^2 (\sum_{j=1}^{M_i} N_{ij})^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} a_{ijk} [\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \gamma_{ij} + \sum_{j=1}^{M_i} (N_{ij} - n_{ij})(1 - \gamma_{ij}) \gamma_i \\ & + (\sum_{j=1}^{M_i} N_{ij} - \sum_{j=1}^{m_i} N_{ij}) \gamma_i - \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) - \sum_{j=1}^{M_i} N_{ij} + \sum_{j=1}^{m_i} N_{ij}] \end{aligned}$$

which reduces to zero as desired, upon application of the identity $\sigma_\xi^2 (1 - \gamma_i) = \sigma_u^2 \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1}$. ■

Remark 3.5 If $m_i M_i^{-1} \rightarrow 0$, then $t_i^F(\sigma^2, \mathbf{y})$, the BLUP for the finite population case, tends to $t_i(\sigma^2, \mathbf{y}) = \bar{\mathbf{X}}_i' \tilde{\beta} + \tilde{\xi}_i$, the BLUP for the infinite population case. Note that a negligible sampling fraction makes sense in the small area context for some applications, since the size of the PSU may be large even though the samples size within the PSU is small.

3.4 Estimation of the Variance Components

The BLUPs given by (3.6) and (3.9) are obtained under the assumption that σ^2 , the vector of variance components, is known. In practice, σ^2 is unknown making it necessary to estimate σ^2 and then substitute the estimators into (3.6) and (3.9) to get the EBLUPs. We estimate σ^2 using Henderson's method of fitting constants, just as we did for the one-fold model in Chapter 2. Again, we show these estimates to be equivalent to those obtained using the Fuller and Battese transformation method.

3.4.1 Estimation of σ_ϵ^2 , σ_u^2 and σ_c^2

The method of fitting constants involves performing three OLS regressions on the model (3.3), since there are three unknown variance components. The method of moments is then used to form estimators, one from each regression.

Estimating σ_ϵ^2 involves a regression of $\mathbf{z}$ on $\mathbf{X}_1$ where $\mathbf{X}_1$ is the matrix $\mathbf{W}$ augmented by the matrix $\mathbf{D}$, i.e., $\mathbf{X}_1 = (\mathbf{W}|\mathbf{D}) = (\mathbf{W}|\mathbf{D}_1|\mathbf{D}_2)$. This is tantamount to treating both β and

$$\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}$$

as fixed effects. We calculate the residual sum of squares for this model, given by

$$\text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2) = \mathbf{z}'\mathbf{z} - \mathbf{z}'\mathbf{X}_1 (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{z}$$

as follows: First note that $\text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2)$ is unique for any choice of basis $\tilde{\mathbf{X}}_1$ for $\mathbf{X}_1$. In particular, we can choose $\tilde{\mathbf{X}}_1 = (\mathbf{W}|\mathbf{D}_2)$ if the model (3.3) does not contain an intercept term and $\tilde{\mathbf{X}}_1 = (\mathbf{W}_1|\mathbf{D}_2)$ if the model does contain an intercept term, where $\mathbf{W} = \left(\text{col}_{1 \leq i \leq t} (\mathbf{f}_i^{-1}) | \mathbf{W}_1 \right)$. Thus

$$\text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2) = \mathbf{z}'\mathbf{z} - \mathbf{z}'\tilde{\mathbf{X}}_1 (\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1)^{-1}\tilde{\mathbf{X}}_1'\mathbf{z},$$

noting that $\tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_1$ is nonsingular.

If an intercept is not included in the model, then applying the formula for the inverse of a partitioned matrix (see Rao (1965), pp. 33) to

$$\tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_1 = \begin{pmatrix} \mathbf{W}'\mathbf{W} & \mathbf{W}'\mathbf{D}_2 \\ \mathbf{D}_2'\mathbf{W} & \mathbf{D}_2'\mathbf{D}_2 \end{pmatrix},$$

we obtain after the simplification

$$\begin{aligned} \text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2) &= \mathbf{z}'\mathbf{z} - \mathbf{z}'\mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{z} \\ &\quad - (\mathbf{W}'\mathbf{z} - \mathbf{W}'\mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{z})'\mathbf{H}^{-1}(\mathbf{W}'\mathbf{z} - \mathbf{W}'\mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{z}) \end{aligned}$$

where

$$\mathbf{H} = \mathbf{W}'\mathbf{W} - \mathbf{W}'\mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{W}.$$

Now drawing on the fact that

$$\mathbf{z}'\mathbf{z} - \mathbf{z}'\mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{z} = (\mathbf{z} - \mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{z})'(\mathbf{z} - \mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{z}),$$

$$\mathbf{z}'\mathbf{W} - \mathbf{z}'\mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{W} = (\mathbf{z} - \mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{z})'(\mathbf{W} - \mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{W}),$$

and

$$\mathbf{H} = (\mathbf{W} - \mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{W})'(\mathbf{W} - \mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{W}),$$

we can finally write

$$\text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2) = \mathbf{z}_1^{*'}\mathbf{z}_1^* - \mathbf{z}_1^{*'}\mathbf{W}_1^*(\mathbf{W}_1^{*'}\mathbf{W}_1^*)^{-1}\mathbf{W}_1^{*'}\mathbf{z}_1^*$$

where

$$\mathbf{z}_1^* = \mathbf{z} - \mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{z}$$

and

$$\mathbf{W}_1^* = \mathbf{W} - \mathbf{D}_2(\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\mathbf{W}.$$

If the model includes an intercept term, then we simply replace $\mathbf{W}$ by $\mathbf{W}_1$ in $\mathbf{W}_1^*$.

Therefore, $\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1, \mathbf{u}_2)$ can also be thought of as the residual sum of squares due to the regression of $z_{ijk}^* = f_{ijk}^{-1}(y_{ijk} - y_{ijw})$ on $\mathbf{w}_{ijk}^* = f_{ijk}^{-1}(\mathbf{x}_{ijk} - \mathbf{x}_{ijw})$; $i = 1, \dots, t$; $j = 1, \dots, m_i$; $k = 1, \dots, n_{ij}$. This is indeed the regression to be employed as a logical extension of the transformation method of estimation of variance components, as suggested by Fuller and Battese (1973) for the case $f_{ijk} = 1$. Therefore, the two methods produce identical estimators for σ_e^2 .

The unbiased estimator $\hat{\sigma}_e^2$ of σ_e^2 is given by

$$\hat{\sigma}_e^2 = (n - p - \sum_{i=1}^t m_i + \lambda)^{-1} \text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1, \mathbf{u}_2) \quad (3.10)$$

where $\lambda = 1$ for the intercept model and $\lambda = 0$ for the no-intercept model. This follows by noting that

$$E[\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1, \mathbf{u}_2)] = (n - p - \sum_{i=1}^t m_i + \lambda)\sigma_e^2.$$

Estimating σ_u^2 involves a regression of $\mathbf{z}$ on $\mathbf{X}_2$ where $\mathbf{X}_2$ is the matrix $\mathbf{W}$ augmented by the matrix $\mathbf{D}_1$, i.e., $\mathbf{X}_2 = (\mathbf{W}|\mathbf{D}_1)$. This is tantamount to treating both $\boldsymbol{\beta}$ and $\mathbf{u}_1$ as fixed effects. We calculate the residual sum of squares for this model, given by

$$\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1) = \mathbf{z}'\mathbf{z} - \mathbf{z}'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{z},$$

much the same way as we calculate $\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1, \mathbf{u}_2)$ in deriving the estimator $\hat{\sigma}_e^2$ for σ_e^2 . Once again, we note that $\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1)$ is unique for any choice of basis $\mathbf{X}_2$ for $\mathbf{X}_2$. In particular, we can choose $\tilde{\mathbf{X}}_2 = (\mathbf{W}|\mathbf{D}_1)$ if the model (3.3) does not include an intercept term and $\tilde{\mathbf{X}}_2 = (\mathbf{W}_1|\mathbf{D}_1)$ if it does include one. Thus

$$\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1) = \mathbf{z}'\mathbf{z} - \mathbf{z}'\tilde{\mathbf{X}}_2(\tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_2)^{-1}\tilde{\mathbf{X}}_2'\mathbf{z},$$

noting that $\hat{\mathbf{X}}_2' \hat{\mathbf{X}}_2$ is nonsingular.

Now proceeding as above, we get, in the no-intercept case

$$\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1) = \mathbf{z}_2^{*\prime} \mathbf{z}_2^* - \mathbf{z}_2^{*\prime} \mathbf{W}_2^* (\mathbf{W}_2^{*\prime} \mathbf{W}_2^*)^{-1} \mathbf{W}_2^{*\prime} \mathbf{z}_2^*$$

where

$$\mathbf{z}_2^* = \mathbf{z} - \mathbf{D}_1 (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{D}_1' \mathbf{z}$$

and

$$\mathbf{W}_2^* = \mathbf{W} - \mathbf{D}_1 (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{D}_1' \mathbf{W}.$$

If the model includes an intercept term, then we simply replace $\mathbf{W}$ by $\mathbf{W}_1$ in $\mathbf{W}_2^*$.

Therefore, $\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1)$ can also be thought of as the residual sum of squares due to the regression of $z_{ijk}^{**} = f_{ijk}^{-1}(y_{ijk} - \bar{y}_{i\cdot})$ on $\mathbf{w}_{ijk}^{**} = f_{ijk}^{-1}(\mathbf{x}_{ijk} - \bar{\mathbf{x}}_{i\cdot})$; $i = 1, \dots, t$; $j = 1, \dots, m_i$; $k = 1, \dots, n_{ij}$ where

$$y_{i\cdot} = \eta_i^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} f_{ijk}^{-1} z_{ijk},$$

$$\bar{\mathbf{x}}_{i\cdot} = \eta_i^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} f_{ijk}^{-1} \mathbf{w}_{ijk},$$

and

$$\eta_i = \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} f_{ijk}^{-2} = \sum_{j=1}^{m_i} \eta_{ij}.$$

Again, this is the appropriate regression to be employed as a logical extension of the transformation method of estimation, as suggested by Fuller and Battese (1973), for the case $f_{ijk} = 1$. Therefore, the two methods produce identical estimators for σ_u^2 .

The unbiased estimator $\hat{\sigma}_u^2$ of σ_u^2 is given by

$$\begin{aligned} \hat{\sigma}_u^2 &= \eta_*^{-1} [\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1) - \hat{\sigma}_e^2(n - p - t + \lambda)] \\ &= \eta_*^{-1} [\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1) - \text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1, \mathbf{u}_2) - \hat{\sigma}_e^2(\sum_i m_i - t)] \end{aligned} \quad (3.11)$$

where

$$\eta_{\bullet} = \sum_{i=1}^t \eta_i - \sum_{i=1}^t \eta_i^{-1} \sum_{j=1}^{m_i} \eta_{ij}^2 - \sum_{i=1}^t \sum_{j=1}^{m_i} \eta_{ij}^2 (\mathbf{x}_{ijw} - \mathbf{x}_{iw})' (\mathbf{W}_2^{*'} \mathbf{W}_2^*)^{-1} (\mathbf{x}_{ijw} - \mathbf{x}_{iw}).$$

This follows by using Lemma 3.5 ahead, which offers an expression for the expectation of a quadratic form, and by noting that

$$\begin{aligned} E[\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1)] &= E[\mathbf{z}'(\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2') \mathbf{z}] \\ &= \text{trace} [(\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2') \mathbf{V}] \\ &= \eta_{\bullet} \sigma_{\mathbf{u}}^2 + (n - p - t + \lambda) \sigma_{\epsilon}^2 \end{aligned} \quad (3.12)$$

where $\mathbf{V}$ is given by (3.4).

Estimating σ_{ξ}^2 involves a regression of $\mathbf{z}$ on $\mathbf{W}$. Therefore

$$\text{SSE}(\boldsymbol{\beta}) = \mathbf{z}' \mathbf{z} - \mathbf{z}' \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \mathbf{z}$$

is the residual sum of squares due to the regression of z_{ijk} on $\mathbf{w}_{ijk}$.

The unbiased estimator $\hat{\sigma}_{\xi}^2$ of σ_{ξ}^2 is given by

$$\begin{aligned} \hat{\sigma}_{\xi}^2 &= \eta_{\bullet\bullet}^{-1} [\text{SSE}(\boldsymbol{\beta}) - \hat{\sigma}_{\epsilon}^2(n - p) - \hat{\sigma}_{\mathbf{u}}^2 \eta_{\bullet\bullet\bullet}] \\ &= \eta_{\bullet\bullet}^{-1} [\text{SSE}(\boldsymbol{\beta}) - \text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1, \mathbf{u}_2) - \hat{\sigma}_{\mathbf{u}}^2 \eta_{\bullet\bullet\bullet} - \hat{\sigma}_{\epsilon}^2 (\sum_i m_i - \lambda)] \end{aligned} \quad (3.13)$$

where

$$\eta_{\bullet\bullet} = \sum_{i=1}^t \eta_i - \sum_{i=1}^t \eta_i^2 \mathbf{x}_{iw}' (\mathbf{W}' \mathbf{W})^{-1} \mathbf{x}_{iw}$$

and

$$\eta_{\bullet\bullet\bullet} = \sum_{i=1}^t \sum_{j=1}^{m_i} \eta_{ij} - \sum_{i=1}^t \sum_{j=1}^{m_i} \eta_{ij}^2 \mathbf{x}_{ijw}' (\mathbf{W}' \mathbf{W})^{-1} \mathbf{x}_{ijw}.$$

This follows by applying Lemma 3.5 again, and by noting that

$$\begin{aligned}
E[\text{SSE}(\beta)] &= E[\mathbf{z}'(\mathbf{I}_n - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{z}] \\
&= \text{trace}[(\mathbf{I}_n - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{V}] \\
&= \eta_{\bullet\bullet}\sigma_\xi^2 + \eta_{\bullet\bullet\bullet}\sigma_{\mathbf{u}}^2 + (n-p)\sigma_\epsilon^2.
\end{aligned} \tag{3.14}$$

Any unbiased quadratic estimators for $\sigma_{\mathbf{u}}^2$ and σ_ξ^2 can take on negative values. We therefore truncate the estimator of $\sigma_{\mathbf{u}}^2$ to zero whenever $\hat{\sigma}_{\mathbf{u}}^2 < 0$, and truncate the estimator of σ_ξ^2 to zero whenever $\hat{\sigma}_\xi^2 < 0$. We define the truncated estimators as follows:

$$\hat{\sigma}_{\mathbf{u}}^2 = \max(0, \tilde{\sigma}_{\mathbf{u}}^2)$$

and

$$\hat{\sigma}_\xi^2 = \max(0, \tilde{\sigma}_\xi^2).$$

We set $\hat{\sigma}_\epsilon^2 = \tilde{\sigma}_\epsilon^2$ for consistency of notation.

Remark 3.6 The estimators $\hat{\sigma}_\xi^2$, $\hat{\sigma}_{\mathbf{u}}^2$ and $\hat{\sigma}_\epsilon^2$ can be expressed as quadratic forms as follows:

$$\hat{\sigma}_\epsilon^2 = \mathbf{z}'\mathbf{C}_1\mathbf{z}$$

where

$$\mathbf{C}_1 = (n - p - \sum_{i=1}^t m_i + \lambda)^{-1}[\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'],$$

$$\hat{\sigma}_{\mathbf{u}}^2 = \mathbf{z}'\mathbf{C}_2\mathbf{z}$$

where

$$\mathbf{C}_2 = \eta_\bullet^{-1}[\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_2' - (n - p - t + \lambda)\mathbf{C}_1],$$

and

$$\hat{\sigma}_{\xi}^2 = \mathbf{z}' \mathbf{C}_3 \mathbf{z}$$

where

$$\mathbf{C}_3 = \eta_{...}^{-1} [\mathbf{I}_n - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' - (n-p)\mathbf{C}_1 - \eta_{...}\mathbf{C}_2].$$

These forms are needed in Section 3.5.

3.5 A Second Order Approximation to the MSE of the EBLUP

Following the method of Section 2.5 to obtain an exact expression for the MSE of $t_i(\boldsymbol{\sigma}^2, \mathbf{y})$, we invoke Theorem 2.3 with $\mathbf{k} = \bar{\mathbf{X}}_{i..}$ and

$$\mathbf{m} = \begin{pmatrix} \text{col}_{1 \leq t \leq t} (\delta_{it}) \\ \mathbf{0}_{\Sigma m_i} \end{pmatrix},$$

and obtain after simplification

$$\begin{aligned} \text{MSE}(t_i(\boldsymbol{\sigma}^2, \mathbf{y})) &= \sigma_{\xi}^2(1 - \gamma_i) \\ &\quad + [\bar{\mathbf{X}}_{i..} - (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \gamma_i \sum_{j=1}^{m_i} \gamma_{ij} \mathbf{x}_{ijw}]' (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1} \\ &\quad \cdot [\bar{\mathbf{X}}_{i..} - (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \gamma_i \sum_{j=1}^{m_i} \gamma_{ij} \mathbf{x}_{ijw}] \\ &= g_{1i}(\boldsymbol{\sigma}^2) + g_{2i}(\boldsymbol{\sigma}^2), \end{aligned} \quad (3.15)$$

say. Once again, we can interpret $g_{1i}(\boldsymbol{\sigma}^2)$ as the MSE when $\boldsymbol{\beta}$ is known, and $g_{2i}(\boldsymbol{\sigma}^2)$ as the inflation due to estimating $\boldsymbol{\beta}$.

When $\boldsymbol{\sigma}^2$ is unknown, we can find an approximation correct to order $O(t^{-1})$ for the MSE of $t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y})$, the EBLUP of $\bar{Y}_{i..}$, in the same manner as we did for the one-fold case in Section 2.5. Once again, we have

$$\text{MSE}[t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y})] = \text{MSE}[t_i(\boldsymbol{\sigma}^2, \mathbf{y})] + E[t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y}) - t_i(\boldsymbol{\sigma}^2, \mathbf{y})]^2 \quad (3.16)$$

since the crossproduct term is zero under normality, where $\hat{\sigma}^2$, is a translation invariant estimator of σ^2 , and $\text{MSE}[t_i(\sigma^2, \mathbf{y})]$ is given by (3.15). As before, we consider regularity conditions equivalent to (C1)–(C4) cited in Chapter 2, but applied to model (3.3). Condition (C2) holds since, from Remark 3.2, we have that $n_i = n_i$, and

$\mathbf{R}_i = \sigma_c^2 \mathbf{I}_{m_i}$, and we can always lexicographically reorder $\mathbf{u}$ so that $\mathbf{G} = \text{col}_{1 \leq i \leq t} (\mathbf{G}_i)$ where

$$\mathbf{G}_i = \begin{pmatrix} \sigma_\xi^2 & 0 \\ 0 & \sigma_u^2 \mathbf{I}_{m_i} \end{pmatrix}$$

and $h_i = 1 + m_i$. Further, a condition equivalent to (C4) holds via Remark 3.6, since for model (3.3), $r = 3$, $\hat{\sigma}_1^2 = \hat{\sigma}_\xi^2$, $\hat{\sigma}_2^2 = \hat{\sigma}_u^2$ and $\hat{\sigma}_3^2 = \hat{\sigma}_c^2$. Hence we need only assume regularity conditions equivalent to (C1) and (C3) given by the following:

(C1)' The elements of $\mathbf{W}$ and $\mathbf{D}$ are uniformly bounded such that $\mathbf{W}'\mathbf{V}^{-1}\mathbf{W} = O(t)$.

(C3)' (i) $\sup_{1 \leq i \leq t} \sup_{1 \leq j \leq m_i} n_{ij} < \infty$ and
(ii) $\sup_{1 \leq i \leq t} m_i < \infty$.

Hence, assuming regularity conditions (C1)' and (C3)' on model (3.3), the Prasad-Rao approximation is given by

$$E[t_i(\hat{\sigma}^2, \mathbf{y}) - t_i(\sigma^2, \mathbf{y})]^2 \doteq \text{trace}\{(\nabla \mathbf{b})\mathbf{V}(\nabla \mathbf{b}')E[(\hat{\sigma}^2 - \sigma^2)(\hat{\sigma}^2 - \sigma^2)']\} \quad (3.17)$$

where

$$\nabla \mathbf{b}' = \text{col}_{1 \leq j \leq 3} (\partial \mathbf{b}' / \partial \sigma_j^2)$$

and

$$\mathbf{b}' = \mathbf{m}'\mathbf{G}\mathbf{D}'\mathbf{V}^{-1}.$$

For our model, the trace term simplifies to

$$\begin{aligned}
& \text{trace} \{ (\nabla \mathbf{b}) \mathbf{V} (\nabla \mathbf{b}') E[(\hat{\sigma}^2 - \sigma^2)(\hat{\sigma}^2 - \sigma^2)'] \} \\
&= \sigma_u^{-2} \sigma_\xi^4 (1 - \gamma_i)^2 \left\{ (1 - \gamma_i) \sum_{j=1}^{m_i} \gamma_{ij} \text{var}[(\sigma_\xi^{-2} \hat{\sigma}_\xi^2 - \sigma_\epsilon^{-2} \hat{\sigma}_\epsilon^2) \right. \\
&\quad \left. - (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij}^2 (\sigma_u^{-2} \hat{\sigma}_u^2 - \sigma_\epsilon^{-2} \hat{\sigma}_\epsilon^2)] \right. \\
&\quad \left. + (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} [(\sum_{j=1}^{m_i} \gamma_{ij}^3)(\sum_{j=1}^{m_i} \gamma_{ij}) - (\sum_{j=1}^{m_i} \gamma_{ij}^2)^2] \text{var}(\sigma_u^{-2} \hat{\sigma}_u^2 - \sigma_\epsilon^{-2} \hat{\sigma}_\epsilon^2) \right\} \\
&= g_{31}(\sigma^2), \tag{3.18}
\end{aligned}$$

say, where the expressions for $\text{var}(\hat{\sigma}_\epsilon^2)$, $\text{var}(\hat{\sigma}_u^2)$, $\text{var}(\hat{\sigma}_\xi^2)$, $\text{cov}(\hat{\sigma}_\epsilon^2, \hat{\sigma}_u^2)$, $\text{cov}(\hat{\sigma}_\epsilon^2, \hat{\sigma}_\xi^2)$ and $\text{cov}(\hat{\sigma}_u^2, \hat{\sigma}_\xi^2)$ are derived under the assumption of normality of the random errors, in Section 3.5.1. We obtain the above expression for the trace term upon noting that

$$\mathbf{b} = \sigma_\epsilon^{-2} \sigma_\xi^2 (1 - \gamma_i) \left[\text{col}_{1 \leq \ell \leq \ell} (\delta_{i\ell} - \text{col}_{1 \leq j \leq m_i} (1 - \gamma_{ij})(\Gamma_{ij}^{-1})) \right]$$

where $\delta_{i\ell} = \mathbf{1}_{n_i}$, if $\ell = i$ and $\mathbf{0}_{n_i}$, if $\ell \neq i$, and where $\mathbf{1}_{n_i}$ is an $n_i \times 1$ vector consisting entirely of ones, while $\mathbf{0}_{n_i}$ is an $n_i \times 1$ vector consisting entirely of zeros. Also,

$$(\nabla \mathbf{b}) \mathbf{V} (\nabla \mathbf{b}') = \begin{pmatrix} \sum_{j=1}^{m_i} a_{ij} a_{ij}^* \eta_{ij} & \sum_{j=1}^{m_i} b_{ij} a_{ij}^* \eta_{ij} & \sum_{j=1}^{m_i} c_{ij} a_{ij}^* \eta_{ij} \\ \sum_{j=1}^{m_i} a_{ij} b_{ij}^* \eta_{ij} & \sum_{j=1}^{m_i} b_{ij} b_{ij}^* \eta_{ij} & \sum_{j=1}^{m_i} c_{ij} b_{ij}^* \eta_{ij} \\ \sum_{j=1}^{m_i} a_{ij} c_{ij}^* \eta_{ij} & \sum_{j=1}^{m_i} b_{ij} c_{ij}^* \eta_{ij} & \sum_{j=1}^{m_i} c_{ij} c_{ij}^* \eta_{ij} \end{pmatrix}$$

where

$$a_{ij}^* = \sigma_\epsilon^2 a_{ij} + \sigma_u^2 \eta_{ij} a_{ij} + \sigma_\xi^2 \sum_{j=1}^{m_i} \eta_{ij} a_{ij},$$

$$b_{ij}^* = \sigma_\epsilon^2 b_{ij} + \sigma_u^2 \eta_{ij} b_{ij} + \sigma_\xi^2 \sum_{j=1}^{m_i} \eta_{ij} b_{ij},$$

$$c_{ij}^* = \sigma_\epsilon^2 c_{ij} + \sigma_u^2 \eta_{ij} c_{ij} + \sigma_\xi^2 \sum_{j=1}^{m_i} \eta_{ij} c_{ij},$$

and where

$$a_{ij} = \partial[\sigma_e^{-2}\sigma_\xi^2(1-\gamma_i)(1-\gamma_{ij})]/\partial\sigma_\xi^2 = \sigma_e^{-2}(1-\gamma_{ij})(1-\gamma_i)^2,$$

$$b_{ij} = \partial[\sigma_e^{-2}\sigma_\xi^2(1-\gamma_i)(1-\gamma_{ij})]/\partial\sigma_u^2 = \sigma_e^{-2}\sigma_u^{-2}\sigma_\xi^2(1-\gamma_{ij})(1-\gamma_i) \\ \cdot [(\sum_{j=1}^{m_i}\gamma_{ij})^{-1}\gamma_i\sum_{j=1}^{m_i}\gamma_{ij}^2 - \gamma_{ij}]$$

and

$$c_{ij} = \partial[\sigma_e^{-2}\sigma_\xi^2(1-\gamma_i)(1-\gamma_{ij})]/\partial\sigma_e^2 = -\sigma_e^{-2}\sigma_\xi^2c_{ij} - \sigma_e^{-2}\sigma_u^2b_{ij}.$$

Hence a second order approximation to the MSE of $t_i(\hat{\sigma}^2, \mathbf{y})$ correct to order $O(t^{-1})$ is given by

$$\begin{aligned} \text{MSE}(t_i(\hat{\sigma}^2, \mathbf{y})) &\doteq \sigma_\xi^2(1-\gamma_i) \\ &+ (\bar{\mathbf{X}}_{i..} - (\sum_{j=1}^{m_i}\gamma_{ij})^{-1}\gamma_i\sum_{j=1}^{m_i}\gamma_{ij}\bar{\mathbf{x}}_{ijw})'(\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1} \\ &\cdot (\bar{\mathbf{X}}_{i..} - (\sum_{j=1}^{m_i}\gamma_{ij})^{-1}\gamma_i\sum_{j=1}^{m_i}\gamma_{ij}\bar{\mathbf{x}}_{ijw}) \\ &+ \sigma_u^{-2}\sigma_\xi^4(1-\gamma_i)^2\{(1-\gamma_i)\sum_{j=1}^{m_i}\gamma_{ij}\text{var}[(\sigma_\xi^{-2}\hat{\sigma}_\xi^2 - \sigma_e^{-2}\hat{\sigma}_e^2) \\ &- (\sum_{j=1}^{m_i}\gamma_{ij})^{-1}\sum_{j=1}^{m_i}\gamma_{ij}^2(\sigma_u^{-2}\hat{\sigma}_u^2 - \sigma_e^{-2}\hat{\sigma}_e^2)] \\ &+ (\sum_{j=1}^{m_i}\gamma_{ij})^{-1}[(\sum_{j=1}^{m_i}\gamma_{ij}^3)(\sum_{j=1}^{m_i}\gamma_{ij}) - (\sum_{j=1}^{m_i}\gamma_{ij}^2)^2]\text{var}(\sigma_u^{-2}\hat{\sigma}_u^2 - \sigma_e^{-2}\hat{\sigma}_e^2)\}. \\ &\doteq g_{1i}(\sigma^2) + g_{2i}(\sigma^2) + g_{3i}(\sigma^2) \end{aligned} \quad (3.19)$$

where the variances of $\hat{\sigma}_\xi^2$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ and the covariances between them are given in Section 3.5.1.

A second order approximation for the finite population case cannot be obtained in as straightforward a manner as was done for the one-fold case, so will be omitted here.

Due to the complexity of the proof of (3.17), we confine ourselves to model (3.2), specializing to the common mean case with $f_{ijk} = 1$, in the following theorem and ensuing proof.

Theorem 3.2 Assume the common mean model $y_{ijk} = \beta + \xi_i + \mu_{ij} + \epsilon_{ijk}$; $i = 1, \dots, t$; $j = 1, \dots, m_i$; $k = 1, \dots, n_{ij}$, with normal random errors ξ_i , μ_{ij} and ϵ_{ijk} . In addition, assume regularity conditions (C1)' and (C3)' cited earlier, with $f_{ijk} = 1$, $p = 1$, $x_{ijk1} = 1$ and $\beta_1 = \beta$. Assume the specific bounds on the m_i 's and n_{ij} 's to be as follows: $2 \leq \inf_{1 \leq i \leq t} m_i$, $\sup_{1 \leq i \leq t} m_i \leq m_u$, $2 \leq \inf_{1 \leq i \leq t} \inf_{1 \leq j \leq m_i} n_{ij}$ and $\sup_{1 \leq i \leq t} \sup_{1 \leq j \leq m_i} n_{ij} \leq n_u$. Then

$$\begin{aligned} E[t_i(\hat{\sigma}^2, \mathbf{y}) - t_i(\sigma^2, \mathbf{y})]^2 \\ = \text{trace}\{(\nabla \mathbf{b})\mathbf{V}(\nabla \mathbf{b}')E[(\hat{\sigma}^2 - \sigma^2)(\hat{\sigma}^2 - \sigma^2)']\} + o(t^{-1}) \end{aligned}$$

where

$$t_i(\sigma^2, \mathbf{y}) = (1 - \gamma_i)\hat{\beta} + \gamma_i y_{i\gamma}.$$

Here

$$\begin{aligned} \hat{\beta} &= (\sum_{i=1}^t \gamma_i)^{-1} \sum_{i=1}^t \gamma_i y_{i\gamma}, \\ y_{i\gamma} &= (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} y_{ij}, \\ \gamma_{ij} &= \sigma_u^2 (\sigma_u^2 + \sigma_e^2 n_{ij}^{-1})^{-1} \end{aligned}$$

and

$$\gamma_i = \sigma_\xi^2 (\sigma_\xi^2 + \sigma_u^2 (\sum_{j=1}^{m_i} \gamma_{ij})^{-1})^{-1}.$$

Proof Firstly, $t_i(\sigma^2, \mathbf{y})$, the BLUP of $\bar{Y}_{i..}$, can be seen to be the expression given in the body of the theorem by a direct application of Theorem 2.1 with $k = 1$,

$$\mathbf{m} = \begin{pmatrix} \text{col}_{1 \leq i \leq t} (\delta_{it}) \\ \mathbf{0}_{\Sigma m_i} \end{pmatrix}$$

and $\mathbf{V}^{-1}$ as in (3.7) with $f_{ijk} = 1$.

To verify that the neglected terms in the second order approximation are of order $\alpha(t^{-1})$, we take a Taylor expansion of $t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y})$ about $\boldsymbol{\sigma}^2$. Thus

$$t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y}) - t_i(\boldsymbol{\sigma}^2, \mathbf{y}) = (\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{d}_i(\boldsymbol{\sigma}^2) + \frac{1}{2}(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{D}_i(\boldsymbol{\sigma}^{*2})(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)$$

where $|\boldsymbol{\sigma}^{*2} - \boldsymbol{\sigma}^2| \leq |\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2|$, $\mathbf{d}_i(\boldsymbol{\sigma}^2)$ is the 3×1 vector of first order derivatives of $t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y})$ with respect to $\hat{\boldsymbol{\sigma}}^2$, evaluated at $\hat{\boldsymbol{\sigma}}^2 = \boldsymbol{\sigma}^2$, and $\mathbf{D}_i(\boldsymbol{\sigma}^{*2})$ is the 3×3 matrix of second order derivatives of $t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y})$ with respect to $\hat{\boldsymbol{\sigma}}^2$, evaluated at $\hat{\boldsymbol{\sigma}}^2 = \boldsymbol{\sigma}^{*2}$.

Squaring and applying the expectation operator to the expansion yields

$$\begin{aligned} E[t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y}) - t_i(\boldsymbol{\sigma}^2, \mathbf{y})]^2 &= E[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{d}_i(\boldsymbol{\sigma}^2)]^2 \\ &\quad + \frac{1}{4} E[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{D}_i(\boldsymbol{\sigma}^{*2})(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)]^2 \\ &\quad + E\{[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{d}_i(\boldsymbol{\sigma}^2)] [(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{D}_i(\boldsymbol{\sigma}^{*2})(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)]\} \\ &= E[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{d}_i(\boldsymbol{\sigma}^2)]^2 + R_1 + R_2, \text{ say.} \end{aligned}$$

Now in Appendix A, Lemma A.1, we show that

$$E[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{d}_i(\boldsymbol{\sigma}^2)]^2 = E[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{d}_i(\boldsymbol{\sigma}^2)]^2 + \alpha(t^{-1}).$$

Also,

$$E[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \mathbf{d}_i(\boldsymbol{\sigma}^2)]^2 = \text{trace}\{(\mathbf{V}\mathbf{b})\mathbf{V}(\mathbf{V}\mathbf{b})' E[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)']\} + \alpha(t^{-1})$$

upon direct application of Theorem A.1 of Prasad and Rao (1990), noting that regularity conditions (C1)' and (C3)' are satisfied for the model at hand.

It remains to show that both R_1 and R_2 are $\alpha(t^{-1})$. For the sake of brevity, we shall only show that $R_2 = \alpha(t^{-1})$ since the proof that $R_1 = \alpha(t^{-1})$ is similar in

nature. Now, R_2 can be written $R_2 = E(U_2)$ where

$$\begin{aligned}
U_2 &= [(\hat{\sigma}^2 - \sigma^2)' d_1(\sigma^2)] [(\hat{\sigma}^2 - \sigma^2)' D_1(\sigma^{*2})(\hat{\sigma}^2 - \sigma^2)] \\
&= \left(\sum_{m=1}^3 (\hat{\sigma}_m^2 - \sigma_m^2) \partial t_1(\hat{\sigma}^2, y) / \partial \hat{\sigma}_m^2 \right) \Big|_{\sigma_m^2 = \sigma_m^2} \\
&\quad \cdot \left(\sum_{\ell=1}^3 \sum_{s=1}^3 (\hat{\sigma}_\ell^2 - \sigma_\ell^2)(\hat{\sigma}_s^2 - \sigma_s^2) \partial^2 t_1(\hat{\sigma}^2, y) / \partial \hat{\sigma}_\ell^2 \partial \hat{\sigma}_s^2 \right) \Big|_{\substack{\hat{\sigma}_\ell^2 = \sigma_\ell^{*2} \\ \hat{\sigma}_s^2 = \sigma_s^{*2}}} \\
&= \left(\sum_{m=1}^3 (\hat{\sigma}_m^2 - \sigma_m^2) t'_{im} \right) \left(\sum_{\ell=1}^3 \sum_{s=1}^3 (\hat{\sigma}_\ell^2 - \sigma_\ell^2)(\hat{\sigma}_s^2 - \sigma_s^2) t''_{is} \right),
\end{aligned}$$

say, where $\sigma_1^2 = \sigma_\xi^2$, $\sigma_2^2 = \sigma_u^2$ and $\sigma_3^2 = \sigma_\epsilon^2$. To proceed further, we require the use of the following well-known lemma:

Lemma 3.2 (Serfling (1980), Theorem C, pp. 15) *If $\{X_n\}$ is a sequence of random variables such that $X_n = o_p(1)$ and $\sup_{n \geq 1} E|X_n|^{1+\delta} < \infty$ for some $0 < \delta < 1$ then $E(X_n) = o(1)$.*

Now, appealing to the above lemma, we need only show that

- (a) $U_2 = o_p(t^{-1})$ and
- (b) $E|tU_2|^{1+\delta_1} < \infty$ for some $0 < \delta_1 < 1$.

Proof of (a) It suffices to deal with U_2 in a componentwise fashion, showing that every term in U_2 has the desired property. Hence, we must show that

$$(\hat{\sigma}_m^2 - \sigma_m^2) t'_{im} (\hat{\sigma}_\ell^2 - \sigma_\ell^2) (\hat{\sigma}_s^2 - \sigma_s^2) t''_{is} = o_p(t^{-1}) \text{ for } m, \ell, s = 1, 2, 3.$$

Now there are three possible cases:

- (i) $m = \ell = s$
- (ii) $m = \ell$ or $m = s$ or $\ell = s$
- (iii) $m \neq \ell \neq s$.

We shall deal only with case (i), since the proofs for all three cases are similar in nature. Directing our attention then to case (i), we must show that

$$(\hat{\sigma}_m^2 - \sigma_m^2)^3 t'_{im} t''_{im} = o_p(t^{-1}) \text{ for } m = 1, 2, 3.$$

Focusing on the first term, if we can show that $\hat{\sigma}_m^2 - \sigma_m^2 = o_p(t^{-\epsilon})$ for $0 < \epsilon < \frac{1}{2}$ and $m = 1, 2, 3$ then $(\hat{\sigma}_m^2 - \sigma_m^2)^3 = o_p(t^{-1})$ by choosing $\epsilon = \frac{1}{3}$. The proof that $\hat{\sigma}_m^2 - \sigma_m^2 = o_p(t^{-\epsilon})$ for $0 < \epsilon < \frac{1}{2}$ is relegated to Appendix A, Lemma A.2 (c).

Next, we must show that $t'_{im} = O_p(1)$ for $m = 1, 2, 3$. In the body of the proof of part b), we show that $E|t'_{im}|^8 = O(1)$ for $m = 1, 2, 3$. This, in turn, implies that $E|t'_{im}| = O(1)$ for $m = 1, 2, 3$ since any lower order moment is also bounded. Appealing to the well-known lemma below, we have that $t'_{im} = O_p(1)$ for $m = 1, 2, 3$.

Lemma 3.3 (Bhat (1981), Lemma (5.2), pp. 85) *Let $\{X_n\}$ be a sequence of random variables. Then $\sup_{n \geq 1} E|X_n| < \infty$ implies $X_n = O_p(1)$.*

Similarly, we can show that $t''_{im} = O_p(1)$ for $m = 1, 2, 3$. Piecing the three terms together, we have the desired result.

Proof of (b) Again it suffices to deal with U_2 in a componentwise fashion, showing that every term in U_2 has the desired property. Hence, we must show that

$$t^{1+\delta_1} E\{|\hat{\sigma}_m^2 - \sigma_m^2|^{1+\delta_1} |t'_{im}|^{1+\delta_1} |\hat{\sigma}_\ell^2 - \sigma_\ell^2|^{1+\delta_1} |\hat{\sigma}_s^2 - \sigma_s^2|^{1+\delta_1} |t''_{i\ell s}|^{1+\delta_1}\} \leq c_0$$

for some $c_0 < \infty$ and for $m, \ell, s = 1, 2, 3$. Again, there are three possible cases:

- (i) $m = \ell = s$
- (ii) $m = \ell$ or $m = s$ or $\ell = s$
- (iii) $m \neq \ell \neq s$.

We shall deal only with case (i), since the proof for all three cases are similar in nature. Directing our attention, then, to case (i), we must show that

$$t^{1+\delta_1} E\{|\hat{\sigma}_m^2 - \sigma_m^2|^{3+\delta_2} |t'_{im}|^{1+\delta_1} |t''_{imm}|^{1+\delta_1}\} = E(U_2^*) \leq c_0$$

for some $c_0 < \infty$, for $m = 1, 2, 3$ and for some $0 < \delta_1 < 1$ and $\delta_2 = 3\delta_1$.

We shall shortly make use of the following four well-known inequalities, summarized in Lemma 3.4.

Lemma 3.4 (a) **Holder's Inequality:** Let X and Y be random variables. Then $E|XY| \leq (E|X|^r)^{1/r} (E|Y|^s)^{1/s}$ for $r^{-1} + s^{-1} = 1$ and $r, s > 1$.

(b) **C_r Inequality:** For random variables X and Y , we have $E|X + Y|^r \leq C_r(E|X|^r + E|Y|^r)$ where $C_r = 1$ if $r < 1$ and $C_r = 2^{r-1}$ if $r \geq 1$.

(c) **Minkowski's Inequality:** Let $\{X_i\}$ be a sequence of random variables. Then $(E|\sum_i X_i|^p)^{1/p} \leq \sum_i (E|X_i|^p)^{1/p}$ for $p \geq 1$.

(d) **Markov's Inequality:** If X is a random variable then $Pr(|X| > a) \leq E|X|/a$ for $a > 0$.

Now, appealing to Holders' inequality twice, the first time with $r = s = 2$, and the second time with $r = 3/2$, $s = 3$, we have

$$\begin{aligned} E(U_2^*) &\leq t^{1+\delta_1} (E|\hat{\sigma}_m^2 - \sigma_m^2|^{2(3+\delta_2)})^{\frac{1}{2}} (E|t''_{imm}|^{2(1+\delta_1)\frac{3}{2}})^{\frac{1}{2} \cdot \frac{2}{3}} \\ &\quad \cdot (E|t'_{im}|^{2(1+\delta_1)3})^{\frac{1}{2} \cdot \frac{1}{3}}. \end{aligned}$$

Choosing $\delta_1 = 1/3$ and hence $\delta_2 = 1$, we need only show that

$$t^{\frac{4}{3}} (E|\hat{\sigma}_m^2 - \sigma_m^2|^8)^{\frac{1}{2}} (E|t'_{im}|^6)^{\frac{1}{6}} (E|t''_{imm}|^4)^{\frac{1}{3}} = O(1). \quad (3.20)$$

Now, we show that $t^{\frac{4}{3}}(E|\hat{\sigma}_m^2 - \sigma_m^2|^8)^{\frac{1}{3}} = O(t^{-2/3})$ by showing that $E|\hat{\sigma}_m^2 - \sigma_m^2|^8 = O(t^{-4})$ as follows:

$$\begin{aligned} E|\hat{\sigma}_m^2 - \sigma_m^2|^8 &= E\{|\hat{\sigma}_m^2 - \sigma_m^2|^8 [I(\hat{\sigma}_m^2 \leq 0) + I(\hat{\sigma}_m^2 > 0)]\} \\ &= |\sigma_m^2|^8 Pr(\hat{\sigma}_m^2 \leq 0) + E\{|\hat{\sigma}_m^2 - \sigma_m^2|^8 I(\hat{\sigma}_m^2 > 0)\} \\ &\leq |\sigma_m^2|^8 Pr(\hat{\sigma}_m^2 \leq 0) + E|\hat{\sigma}_m^2 - \sigma_m^2|^8 \end{aligned}$$

where $I(\hat{\sigma}_m^2 > 0) = 1$ if $\hat{\sigma}_m^2 > 0$ and zero otherwise; $I(\hat{\sigma}_m^2 \leq 0)$ is defined analogously. Now $|\sigma_m^2|^8$ is bounded, and $Pr(\hat{\sigma}_m^2 \leq 0) = O(t^{-4})$ using Lemma A.2(b) from Appendix A with $s = 4$. Finally, $E|\hat{\sigma}_m^2 - \sigma_m^2|^8 = O(t^{-4})$ by Lemma A.3 from Appendix A with $s = 4$.

Turning our attention to the second term of (3.20), we need to show that $E|t'_{im}|^8$ is bounded. In order to show this, we need the following result:

$$|t'_{im}| \leq a_m |\bar{y}_{i\gamma}| + b_m t^{-1} \sum_{i=1}^i |\bar{y}_{i\gamma}| \text{ for } m = 1, 2, 3,$$

the proof of which is relegated to Appendix A, Lemma A.4. Now, applying this result, followed by the C_r and Minkowski inequalities successively, we obtain

$$\begin{aligned} E|t'_{im}|^8 &\leq E|a_m |\bar{y}_{i\gamma}| + b_m t^{-1} \sum_{i=1}^i |\bar{y}_{i\gamma}||^8 \\ &\leq 2^7 [|a_m|^8 E|\bar{y}_{i\gamma}|^8 + |b_m|^8 E|t^{-1} \sum_{i=1}^i |\bar{y}_{i\gamma}||^8] \\ &\leq 2^7 |a_m|^8 E|\bar{y}_{i\gamma}|^8 + 2^7 |b_m|^8 t^{-8} (\sum_{i=1}^i (E|\bar{y}_{i\gamma}|^8)^{1/8})^8. \end{aligned}$$

To show that the above expression is bounded, we need only show that $E|y_{i\gamma}|^{8+\delta}$ is bounded. Noting that $(\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \gamma_{ij} \leq 1$ and then applying Minkowski's inequality we get

$$\begin{aligned} E|y_{i\gamma}|^8 &= E|(\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{y}_{ij}|^8 \leq E(\sum_{j=1}^{m_i} |\bar{y}_{ij}|)^8 \\ &\leq (\sum_{j=1}^{m_i} (E|\bar{y}_{ij}|^8)^{1/8})^8 \end{aligned}$$

which is bounded if $E|y_{ij}|^8$ is bounded. Applying Minkowski's inequality again, we have

$$\begin{aligned} E|y_{ij}|^8 &= (n_{ij})^{-8} E\left|\sum_{k=1}^{n_{ij}} y_{ijk}\right|^8 \\ &\leq (n_{ij})^{-8} \left(\sum_{k=1}^{n_{ij}} (E|y_{ijk}|^8)^{1/8}\right)^8 < \infty \end{aligned}$$

since $n_{ij} \geq 2 \forall (i, j)$, and since under the assumption of normality of the random errors, all moments of the y_{ijk} 's exist and are finite.

Lastly, we turn our attention to the third term of (3.20). The proof that $E|t''_{imn}|^4 < \infty$ is omitted since it is identical in nature to that of the second term. We must however make mention of the fact that the proof requires the following result:

$$|t''_{i\ell s}| \leq a_{\ell s}|y_{i\gamma}| + b_{\ell s}t^{-1} \sum_{\gamma=1}^t |y_{i\gamma}| \text{ for } \ell, s = 1, 2, 3.$$

The proof of this is relegated to Appendix A, Lemma A.5. ■

3.5.1 The Variances and Covariances of the Henderson Estimators Under the Assumption of Normality

We are interested in finding expressions for the variances of the Henderson estimates, as well as the covariances between these estimates, under model (3.2). This enables the evaluation of the term $g_{31}(\sigma^2)$ given by expression (3.18). We assume normality of the random errors ξ_i , u_{ij} and e_{ijk} , and use the lemma that follows.

Lemma 3.5 (Searle (1971); Theorem 1, pp. 55; Corollary 1.2, pp. 57; pp. 451) *If $\mathbf{z}$ is an $n \times 1$ random vector such that $\mathbf{z} \sim N(\boldsymbol{\mu}, \mathbf{V})$ then $E(\mathbf{z}'\mathbf{P}\mathbf{z}) = \text{trace}(\mathbf{P}\mathbf{V}) + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu}$, $\text{var}(\mathbf{z}'\mathbf{P}\mathbf{z}) = 2 \text{trace}(\mathbf{P}\mathbf{V})^2 + 4\boldsymbol{\mu}'\mathbf{P}\mathbf{V}\mathbf{P}\boldsymbol{\mu}$ and $\text{cov}(\mathbf{z}'\mathbf{P}\mathbf{z}, \mathbf{z}'\mathbf{Q}\mathbf{z}) = 2 \text{trace}(\mathbf{P}\mathbf{V}\mathbf{Q}\mathbf{V}) + 4\boldsymbol{\mu}'\mathbf{P}\mathbf{V}\mathbf{Q}\boldsymbol{\mu}$ where $\mathbf{P}$ and $\mathbf{Q}$ are $n \times n$ known matrices. The first result on the expectation of a quadratic form is also true when $\mathbf{z}$ is nonnormal.*

We shall deal with each variance and covariance term separately.

(i) The unbiased estimator $\hat{\sigma}_\epsilon^2$ of σ_ϵ^2 is given by (3.10) as

$$\hat{\sigma}_\epsilon^2 = (n - p - \sum_{i=1}^t m_i + \lambda)^{-1} \text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2)$$

where

$$\text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2) = \mathbf{z}'[\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'] \mathbf{z} = \mathbf{z}' \mathbf{M}_1 \mathbf{z},$$

say, and $\mathbf{X}_1 = (\mathbf{W}|\mathbf{D}_1|\mathbf{D}_2)$ as before. Now $\mathbf{V}$, from (3.4), can also be written as $\mathbf{V} = \sigma_\epsilon^2 \mathbf{I}_n + \sigma_u^2 \mathbf{D}_2 \mathbf{D}_2' + \sigma_\epsilon^2 \mathbf{D}_1 \mathbf{D}_1'$. Invoking Lemma 3.5 with this $\mathbf{V}$, we have

$$\text{var}(\text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2)) = 2 \text{trace}(\mathbf{M}_1 \mathbf{V} \mathbf{M}_1 \mathbf{V}) + 4(\mathbf{W}\beta)' \mathbf{M}_1 \mathbf{V} \mathbf{M}_1 (\mathbf{W}\beta).$$

The second term is zero since $\mathbf{M}_1 \mathbf{W} = \mathbf{0}$, and, focusing on the first term, one can show that $\mathbf{M}_1 \mathbf{V} = \sigma_\epsilon^2 \mathbf{M}_1$ using the fact that $\mathbf{M}_1 \mathbf{D}_1 = \mathbf{0}$ and $\mathbf{M}_1 \mathbf{D}_2 = \mathbf{0}$. Thus

$$\text{var}(\text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2)) = 2\sigma_\epsilon^4 \text{trace}(\mathbf{M}_1) = 2\sigma_\epsilon^4(n - p - \sum_{i=1}^t m_i + \lambda)$$

since $\mathbf{M}_1$ is idempotent. Therefore

$$\text{var}(\hat{\sigma}_\epsilon^2) = 2\sigma_\epsilon^4(n - p - \sum_{i=1}^t m_i + \lambda)^{-1}. \quad (3.21)$$

This result also follows by noting that $\text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2)/\sigma_\epsilon^2$ is a chi-square variable with $n - p - \sum_{i=1}^t m_i + \lambda$ degrees of freedom.

(ii) The unbiased estimator $\hat{\sigma}_u^2$ of σ_u^2 is given in (3.11) as

$$\hat{\sigma}_u^2 = \eta_*^{-1}[\text{SSE}(\beta, \mathbf{u}_1) - \hat{\sigma}_\epsilon^2(n - p - t + \lambda)]$$

where

$$\eta_* = \sum_{i=1}^t \eta_i - \sum_{i=1}^t \sum_{j=1}^{m_i} \eta_i^{-1} \eta_{ij}^2 - \sum_{i=1}^t \sum_{j=1}^{m_i} \eta_{ij}^2 (\bar{\mathbf{x}}_{ijw} - \bar{\mathbf{x}}_{iww})' (\mathbf{W}_2^{*'} \mathbf{W}_2^*)^{-1} (\bar{\mathbf{x}}_{ijw} - \bar{\mathbf{x}}_{iww}),$$

$$\mathbf{W}_2^* = \text{col}_{1 \leq i \leq t} \text{col}_{1 \leq j \leq m_i} \text{col}_{1 \leq k \leq n_{ij}} (f_{ijk}^{-1}(\mathbf{x}_{ijk} - \bar{\mathbf{x}}_{iww})),$$

$$\text{SSE}(\beta, \mathbf{u}_1) = \mathbf{z}'[\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2'] \mathbf{z} = \mathbf{z}' \mathbf{M}_2 \mathbf{z},$$

say $\mathbf{X}_2 = (\mathbf{W}|\mathbf{D}_1)$ as before. Now $\tilde{\sigma}_{\mathbf{u}}^2$ can also be written as

$$\tilde{\sigma}_{\mathbf{u}}^2 = \eta_{\bullet}^{-1} [SS_1 - \tilde{\sigma}_{\epsilon}^2 (\sum_{i=1}^t m_i - t)]$$

where $SS_1 = \text{SSE}(\beta, \mathbf{u}_1) - \text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2) = \mathbf{z}'(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{z}$. We can show that SS_1 and $\tilde{\sigma}_{\epsilon}^2$ are independent by invoking the lemma that follows.

Lemma 3.6 (Searle (1971), Theorem 4, pp. 59) *When $\mathbf{z} \sim N(\mu, \mathbf{V})$, the quadratic forms $\mathbf{z}'\mathbf{A}\mathbf{z}$ and $\mathbf{z}'\mathbf{B}\mathbf{z}$ are distributed independently iff $\mathbf{A}\mathbf{V}\mathbf{B} = \mathbf{0}$ or, equivalently, $\mathbf{B}\mathbf{V}\mathbf{A} = \mathbf{0}$.*

Now, from Lemma 3.6, we see that $\mathbf{z}'(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{z}$ and $\mathbf{z}'\mathbf{M}_1\mathbf{z}$ are independent or that $(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{V}\mathbf{M}_1 = \mathbf{0}$ by noting that $\mathbf{V}\mathbf{M}_1 = \sigma_{\epsilon}^2\mathbf{M}_1$, that $\mathbf{M}_2\mathbf{M}_1 = \mathbf{M}_1$ and that $\mathbf{M}_1$ is idempotent. Therefore

$$\text{var}(\tilde{\sigma}_{\mathbf{u}}^2) = \eta_{\bullet}^{-2} [\text{var}(SS_1) + (\sum_{i=1}^t m_i - t)^2 \text{var}(\tilde{\sigma}_{\epsilon}^2)].$$

We invoke Lemma 3.5 once again to enable us to find $\text{var}(SS_1)$ and upon doing so, we have that

$$\text{var}(SS_1) = 2 \text{ trace } [(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{V}]^2 + 4(\mathbf{W}\beta)'(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{V}(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{W}\beta.$$

Once again, the second term is zero since both $\mathbf{M}_1\mathbf{W} = \mathbf{0}$ and $\mathbf{M}_2\mathbf{W} = \mathbf{0}$.

Focusing on the first term, we see that $(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{V}\mathbf{M}_1\mathbf{V} = \mathbf{0}$ from the independence of SS_1 and $\tilde{\sigma}_{\epsilon}^2$. Also,

$$\text{trace } (\mathbf{M}_2 - \mathbf{M}_1)\mathbf{V}\mathbf{M}_2\mathbf{V} = \text{trace } (\mathbf{M}_2\mathbf{V}\mathbf{M}_2\mathbf{V}) - \sigma_{\epsilon}^4(n - p - \sum_{i=1}^t m_i + \lambda)$$

since $\mathbf{M}_1\mathbf{V} = \sigma_{\epsilon}^2\mathbf{M}_1$ and $\mathbf{M}_1\mathbf{M}_2 = \mathbf{M}_1$. Finally, $\mathbf{M}_2\mathbf{V} = \sigma_{\epsilon}^2\mathbf{M}_2 + \sigma_{\mathbf{u}}^2\mathbf{M}_2\mathbf{D}_2\mathbf{D}_2'$ since $\mathbf{M}_2\mathbf{D}_1 = \mathbf{0}$. Therefore

$$\text{trace}(\mathbf{M}_2\mathbf{V}_2)^2 = \sigma_{\epsilon}^4 \text{ trace}(\mathbf{M}_2) + \sigma_{\mathbf{u}}^4 \text{ trace}(\mathbf{M}_2\mathbf{D}_2\mathbf{D}_2')^2 + 2\sigma_{\epsilon}^2\sigma_{\mathbf{u}}^2 \text{ trace } (\mathbf{M}_2\mathbf{D}_2\mathbf{D}_2').$$

Now, it is easy to see that $\text{trace}(\mathbf{M}_2) = n - p - t + \lambda$ and that $\text{trace}(\mathbf{M}_2 \mathbf{D}_2 \mathbf{D}_2') = \eta_*$ by using the fact that

$$\begin{aligned} E(\text{SSE}(\boldsymbol{\beta}, \mathbf{u}_1)) &= \text{trace}(\mathbf{M}_2 \mathbf{V}) = \sigma_\epsilon^2 \text{trace}(\mathbf{M}_2) + \sigma_{\mathbf{u}}^2 \text{trace}(\mathbf{M}_2 \mathbf{D}_2 \mathbf{D}_2') \\ &= \sigma_\epsilon^2(n - p - t + \lambda) + \sigma_{\mathbf{u}}^2 \eta_*, \end{aligned}$$

which follows from Lemma 3.5, and from (3.12).

Piecing the above results together, we have

$$\text{var}(SS_1) = 2\sigma_\epsilon^4 \left(\sum_{i=1}^t m_i - t \right) + 2\sigma_{\mathbf{u}}^4 \text{trace}(\mathbf{M}_2 \mathbf{D}_2 \mathbf{D}_2')^2 + 4\sigma_\epsilon^2 \sigma_{\mathbf{u}}^2 \eta_*$$

and hence

$$\begin{aligned} \text{var}(\hat{\sigma}_{\mathbf{u}}^2) &= 2\eta_*^{-2} \left\{ \sigma_\epsilon^4 \left(\sum_{i=1}^t m_i - t \right) (n - p - t + \lambda) \left(n - p - \sum_{i=1}^t m_i + \lambda \right)^{-1} \right. \\ &\quad \left. + \sigma_{\mathbf{u}}^4 \text{trace}(\mathbf{M}_2 \mathbf{D}_2 \mathbf{D}_2')^2 + 2\sigma_\epsilon^2 \sigma_{\mathbf{u}}^2 \eta_* \right\}. \end{aligned} \quad (3.22)$$

(iii) Using the alternative expression for $\hat{\sigma}_{\mathbf{u}}^2$ and the independence of SS_1 and $\hat{\sigma}_\epsilon^2$, we automatically have

$$\text{cov}(\hat{\sigma}_{\mathbf{u}}^2, \hat{\sigma}_\epsilon^2) = -\eta_*^{-1} \left(\sum_{i=1}^t m_i - t \right) \text{var}(\hat{\sigma}_\epsilon^2). \quad (3.23)$$

(iv) The unbiased estimator $\hat{\sigma}_\xi^2$ of σ_ξ^2 is given in (3.13) as

$$\hat{\sigma}_\xi^2 = n_{**}^{-1} [\text{SSE}(\boldsymbol{\beta}) - (n - p) \hat{\sigma}_\epsilon^2 - \eta_{***} \hat{\sigma}_{\mathbf{u}}^2]$$

where

$$\begin{aligned} \eta_{**} &= \sum_{i=1}^t \eta_i - \sum_{i=1}^t \eta_i^2 \bar{\mathbf{x}}_{i**}' (\mathbf{W}' \mathbf{W})^{-1} \bar{\mathbf{x}}_{i**}, \\ \eta_{***} &= \sum_{i=1}^t \sum_{j=1}^{m_i} \eta_{ij} - \sum_{i=1}^t \sum_{j=1}^{m_i} \eta_{ij}^2 \bar{\mathbf{x}}_{ijw}' (\mathbf{W}' \mathbf{W})^{-1} \bar{\mathbf{x}}_{ijw}, \end{aligned}$$

and

$$\text{SSE}(\boldsymbol{\beta}) = \mathbf{z}' [\mathbf{I}_n - \mathbf{W}(\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}'] \mathbf{z} = \mathbf{z}' \mathbf{M}_3 \mathbf{z},$$

say. Now $\hat{\sigma}_\xi^2$ can also be written as

$$\hat{\sigma}_\xi^2 = \eta_{**}^{-1} [SS_2 + c_1 \hat{\sigma}_\epsilon^2 - c_2 SS_1]$$

where

$$SS_2 = \text{SSE}(\beta) - \text{SSE}(\beta, \mathbf{u}_1, \mathbf{u}_2) = \mathbf{z}'(\mathbf{M}_3 - \mathbf{M}_1)\mathbf{z},$$

$$c_1 = -(\sum_{i=1}^l m_i - \lambda) + \eta_{**}^{-1} \eta_{***} (\sum_{i=1}^l m_i - t)$$

and

$$c_2 = \eta_{**}^{-1} \eta_{***}.$$

From part (ii), SS_1 and $\hat{\sigma}_\epsilon^2$ are independent. We show that SS_2 and $\hat{\sigma}_\epsilon^2$ are independent by invoking Lemma 3.6. Thus $(\mathbf{M}_3 - \mathbf{M}_1)\mathbf{V}\mathbf{M}_1 = \mathbf{0}$ since $\mathbf{V}\mathbf{M}_1 = \sigma_\epsilon^2 \mathbf{M}_1$, $\mathbf{M}_3\mathbf{M}_1 = \mathbf{M}_1$ and $\mathbf{M}_1$ is idempotent. Therefore,

$$\text{var}(\hat{\sigma}_\xi^2) = \eta_{**}^{-2} [\text{var}(SS_2) + c_1^2 \text{var}(\hat{\sigma}_\epsilon^2) + c_2^2 \text{var}(SS_1) - 2c_2 \text{cov}(SS_1, SS_2)].$$

But we can show that $\text{cov}(SS_1, SS_2) = \text{var}(SS_1)$ by using as an intermediate step that $SS_3 = \text{SSE}(\beta) - \text{SSE}(\beta, \mathbf{u}_1)$ and $\text{SSE}(\beta, \mathbf{u}_1)$ are independent. Therefore

$$\text{var}(\hat{\sigma}_\xi^2) = \eta_{**}^{-2} [\text{var}(SS_2) + c_1^2 \text{var}(\hat{\sigma}_\epsilon^2) + c_2(c_2 - 2) \text{var}(SS_1)].$$

We need an expression for $\text{var}(SS_2)$. Invoking Lemma 3.5, we have

$$\text{var}(SS_2) = 2 \text{trace}[(\mathbf{M}_3 - \mathbf{M}_1)\mathbf{V}]^2.$$

Now $(\mathbf{M}_3 - \mathbf{M}_1)\mathbf{V}\mathbf{M}_1\mathbf{V} = \mathbf{0}$ since SS_2 and $\hat{\sigma}_\epsilon^2$ are independent. In addition,

$$\begin{aligned} \text{trace}[(\mathbf{M}_3 - \mathbf{M}_1)\mathbf{V}\mathbf{M}_3\mathbf{V}] &= \text{trace}(\mathbf{M}_3\mathbf{V})^2 - \text{trace}(\mathbf{M}_1\mathbf{V}\mathbf{M}_1\mathbf{V}) \\ &= \text{trace}(\mathbf{M}_3\mathbf{V})^2 - \sigma_\epsilon^4(n - p - \sum_{i=1}^l m_i + \lambda) \end{aligned}$$

since $\mathbf{M}_1\mathbf{V} = \sigma_\epsilon^2 \mathbf{M}_1$ and $\mathbf{M}_1\mathbf{M}_3 = \mathbf{M}_1$. Now $\mathbf{M}_3\mathbf{V} = \sigma_\epsilon^2 \mathbf{M}_3 + \sigma_u^2 \mathbf{M}_3 \mathbf{D}_2 \mathbf{D}_2' + \sigma_\xi^2 \mathbf{M}_3 \mathbf{D}_1 \mathbf{D}_1'$, so that

$$\begin{aligned} \text{trace}(\mathbf{M}_3\mathbf{V})^2 &= \sigma_\epsilon^4 \text{trace}(\mathbf{M}_3) + \sigma_u^4 \text{trace}(\mathbf{M}_3 \mathbf{D}_2 \mathbf{D}_2')^2 + \sigma_\xi^4 \text{trace}(\mathbf{M}_3 \mathbf{D}_1 \mathbf{D}_1')^2 \\ &\quad + 2\sigma_\epsilon^2 \sigma_u^2 \text{trace}(\mathbf{M}_3 \mathbf{D}_2 \mathbf{D}_2') + 2\sigma_\epsilon^2 \sigma_\xi^2 \text{trace}(\mathbf{M}_3 \mathbf{D}_1 \mathbf{D}_1') \\ &\quad + 2\sigma_u^2 \sigma_\xi^2 \text{trace}(\mathbf{M}_3 \mathbf{D}_2 \mathbf{D}_2' \mathbf{M}_3 \mathbf{D}_1 \mathbf{D}_1'). \end{aligned}$$

It is easy to see that $\text{trace}(\mathbf{M}_3) = n - p$, $\text{trace}(\mathbf{M}_3 \mathbf{D}_2 \mathbf{D}_2') = \eta_{\bullet\bullet\bullet}$ and that $\text{trace}(\mathbf{M}_3 \mathbf{D}_1 \mathbf{D}_1') = \eta_{\bullet\bullet}$ by using the fact that

$$\begin{aligned} E(\text{SSE}(\boldsymbol{\beta})) &= \text{trace}(\mathbf{M}_3 \mathbf{V}) = \sigma_\epsilon^2 \text{trace}(\mathbf{M}_3) + \sigma_{\mathbf{u}}^2 \text{trace}(\mathbf{M}_3 \mathbf{D}_2 \mathbf{D}_2') \\ &\quad + \sigma_\xi^2 \text{trace}(\mathbf{M}_3 \mathbf{D}_1 \mathbf{D}_1') = \sigma_\epsilon^2(n - p) + \sigma_{\mathbf{u}}^2 \eta_{\bullet\bullet\bullet} + \sigma_\xi^2 \eta_{\bullet\bullet}, \end{aligned}$$

which follows from Lemma 3.5, and from (3.14).

Piecing the above results together, we have

$$\begin{aligned} \text{var}(\tilde{\sigma}_\xi^2) &= 2\eta_{\bullet\bullet\bullet}^{-2} \{ \sigma_\epsilon^4 (n - p - \sum_{i=1}^t m_i + \lambda)^{-1} [(\sum_{i=1}^t m_i - \lambda)(n - p) \\ &\quad + c_2^2 (\sum_{i=1}^t m_i - t)(n - p - t + \lambda) - 2c_2 (\sum_{i=1}^t m_i - t)(n - p)] \\ &\quad + \sigma_{\mathbf{u}}^4 [c_2(c_2 - 2) \text{trace}(\mathbf{M}_2 \mathbf{D}_2 \mathbf{D}_2')^2 + \text{trace}(\mathbf{M}_3 \mathbf{D}_2 \mathbf{D}_2')^2] \\ &\quad + \sigma_\xi^4 \text{trace}(\mathbf{M}_3 \mathbf{D}_1 \mathbf{D}_1')^2 + 2\sigma_\epsilon^2 \sigma_{\mathbf{u}}^2 [c_2(c_2 - 2) \eta_{\bullet\bullet} + \eta_{\bullet\bullet\bullet}] \\ &\quad + 2\sigma_\epsilon^2 \sigma_\xi^2 \eta_{\bullet\bullet} + 2\sigma_{\mathbf{u}}^2 \sigma_\xi^2 \text{trace}(\mathbf{M}_3 \mathbf{D}_2 \mathbf{D}_2' \mathbf{M}_3 \mathbf{D}_1 \mathbf{D}_1') \}. \end{aligned} \quad (3.24)$$

(v) Using the alternative expression for $\tilde{\sigma}_\xi^2$ and the independence of $\tilde{\sigma}_\epsilon^2$ from both SS_1 and SS_2 , we automatically have that

$$\text{cov}(\tilde{\sigma}_\xi^2, \tilde{\sigma}_\epsilon^2) = \eta_{\bullet\bullet\bullet}^{-1} c_1 \text{var}(\tilde{\sigma}_\epsilon^2). \quad (3.25)$$

(vi) Using the alternative expressions for $\tilde{\sigma}_\xi^2$ and $\tilde{\sigma}_{\mathbf{u}}^2$ and the independence of $\tilde{\sigma}_\epsilon^2$ from both SS_1 and SS_2 , we also have that

$$\begin{aligned} \text{cov}(\tilde{\sigma}_\xi^2, \tilde{\sigma}_{\mathbf{u}}^2) &= (\eta_{\bullet\bullet} \eta_{\bullet\bullet\bullet})^{-1} \{ (1 - c_2) [2\sigma_\epsilon^4 (\sum_{i=1}^t m_i - t) \\ &\quad + 2\sigma_{\mathbf{u}}^4 \text{trace}(\mathbf{M}_2 \mathbf{D}_2 \mathbf{D}_2')^2 + 4\sigma_\epsilon^2 \sigma_{\mathbf{u}}^2 \eta_{\bullet\bullet}] \\ &\quad - c_1 (\sum_{i=1}^t m_i - t) \text{var}(\tilde{\sigma}_\epsilon^2) \}. \end{aligned} \quad (3.26)$$

3.6 Estimation of the MSE of the EBLUP

Mimicking the methods of Section 2.6, we need to find an estimator of the second order approximation to $\text{MSE}(t_i(\hat{\sigma}^2, \mathbf{y}))$, given by $g_{1i}(\sigma^2) + g_{2i}(\sigma^2) + g_{3i}(\sigma^2)$, such that the estimator has a bias of order $o(t^{-1})$; that is,

$$E[\text{mse}(t_i(\hat{\sigma}^2, \mathbf{y})) - \text{MSE}(t_i(\hat{\sigma}^2, \mathbf{y}))] = o(t^{-1})$$

where $\text{mse}(t_i(\hat{\sigma}^2, \mathbf{y}))$ is the estimator in question.

Now since $g_{2i}(\sigma^2)$ and $g_{3i}(\sigma^2)$ are of order $O(t^{-1})$, it follows that

$$E[g_{2i}(\hat{\sigma}^2)] = g_{2i}(\sigma^2) + o(t^{-1})$$

and

$$E[g_{3i}(\hat{\sigma}^2)] = g_{3i}(\sigma^2) + o(t^{-1})$$

where

$$g_{2i}(\hat{\sigma}^2) = g_{2i}(\sigma^2)|_{\sigma^2 = \hat{\sigma}^2}$$

and

$$g_{3i}(\hat{\sigma}^2) = g_{3i}(\sigma^2)|_{\sigma^2 = \hat{\sigma}^2}.$$

Therefore $g_{2i}(\hat{\sigma}^2)$ and $g_{3i}(\hat{\sigma}^2)$ are the correct estimators of $g_{2i}(\sigma^2)$ and $g_{3i}(\sigma^2)$ respectively, each with bias $o(t^{-1})$.

Now $g_{1i}(\sigma^2)$ is of order $O(1)$, so $g_{1i}(\hat{\sigma}^2) = g_{1i}(\sigma^2)|_{\sigma^2 = \hat{\sigma}^2}$ is not the correct estimator of $g_{1i}(\sigma^2)$, since its bias is of order $O(t^{-1})$. Thus, the bias in $g_{1i}(\hat{\sigma}^2)$ must be estimated to the correct order. We now show that

$$E[g_{1i}(\hat{\sigma}^2) + g_{3i}(\hat{\sigma}^2)] = g_{1i}(\sigma^2) + o(t^{-1}).$$

The Taylor expansion of $g_{11}(\hat{\sigma}^2)$ about the point σ^2 yields

$$g_{11}(\hat{\sigma}^2) = g_{11}(\sigma^2) + (\hat{\sigma}^2 - \sigma^2)' \nabla g_{11}(\sigma^2) + 2^{-1} (\hat{\sigma}^2 - \sigma^2)' \nabla^2 g_{11}(\sigma^2) (\hat{\sigma}^2 - \sigma^2) + O_p(t^{-3/2})$$

where $\nabla g_{11}(\sigma^2)$ is a 3×1 vector of first order derivatives of $g_{11}(\sigma^2)$ with respect to σ^2 and $\nabla^2 g_{11}(\sigma^2)$ is a 3×3 matrix of second order derivatives of $g_{11}(\sigma^2)$ with respect to σ^2 . Applying the expectation operator to the above expression yields

$$E(g_{11}(\hat{\sigma}^2)) = g_{11}(\sigma^2) + E[(\hat{\sigma}^2 - \sigma^2)' \nabla g_{11}(\sigma^2)] + 2^{-1} E[(\hat{\sigma}^2 - \sigma^2)' \nabla^2 g_{11}(\sigma^2) (\hat{\sigma}^2 - \sigma^2)] + O(t^{-2}),$$

noting that $E[O_p(t^{-3/2})] = O(t^{-2})$.

Now

$$\begin{aligned} E[(\hat{\sigma}^2 - \sigma^2)' \nabla g_{11}(\sigma^2)] &= \sum_{m=1}^3 \nabla g_{11m}(\sigma^2) E(\hat{\sigma}_m^2 - \sigma_m^2) \\ &= \sum_{m=1}^3 \nabla g_{11m}(\sigma^2) E(\hat{\sigma}_m^2 - \sigma_m^2) = o(t^{-1}) \end{aligned}$$

where $\nabla g_{11m}(\sigma^2)$ is the m^{th} element of $\nabla g_{11}(\sigma^2)$. The above expression is $o(t^{-1})$ since $\nabla g_{11m}(\sigma^2) = O(1)$, $E(\hat{\sigma}_m^2) = \sigma_m^2$, and since $E|\hat{\sigma}_m^2 - \sigma_m^2| = o(t^{-1})$ from the proof of Lemma A.2(c). In addition, we can write

$$\begin{aligned} E[(\hat{\sigma}^2 - \sigma^2)' \nabla^2 g_{11}(\sigma^2) (\hat{\sigma}^2 - \sigma^2)] &= \sum_{l=1}^3 \sum_{s=1}^3 \nabla^2 g_{11ls}(\sigma^2) E[(\hat{\sigma}_l^2 - \sigma_l^2)(\hat{\sigma}_s^2 - \sigma_s^2)] \\ &= \sum_{l=1}^3 \sum_{s=1}^3 \nabla^2 g_{11ls}(\sigma^2) E[(\hat{\sigma}_l^2 - \sigma_l^2)(\hat{\sigma}_s^2 - \sigma_s^2)] \\ &\quad + \sum_{l=1}^3 \sum_{s=1}^3 \nabla^2 g_{11ls}(\sigma^2) E[(\hat{\sigma}_l^2 - \hat{\sigma}_l^2)(\hat{\sigma}_s^2 - \sigma_s^2)] \\ &\quad + \sum_{l=1}^3 \sum_{s=1}^3 \nabla^2 g_{11ls}(\sigma^2) E[(\hat{\sigma}_l^2 - \sigma_l^2)(\hat{\sigma}_s^2 - \hat{\sigma}_s^2)] \\ &\quad + \sum_{l=1}^3 \sum_{s=1}^3 \nabla^2 g_{11ls}(\sigma^2) E[(\hat{\sigma}_l^2 - \sigma_l^2)(\hat{\sigma}_s^2 - \hat{\sigma}_s^2)] \\ &= \sum_{l=1}^3 \sum_{s=1}^3 \nabla^2 g_{11ls}(\sigma^2) E[(\hat{\sigma}_l^2 - \sigma_l^2)(\hat{\sigma}_s^2 - \sigma_s^2)] + o(t^{-1}) \end{aligned}$$

where $\nabla^2 g_{1,ts}(\boldsymbol{\sigma}^2)$ is the $(t, s)^{\text{th}}$ element of $\nabla^2 g_{1t}(\boldsymbol{\sigma}^2)$. We can verify that the three neglected terms above are in fact $\alpha(t^{-1})$ by showing that $E[(\hat{\sigma}_t^2 - \tilde{\sigma}_t^2)(\hat{\sigma}_s^2 - \tilde{\sigma}_s^2)] = \alpha(t^{-1})$ and that $E[(\hat{\sigma}_t^2 - \tilde{\sigma}_t^2)(\hat{\sigma}_s^2 - \sigma_s^2)] = \alpha(t^{-1})$, and by noting that $\nabla^2 g_{1,ts}(\boldsymbol{\sigma}^2) = 0(1)$. Now

$$E[(\hat{\sigma}_t^2 - \tilde{\sigma}_t^2)(\hat{\sigma}_s^2 - \tilde{\sigma}_s^2)] \leq (E|\hat{\sigma}_t^2 - \tilde{\sigma}_t^2|^2)^{1/2}(E|\hat{\sigma}_s^2 - \tilde{\sigma}_s^2|^2)^{1/2} = \alpha(t^{-1})$$

upon applying Holder's inequality and noting that $E|\hat{\sigma}_t^2 - \tilde{\sigma}_t^2|^2 = \alpha(t^{-1})$ using an argument similar to one used in the proof that $E|\hat{\sigma}_t^2 - \tilde{\sigma}_t^2| = \alpha(t^{-1})$ from Lemma A.2(c). Similarly, $E[(\hat{\sigma}_t^2 - \tilde{\sigma}_t^2)(\hat{\sigma}_s^2 - \sigma_s^2)] = \alpha(t^{-1})$ by applying Holder's inequality and noting that $E|\hat{\sigma}_s^2 - \sigma_s^2|^2 = \text{var}(\tilde{\sigma}_s^2) = O(t^{-1})$ for $s = 1, 2, 3$.

Piecing the above results together, we have that,

$$\begin{aligned} E(g_{1t}(\hat{\boldsymbol{\sigma}}^2)) &= g_{1t}(\boldsymbol{\sigma}^2) + 2^{-1} E[(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)' \nabla^2 g_{1t}(\boldsymbol{\sigma}^2)(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2)] + \alpha(t^{-1}) \\ &= g_{1t}(\boldsymbol{\sigma}^2) + 2^{-1} \text{trace}[\text{var}(\hat{\boldsymbol{\sigma}}^2) \nabla^2 g_{1t}(\boldsymbol{\sigma}^2)] + \alpha(t^{-1}) \end{aligned}$$

by Lemma 3.5, where

$$\text{var}(\hat{\boldsymbol{\sigma}}^2) = \begin{pmatrix} \text{var}(\hat{\sigma}_\xi^2) & \text{cov}(\hat{\sigma}_\xi^2, \hat{\sigma}_u^2) & \text{cov}(\hat{\sigma}_\xi^2, \hat{\sigma}_\epsilon^2) \\ \text{cov}(\hat{\sigma}_\xi^2, \hat{\sigma}_u^2) & \text{var}(\hat{\sigma}_u^2) & \text{cov}(\hat{\sigma}_u^2, \hat{\sigma}_\epsilon^2) \\ \text{cov}(\hat{\sigma}_\xi^2, \hat{\sigma}_\epsilon^2) & \text{cov}(\hat{\sigma}_u^2, \hat{\sigma}_\epsilon^2) & \text{var}(\hat{\sigma}_\epsilon^2) \end{pmatrix}$$

and the elements of $\text{var}(\hat{\boldsymbol{\sigma}}^2)$ are given by expressions (3.21)–(3.26) assuming normality of the random errors. Furthermore,

$$\nabla^2 g_{1t}(\boldsymbol{\sigma}^2) = \begin{pmatrix} \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_\xi^4 & \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_\xi^2 \partial \sigma_u^2 & \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_\xi^2 \partial \sigma_\epsilon^2 \\ \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_\xi^2 \partial \sigma_u^2 & \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_u^4 & \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_u^2 \partial \sigma_\epsilon^2 \\ \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_\xi^2 \partial \sigma_\epsilon^2 & \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_u^2 \partial \sigma_\epsilon^2 & \partial^2 g_{1t}(\boldsymbol{\sigma}^2)/\partial \sigma_\epsilon^4 \end{pmatrix}$$

where

$$\begin{aligned}
\frac{\partial^2 g_{1i}(\boldsymbol{\sigma}^2)}{\partial \sigma_\xi^4} &= -2(\sigma_\xi^2)^{-1} \gamma_i (1 - \gamma_i)^2, \\
\frac{\partial^2 g_{1i}(\boldsymbol{\sigma}^2)}{\partial \sigma_u^4} &= -2\sigma_\xi^4 (\sigma_u^6)^{-1} (1 - \gamma_i)^2 \left[\sum_{j=1}^{m_i} \gamma_{ij}^3 - \gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \left(\sum_{j=1}^{m_i} \gamma_{ij}^2 \right)^2 \right], \\
\frac{\partial^2 g_{1i}(\boldsymbol{\sigma}^2)}{\partial \sigma_c^4} &= -2\sigma_\xi^4 (\sigma_u^2 \sigma_c^4)^{-1} (1 - \gamma_i)^2 \left[\sum_{j=1}^{m_i} \gamma_{ij} (1 - \gamma_{ij})^2 \right. \\
&\quad \left. - \gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \left(\sum_{j=1}^{m_i} \gamma_{ij} (1 - \gamma_{ij}) \right)^2 \right], \\
\frac{\partial^2 g_{1i}(\boldsymbol{\sigma}^2)}{\partial \sigma_\xi^2 \partial \sigma_u^2} &= -2(\sigma_u^2)^{-1} \gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} (1 - \gamma_i)^2 \sum_{j=1}^{m_i} \gamma_{ij}^2, \\
\frac{\partial^2 g_{1i}(\boldsymbol{\sigma}^2)}{\partial \sigma_\xi^2 \partial \sigma_c^2} &= 2\sigma_\xi^2 (\sigma_c^2 \sigma_u^2)^{-1} (1 - \gamma_i)^3 \sum_{j=1}^{m_i} \gamma_{ij} (1 - \gamma_{ij}),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 g_{1i}(\boldsymbol{\sigma}^2)}{\partial \sigma_u^4 \partial \sigma_c^2} &= -2\sigma_\xi^4 (\sigma_u^4 \sigma_c^2)^{-1} (1 - \gamma_i)^2 \left[(1 - \gamma_i) \sum_{j=1}^{m_i} \gamma_{ij}^2 - \sum_{j=1}^{m_i} \gamma_{ij}^3 \right. \\
&\quad \left. + \gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \left(\sum_{j=1}^{m_i} \gamma_{ij}^2 \right)^2 \right].
\end{aligned}$$

Upon simplification we have that

$$2^{-1} \text{trace} [\text{var}(\hat{\boldsymbol{\sigma}}^2) \nabla^2 g_{1i}(\boldsymbol{\sigma}^2)] = -g_{3i}(\boldsymbol{\sigma}^2)$$

and so

$$E(g_{1i}(\hat{\boldsymbol{\sigma}}^2)) = g_{1i}(\boldsymbol{\sigma}^2) - g_{3i}(\boldsymbol{\sigma}^2) + o(t^{-1}).$$

Therefore $g_{1i}(\hat{\boldsymbol{\sigma}}^2) + g_{3i}(\hat{\boldsymbol{\sigma}}^2)$ is the correct estimator of $g_{1i}(\boldsymbol{\sigma}^2)$, having bias of order $o(t^{-1})$. Combining the above results, we have that

$$\text{mse}(t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y})) = g_{1i}(\hat{\boldsymbol{\sigma}}^2) + g_{2i}(\hat{\boldsymbol{\sigma}}^2) + 2g_{3i}(\hat{\boldsymbol{\sigma}}^2) \quad (3.27)$$

is an estimator of $\text{MSE}(t_i(\boldsymbol{\sigma}^2, \mathbf{y}))$ with bias of order $o(t^{-1})$.

3.7 Monte Carlo Study

In order to investigate the finite sample properties of the estimation strategies proposed in this chapter, a small Monte Carlo study was undertaken, based on the balanced, two-fold nested error mean model with equal error variances, given by

$$y_{ijk} = \beta + \xi_i + u_{ij} + \epsilon_{ijk}; \quad i = 1, \dots, t; \quad j = 1, \dots, m; \quad k = 1, \dots, n. \quad (3.28)$$

Note that this is a special case of model (3.1) with $p = 1$, $x_{ijk1} = 1$, $\beta_1 = \beta$, $f_{ijk} = 1$, $m_i = m$ and $n_{ij} = n$. The following sections provide the objectives of the study and a description of the experiment, as well as a discussion of the findings.

3.7.1 Objectives of the Study

The objectives of this study were threefold. Firstly, it was of interest to determine the accuracy of the second order approximation to the MSE of the EBLUP, denoted $\text{MSE}[t_i(\hat{\sigma}^2, \mathbf{y})]$, by computing its percentage relative error with respect to the true MSE. This is defined by

$$\begin{aligned} \text{R.E.} [\text{MSE}(t_i(\hat{\sigma}^2, \mathbf{y}))] \\ = 100\{\text{MSE}[t_i(\hat{\sigma}^2, \mathbf{y})] - \text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})]\} / \text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})] \end{aligned} \quad (3.29)$$

where

$$\begin{aligned} \text{MSE}[t_i(\hat{\sigma}^2, \mathbf{y})] &= \sigma_\xi^2(1 - \gamma_i) + \sigma_\xi^2 t^{-1} \gamma_i^{-1} (1 - \gamma_i)^2 \\ &\quad + \sigma_u^{-2} \sigma_\xi^4 (1 - \gamma_i)^3 m \gamma_{ij} \text{var}\{(\sigma_\xi^{-2} \tilde{\sigma}_\xi^2 - \sigma_\epsilon^{-2} \tilde{\sigma}_\epsilon^2) - \gamma_{ij}(\sigma_u^{-2} \tilde{\sigma}_u^2 - \sigma_\epsilon^{-2} \tilde{\sigma}_\epsilon^2)\} \\ &= g_{1i}(\sigma^2) + g_{2i}(\sigma^2) + g_{3i}(\sigma^2) \end{aligned} \quad (3.30)$$

and where $\text{var}(\tilde{\sigma}_i^2)$ and $\text{cov}(\tilde{\sigma}_i^2, \tilde{\sigma}_j^2)$ for $i, j = 1, 2, 3$ are given by (3.21)–(3.26), specializing to the case $p = 1$, $x_{ijk1} = 1$, $\beta_1 = \beta$, $f_{ijk} = 1$, $m_i = m$ and

2

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

$n_{ij} = n$. Note that (3.30) is simply expression (3.19) with the above parameters, and $\text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})]$ is a Monte Carlo estimate of the true MSE.

Secondly, it was of interest to determine the percentage relative bias of $\text{mse}[t_i(\hat{\sigma}^2, \mathbf{y})]$, the estimator of the MSE of the EBLUP, with respect to the true MSE, defined by

$$\begin{aligned} \text{R.B.} [\text{mse}(t_i(\hat{\sigma}^2, \mathbf{y}))] \\ = 100\{E_M[\text{mse}(t_i(\hat{\sigma}^2, \mathbf{y}))] - \text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})]\}/\text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})] \end{aligned} \quad (3.31)$$

where

$$\text{mse}[t_i(\hat{\sigma}^2, \mathbf{y})] = g_{1i}(\hat{\sigma}^2) + g_{2i}(\hat{\sigma}^2) + 2g_{3i}(\hat{\sigma}^2) \quad (3.32)$$

and where E_M denotes Monte Carlo expectation.

Lastly, it was of interest to track the bias of $\text{mse}_1[t_i(\hat{\sigma}^2, \mathbf{y})]$, the customary estimator of the MSE of the EBLUP. This estimator is naive in that it ignores the uncertainty in the estimator $\hat{\sigma}^2$. The relative bias of $\text{mse}_1[t_i(\hat{\sigma}^2, \mathbf{y})]$ with respect to the true MSE is defined by

$$\begin{aligned} \text{R.B.} [\text{mse}_1(t_i(\hat{\sigma}^2, \mathbf{y}))] \\ = 100\{E_M[\text{mse}_1(t_i(\hat{\sigma}^2, \mathbf{y}))] - \text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})]\}/\text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})] \end{aligned} \quad (3.33)$$

where

$$\text{mse}_1[t_i(\hat{\sigma}^2, \mathbf{y})] = g_{1i}(\hat{\sigma}^2) + g_{2i}(\hat{\sigma}^2). \quad (3.34)$$

3.7.2 Description of the Experiment

For this study, model (3.28) was used, with $t = 30$ small areas, $m = 2$ PSUs within each small area, and $n = 2$ SSUs in each PSU within each small area.

Without loss of generality we set $\beta = 0$, since (3.30), (3.32) and (3.34) are invariant to the choice of β . The variance component σ_e^2 was fixed at $\sigma_e^2 = 300$, while σ_u^2 and σ_ξ^2 were each assigned the following values: 15, 30, 60, 150, 300 and 600, so that the ratios σ_u^2/σ_e^2 and σ_ξ^2/σ_e^2 took on all 36 possible combinations of the following values: 0.05, 0.1, 0.2, 0.5, 1.0 and 2.0. In addition, 2000 independent sets of normally distributed pseudo random errors $\{\xi_i; i = 1, \dots, 30\}$, $\{u_{ij}; i = 1, \dots, 30; j = 1, 2\}$ and $\{e_{ijk}; i = 1, \dots, 30; j = 1, 2; k = 1, 2\}$ having means zero and respective variances σ_ξ^2, σ_u^2 and σ_e^2 , were generated. From these generated sets, the observations $\{y_{ijk}; i = 1, \dots, 30; j = 1, 2; k = 1, 2\}$ were obtained using the model $y_{ijk} = \xi_i + u_{ij} + e_{ijk}$.

3.7.3 Results and Discussion

The percentage relative error of the MSE approximation, $\text{MSE}[t_i(\hat{\sigma}^2, \mathbf{y})]$, was computed using (3.29) and then averaged over all small areas, for each combination of σ_u^2/σ_e^2 and σ_ξ^2/σ_e^2 . The averaged values are reported in Table 3.1 ahead. From this table, one can see that the percentages vary widely, ranging from a high of 75.8% to a low of -0.55%. In the lower third of this table (where σ_ξ^2/σ_e^2 is large), the MSE approximation tends to understate the true MSE. However, the percentages are very close to zero, indicating that the second order approximation is in fact close to the true MSE at these points. The general tendency of the table seems to be that the percentage relative error increases as σ_u^2/σ_e^2 increases (which is tantamount to σ_u^2 increasing, since σ_e^2 is fixed for this study), and that the percentage relative error decreases as σ_ξ^2/σ_e^2 increases (which is tantamount to σ_ξ^2 increasing). This observation prompted the hypothesis that the reliability of the results depends on the probabilities of negative variance component estimates being small. To investigate the hypothesis, the Monte Carlo probabilities of negative

variance component estimates for both σ_u^2 and σ_ξ^2 were calculated. These probabilities were tabulated for each of the 36 variance component ratio combinations, using the following formulae:

$$P_M(\tilde{\sigma}_u^2 < 0) = \sum_{i=1}^{2000} I(\tilde{\sigma}_{u,i}^2 < 0)/2000 \quad (3.35)$$

and

$$P_M(\tilde{\sigma}_\xi^2 < 0) = \sum_{i=1}^{2000} I(\tilde{\sigma}_{\xi,i}^2 < 0)/2000 \quad (3.36)$$

where P_M denotes Monte Carlo probability. Note that $I(\tilde{\sigma}_{u,i}^2 < 0) = 1$ if $\tilde{\sigma}_{u,i}^2 < 0$ and $I(\tilde{\sigma}_{u,i}^2 < 0)$ is zero otherwise, where $\tilde{\sigma}_{u,i}^2$ is the Henderson estimate $\hat{\sigma}_u^2$ for the i^{th} trial; $i = 1, \dots, 2000$. $I(\tilde{\sigma}_{\xi,i}^2 < 0)$ is defined analogously. These probabilities are reported in Table 3.2 ahead. Interpretation of this table hinges on the following facts:

- (a) $\tilde{\sigma}_e^2 = \text{MSE}$ where $\text{MSE} = [tm(n-1)]^{-1} \sum_i \sum_j \sum_k (y_{ijk} - y_{ij.})^2$
and $E(\text{MSE}) = \sigma_e^2$.
- (b) $\tilde{\sigma}_u^2 = (\text{MSU} - \text{MSE})/n$ where $\text{MSU} = [t(m-1)]^{-1} \sum_i \sum_j \sum_k (y_{ijk} - y_{i..})^2$
and $E(\text{MSU}) = \sigma_e^2 + n\sigma_u^2$.
- (c) $\tilde{\sigma}_\xi^2 = (\text{MS}\xi - \text{MSU})/nm$ where $\text{MS}\xi = (t-1)^{-1} \sum_i \sum_j \sum_k (y_{i..} - y_{...})^2$
and $E(\text{MS}\xi) = \sigma_e^2 + n\sigma_u^2 + nm\sigma_\xi^2$.

Now as σ_u^2 increases, or equivalently as σ_u^2/σ_e^2 increases, $\tilde{\sigma}_u^2$ should increase since $E[\text{MSU} - \text{MSE}] = n\sigma_u^2$ is independent of σ_ξ^2 , and thus $P_M(\tilde{\sigma}_u^2 < 0)$ should decrease. As σ_u^2 increases, $P_M(\tilde{\sigma}_\xi^2 < 0)$ should also increase due to the increased fluctuations in the difference between the estimated mean squares $\text{MS}\xi$ and MSU . That is, moving left to right along a row of the table $P_M(\tilde{\sigma}_u^2 < 0)$ should decrease while $P_M(\tilde{\sigma}_\xi^2 < 0)$ should increase. As σ_ξ^2 increases, or equivalently as σ_ξ^2/σ_e^2 increases,

$\hat{\sigma}_\xi^2$ should increase since $E[\text{MS}\xi - \text{MSU}] = nm\sigma_\xi^2$, and thus $P_M(\hat{\sigma}_\xi^2 < 0)$ should decrease. As σ_ξ^2 increases, $P_M(\hat{\sigma}_u^2 < 0)$ should remain fixed since MSU and MSE are independent of σ_ξ^2 . That is, moving top to bottom along a column, $P_M(\hat{\sigma}_\xi^2 < 0)$ should decrease while $P_M(\hat{\sigma}_u^2 < 0)$ should remain fixed. The table values bear out these observations. Bearing in mind that the general theory is based on small probabilities of negative variance component estimates, it is clear that the relative errors reported in Table 3.1 may become unreliable when $P_M(\hat{\sigma}_\xi^2 < 0)$ is large. Thus, the theory only seems to hold up well for $\sigma_\xi^2/\sigma_e^2 \geq 0.2$, $\sigma_u^2/\sigma_e^2 \leq 0.2$ and for $\sigma_\xi^2/\sigma_e^2 \geq 0.5$, $\sigma_u^2/\sigma_e^2 > 0.2$.

These observations are also born out in Table 3.3. Here, the percentage relative bias of the estimator of the MSE of the EBLUP with respect to the true MSE was computed using (3.31) and then averaged over all small areas for each combination of σ_u^2/σ_e^2 and σ_ξ^2/σ_e^2 . The averaged values are reported in Table 3.3 ahead. Again, the values vary widely, but are always on the side of overestimation, ranging from as high as 233.20% for $\sigma_u^2/\sigma_e^2 = 2.0$, $\sigma_\xi^2/\sigma_e^2 = 0.05$ to as low as 0.97% for $\sigma_u^2/\sigma_e^2 = 0.5$, $\sigma_\xi^2/\sigma_e^2 = 2.0$. Again, the general tendency seems to be an increase in the percentages for increased σ_u^2/σ_e^2 values, and a decrease in percentages for increased σ_ξ^2/σ_e^2 values. Thus, the values seem to level off and become reasonable for $\sigma_\xi^2/\sigma_e^2 \geq 0.5$, $\sigma_u^2/\sigma_e^2 \leq 0.2$ and for $\sigma_\xi^2/\sigma_e^2 \geq 1.0$, $\sigma_u^2/\sigma_e^2 > 0.2$.

To investigate the behaviour of $\text{mse}_1[t, (\hat{\sigma}^2, \mathbf{y})]$, the estimator often used in practice, its relative bias with respect to the true MSE was calculated. The values were obtained from (3.33), and then averaged over all small areas for each combination of σ_u^2/σ_e^2 and σ_ξ^2/σ_e^2 . The averaged values are reported in Table 3.4. As expected, $\text{mse}_1[t, (\hat{\sigma}^2, \mathbf{y})]$ almost always underestimates the true MSE, where the percentage underestimation ranges from a high of 27.41% to a low of 1.4%. The

exceptions to this, i.e., where $\text{mse}_1[t_i(\hat{\sigma}^2, \mathbf{y})]$ actually overestimates the true MSE seem to occur for high $\sigma_u^2/\text{low } \sigma_\xi^2$ values and for high $\sigma_\xi^2/\text{low } \sigma_u^2$ values.

Observing the tendency of $\text{mse}_1[t_i(\hat{\sigma}^2, \mathbf{y})]$ to underestimate the true MSE, and $\text{mse}[t_i(\hat{\sigma}^2, \mathbf{y})]$ to overestimate the true MSE for small σ_ξ^2/σ_e^2 values prompted the formation of an adaptive estimator of the MSE of the EBLUP, given by,

$$\text{mse}_2[t_i(\hat{\sigma}^2, \mathbf{y})] = \begin{cases} g_{1i}(\hat{\sigma}^2) + g_{2i}(\hat{\sigma}^2) + 2g_{3i}(\hat{\sigma}^2) & \text{if } \hat{\sigma}_\xi^2/\hat{\sigma}_e^2 \geq 0.1 \\ g_{1i}(\hat{\sigma}^2) + g_{2i}(\hat{\sigma}^2) & \text{if } \hat{\sigma}_\xi^2/\hat{\sigma}_e^2 < 0.1, \end{cases} \quad (3.37)$$

where the value 0.1 is chosen in a somewhat ad hoc fashion, after examining Tables 3.3 and 3.4. The relative bias of $\text{mse}_2[t_i(\hat{\sigma}^2, \mathbf{y})]$ with respect to the true MSE was then defined by

$$\begin{aligned} \text{R.B. } [\text{mse}_2[t_i(\hat{\sigma}^2, \mathbf{y})]] \\ = 100\{E_M[\text{mse}_2(t_i(\hat{\sigma}^2, \mathbf{y}))] - \text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})]\}/\text{MSE}_M[t_i(\hat{\sigma}^2, \mathbf{y})]. \end{aligned} \quad (3.38)$$

The values obtained from (3.38) were averaged over all small areas for each combination of σ_u^2/σ_e^2 and σ_ξ^2/σ_e^2 . The averaged values are reported in Table 3.5. Comparing the values in Table 3.5 to those in Table 3.3, one can immediately see a great improvement in bias reduction, resulting from using the estimator $\text{mse}_2[t_i(\hat{\sigma}^2, \mathbf{y})]$ over the estimator $\text{mse}[t_i(\hat{\sigma}^2, \mathbf{y})]$. The reduction is substantial across the entire table, but particularly in the upper half of the table, i.e., for small σ_ξ^2 values. The percentage overestimation is reduced from a high of 233.2% in Table 3.3 to one of 84.17% in Table 3.5. The low value in percentage overestimation in Table 3.5 is 0.43%. Only one point still shows $\text{mse}_2[t_i(\hat{\sigma}^2, \mathbf{y})]$ underestimating the true MSE by 1.45%, but this is so close to zero as to be negligible. Thus, Table 3.5 seems to behave well in the regions $\sigma_\xi^2/\sigma_e^2 \geq 0.1$, $\sigma_u^2/\sigma_e^2 \leq 0.5$ and $\sigma_\xi^2/\sigma_e^2 \geq 0.5$, $\sigma_u^2/\sigma_e^2 > 0.5$. Although there is no theoretical justification for using

$\text{mse}_2(t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y}))$ over $\text{mse}(t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y}))$, the former appears to behave well in practice, balancing the overestimation of the true MSE coming from $\text{mse}[t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y})]$ with the underestimation of the true MSE coming from $\text{mse}_1[t_i(\hat{\boldsymbol{\sigma}}^2, \mathbf{y})]$.

In conclusion, our study indicates that the asymptotic theory holds up well if the probability of negative $\hat{\sigma}_{\xi}^2$ is small. This occurs roughly on the lower diagonal portion of each table, i.e., where $\sigma_{\xi}^2/\sigma_{\epsilon}^2 \geq \sigma_u^2/\sigma_{\epsilon}^2$. From a practical standpoint, this is the region of interest from each table, since we expect the variation between small areas to be greater than the variation between PSUs.

Table 3.1 Percentage Relative Error of $\text{MSE}[t_1(\hat{\sigma}^2, y)]$, the Second Order Approximation to the MSE of the EBLUP, With Respect to the True MSE

	σ_u^2/σ_e^2					
	.05	0.1	0.2	0.5	1.0	2.0
0.5	34.20	35.58	38.28	46.96	60.11	75.80
0.1	15.49	16.64	19.09	26.97	38.96	55.36
σ_ξ^2/σ_e^2 0.2	4.33	4.87	6.22	11.25	20.21	34.07
0.5	-0.03	0.04	0.13	1.08	4.32	11.85
1.0	-0.37	-0.37	-0.44	-0.44	0.28	2.93
2.0	-0.41	-0.40	-0.43	-0.55	-0.51	0.05

Table 3.2 Monte Carlo Probabilities of Negative Variance Component Estimates for $\hat{\sigma}_\xi^2$ and $\hat{\sigma}_u^2$ where Each Pair Represents $(P_M(\hat{\sigma}_\xi^2 < 0), P_M(\hat{\sigma}_u^2 < 0))$

	σ_u^2/σ_e^2					
	.05	0.1	0.2	0.5	1.0	2.0
0.5	(.34, .38)	(.35, .28)	(.36, .15)	(.41, .02)	(.44, .00)	(.47, .00)
0.1	(.21, .38)	(.22, .28)	(.26, .15)	(.32, .02)	(.38, .00)	(.43, .00)
σ_ξ^2/σ_e^2 0.2	(.08, .38)	(.09, .28)	(.12, .15)	(.19, .02)	(.27, .00)	(.35, .00)
0.5	(.00, .38)	(.01, .28)	(.01, .15)	(.04, .02)	(.09, .00)	(.19, .00)
1.0	(.00, .38)	(.00, .28)	(.00, .15)	(.00, .02)	(.02, .00)	(.06, .00)
2.0	(.00, .38)	(.00, .28)	(.00, .15)	(.00, .02)	(.00, .00)	(.01, .00)

Table 3.3 Percentage Relative Bias of $mse[t_1(\hat{\sigma}^2, y)]$, the Estimator of the MSE of the EBLUP, With Respect to the True MSE

		σ_u^2/σ_e^2					
		.05	0.1	0.2	0.5	1.0	2.0
σ_ξ^2/σ_e^2	.05	101.80	103.60	110.10	137.70	180.50	233.20
	0.1	46.75	47.60	52.22	73.97	110.30	163.10
	0.2	18.49	17.59	18.59	29.81	53.53	94.19
	0.5	8.63	6.52	4.51	5.29	12.51	30.76
	1.0	7.77	5.39	2.90	1.64	3.28	9.15
	2.0	7.65	5.16	2.57	0.97	1.29	2.63

Table 3.4 Percentage Relative Bias of $mse_1[t_1(\hat{\sigma}^2, y)]$, the Naive Estimator, With Respect to the True MSE

		σ_u^2/σ_e^2					
		.05	0.1	0.2	0.5	1.0	2.0
σ_ξ^2/σ_e^2	.05	-10.17	-10.80	-10.91	-6.40	4.22	18.20
	0.1	-20.75	-22.17	-23.27	-20.86	-13.01	0.01
	0.2	-20.12	-22.58	-25.44	-27.41	-24.34	-16.48
	0.5	-8.81	-11.70	-15.70	-21.77	-25.62	-26.76
	1.0	-1.40	-4.20	-7.79	-12.97	-17.81	-23.67
	2.0	2.93	0.23	-2.93	-6.64	-9.91	-15.32

**Table 3.5 Percentage Relative Bias of $\text{mse}_2[t_1(\hat{\sigma}^2, y)]$, the Adaptive Estimator,
With Respect to the True MSE**

		σ_u^2/σ_e^2					
		.05	0.1	0.2	0.5	1.0	2.0
σ_ξ^2/σ_e^2	.05	13.34	13.30	15.31	28.35	52.29	84.17
	0.1	2.10	1.08	1.15	10.20	27.91	55.86
	0.2	3.23	0.43	-1.45	0.92	10.80	28.77
	0.5	7.54	5.26	2.55	0.60	2.07	8.78
	1.0	7.75	5.37	2.86	1.33	1.74	3.47
	2.0	7.65	5.16	2.57	0.97	1.19	1.90

CHAPTER 4

Jackknife Variance Estimation and Robust Confidence Intervals on Functions of Variance Components

In this chapter, we confine ourselves to a special case of the two-fold nested error regression model used in Chapter 3, i.e., an intercept model having one auxiliary variable and constant error variances. In Section 4.1, the general jackknife method is reviewed. In Section 4.2, the Henderson estimates of the variance components obtained in Chapter 3, are modified along the lines of Arvesen (1969). These modified estimators have the desired property that they can be written as weighted sums of estimators formed by deleting each small area (group) in turn. Jackknife estimators of the variances of, and the covariances between, the Arvesen estimators are derived. A detailed proof shows these jackknife estimators to be consistent for large t , the number of groups. Section 4.3 gives jackknife estimators of bounded, nonlinear functions of $\sigma^2 = (\sigma_\epsilon^2, \sigma_u^2, \sigma_c^2)'$, as well as the corresponding jackknife variance estimators. Finally, in Section 4.4, the findings of a small Monte Carlo study using a special case of the above model are reported. In this study, the robustness of confidence intervals of ratios of variance components, based on jackknife estimators are investigated, as well as the bias of the jackknife variance estimators.

4.1 The Jackknife Method

We begin by outlining the general jackknife procedure, as proposed by Quenouille (1949, 1956). Suppose that $\hat{\theta}_n$ denotes an estimator of an unknown parameter θ , based on n i.i.d. observations. Divide the n observations, at random into m groups, each consisting of k observations, i.e., $n = mk$. Let $\hat{\theta}_{m-1}(-i)$;

$i = 1, \dots, m$ be the estimator obtained after deletion of the i^{th} group of observations. Define the pseudo-values as $Q_i = m\hat{\theta}_m - (m-1)\hat{\theta}_{m-1}(-i)$; $i = 1, \dots, m$. The jackknife estimator $\hat{\theta}_J$ of θ is then given by the average of the pseudo-values, i.e., $\hat{\theta}_J = \bar{Q} = m^{-1} \sum_{i=1}^m Q_i$. The jackknife estimator has the property that if the original estimator $\hat{\theta}_n$ of θ has a bias of order $1/n$, then $\hat{\theta}_J$ has a reduced bias of order $1/n^2$. Tukey (1958) proposed $v_J = (m-1)^{-1}m^{-1} \sum_{i=1}^m (Q_i - \bar{Q})^2$ as the jackknife variance estimator of θ_J . He also suggested the use of $t^* = (\hat{\theta}_J - \theta)v_J^{-1/2}$ as a pivotal statistic for obtaining confidence intervals on θ , by treating it as a t -variate with $m-1$ degrees of freedom.

However, when the observations are not i.i.d., such as in the case of the model used in this chapter, modified methods must be used. Miller (1974), Hinkley (1977) and Wu (1984) considered the standard regression model $y = X\beta + e$ where $X = (x_1, \dots, x_n)'$, x_i' is a $1 \times p$ vector and $E(e) = 0$. If $\hat{\beta}$ is the OLS estimator of β given by $\hat{\beta} = (X'X)^{-1}X'y$, then for the special case of 'delete-one' jackknife, i.e., $k = 1$ and $m = n$, $\hat{\beta}$ can be written as

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

where $\hat{\beta}(-i)$ is the estimator obtained upon recomputing $\hat{\beta}$ after deleting the i^{th} observation, and $w_i = 1 - x_i'(X'X)^{-1}x_i$. Hinkley (1977) constructed variance estimators of $\theta = g(\beta)$ based on weighted pseudo-values, given by

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

where $\hat{\theta}(-i) = g(\hat{\beta}(-i))$. A jackknife variance estimator is then given by

$$\begin{aligned} v_{Jn} &= (n(n-p))^{-1} \sum_{i=1}^n (Q_i - \bar{Q})^2 \\ &= n(n-p)^{-1} \sum_{i=1}^n w_i^2 (\hat{\theta} - \hat{\theta}(-i))^2 \end{aligned}$$

where

$$\bar{Q} = n^{-1} \sum_{i=1}^n Q_i.$$

4.2 Jackknife Estimation of Variances and Covariances of Variance Component Estimators

For the sake of simplicity, we confine ourselves to a special case of the model considered in Chapter 3. That is, we focus on the two-fold nested error regression model with one auxiliary variable and constant error variances, i.e., $p = 2$, $r_{ijk1} = 1$, $r_{ijk2} = x_{ijk}$ and $f_{ijk} = 1 \forall (i, j, k)$. The model can be written as

$$y_{ijk} = \alpha + \beta x_{ijk} + \xi_i + u_{ij} + e_{ijk}; \quad i = 1, \dots, t; \quad j = 1, \dots, m_i; \quad k = 1, \dots, n_{ij}. \quad (4.1)$$

Before discussing the jackknife procedure, we modify the estimators $\hat{\sigma}_e^2, \hat{\sigma}_u^2$ and $\hat{\sigma}_\xi^2$, given by (3.10), (3.11) and (3.13), in order to facilitate the construction of weighted pseudo-values, as motivated by Hinkley (1977).

We begin by modifying $\hat{\sigma}_e^2$ along the lines of Arvesen (1969). We perform a regression of $y_{ijk} - \bar{y}_{ij}$ on $x_{ijk} - \bar{x}_{ij}$, where $\bar{y}_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} y_{ijk}$ and $\bar{x}_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} x_{ijk}$, and form the corresponding residuals

$$e_{ijk}^* = (y_{ijk} - \bar{y}_{ij}) - (x_{ijk} - \bar{x}_{ij})/\beta^*$$

where

$$\beta^* = [\sum_{i=1}^t \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{ij})^2]^{-1} \sum_{i=1}^t \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})(x_{ijk} - \bar{x}_{ij}).$$

We note that

$$E[\sum_j \sum_k e_{ijk}^{*2}] = \sigma_e^2 \nu_{11}$$

where

$$\nu_{11} = \sum_j n_{ij} - m_i - [\sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2]^{-1} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2.$$

Using this fact, we form an unbiased estimator of σ_e^2 , denoted by σ_e^{*2} , as follows:

$$\sigma_e^{*2} = (\sum_i \nu_{1i})^{-1} \sum_i \sum_j \sum_k c_{ijk}^2 \quad (4.2)$$

where

$$\sum_i \nu_{1i} = n - \sum_i m_i - 1.$$

Note that $\sigma_e^{*2} = \tilde{\sigma}_e^2$, but we denote the Arvesen estimator of σ_e^2 as σ_e^{*2} for consistency of notation.

We modify $\tilde{\sigma}_u^2$ by performing a regression of $y_{ijk} - \bar{y}_{i..}$ on $x_{ijk} - \bar{x}_{i..}$ where $y_{i..} = n_{i.}^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} y_{ijk}$ and $\bar{x}_{i..} = n_{i.}^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} x_{ijk}$, and form the corresponding residuals

$$u_{ijk}^* = (y_{ijk} - \bar{y}_{i..}) - (x_{ijk} - \bar{x}_{i..})\beta^{**}$$

where

$$\beta^{**} = [\sum_i \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{i..})^2]^{-1} \sum_i \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{i..})(x_{ijk} - \bar{x}_{i..}).$$

We note that

$$E[\sum_j \sum_k u_{ijk}^{*2}] = \nu_{2i} \sigma_e^2 + \tau_{1i} \sigma_u^2$$

where

$$\nu_{2i} = n_{i.} - 1 - [\sum_j \sum_k \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{i..})^2]^{-1} \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2$$

and where

$$\begin{aligned} \tau_{1i} = & n_{i.} - n_{i.}^{-1} \sum_j n_{ij}^2 - 2[\sum_j \sum_k \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{i..})^2]^{-1} \sum_j n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{i..})^2 \\ & + [\sum_j \sum_k \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{i..})^2]^{-2} \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \sum_j \sum_j n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{i..})^2. \end{aligned}$$

Using this fact, we form an unbiased estimator of σ_u^2 , denoted by σ_u^{*2} , as follows:

$$\sigma_u^{*2} = (\sum_i \tau_{1i})^{-1} [\sum_i \sum_j \sum_k u_{ijk}^{*2} - \sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k c_{ijk}^2] \quad (4.3)$$

where

$$\sum_i r_{1i} = n - \sum_i n_i^{-1} \sum_j n_{ij}^2 - [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2]^{-1} \sum_i \sum_j n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{i..})^2.$$

Finally, we modify $\hat{\sigma}_\xi^2$ by performing a regression of $y_{ijk} - \bar{y}_{...}$ on $x_{ijk} - \bar{x}_{...}$ where $y_{...} = n^{-1} \sum_i \sum_j \sum_k y_{ijk}$ and $\bar{x}_{...} = n^{-1} \sum_i \sum_j \sum_k x_{ijk}$, and form the corresponding residuals

$$\xi_{ijk}^* = (y_{ijk} - \bar{y}_{...}) - (x_{ijk} - \bar{x}_{...})\beta^{***}$$

where

$$\beta^{***} = [\sum_{i=1}^t \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{...})^2]^{-1} \sum_{i=1}^t \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{...})(x_{ijk} - \bar{x}_{...}).$$

We note that

$$E[\sum_j \sum_k \xi_{ijk}^{*2}] = \nu_{3i} \sigma_c^2 + r_{2i} \sigma_u^2 + \delta_{1i} \sigma_\xi^2$$

where

$$\begin{aligned} \nu_{3i} &= n_{i.} - n^{-1} n_{i.} - [\sum_j \sum_k \sum_l (x_{ijkl} - \bar{x}_{...})^2]^{-1} \sum_j \sum_k (x_{ijk} - \bar{x}_{...})^2, \\ r_{2i} &= n_{i.} + n^{-2} n_{i.} \sum_j \sum_k n_{ij}^2 - 2n^{-1} \sum_j n_{ij}^2 - 2[\sum_j \sum_k \sum_l (x_{ijkl} - \bar{x}_{...})^2]^{-1} \sum_j n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{...})^2 \\ &\quad + 2[\sum_j \sum_k \sum_l (x_{ijkl} - \bar{x}_{...})^2]^{-1} n_{i.} (\bar{x}_{i..} - \bar{x}_{...}) n^{-1} \sum_j \sum_k n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{...}) \\ &\quad + [\sum_j \sum_k \sum_l (x_{ijkl} - \bar{x}_{...})^2]^{-2} \sum_j \sum_k (x_{ijk} - \bar{x}_{...})^2 \sum_j \sum_k n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{...})^2 \end{aligned}$$

and where

$$\begin{aligned} \delta_{1i} &= n_{i.} + n^{-2} n_{i.} \sum_j n_{ij}^2 - 2n^{-1} n_{i.}^2 - 2[\sum_j \sum_k \sum_l (x_{ijkl} - \bar{x}_{...})^2]^{-1} n_{i.}^2 (\bar{x}_{i..} - \bar{x}_{...})^2 \\ &\quad + [\sum_j \sum_k \sum_l (x_{ijkl} - \bar{x}_{...})^2]^{-2} \sum_j \sum_k (x_{ijk} - \bar{x}_{...})^2 \sum_j n_{ij}^2 (\bar{x}_{i..} - \bar{x}_{...})^2 \\ &\quad + 2[\sum_j \sum_k \sum_l (x_{ijkl} - \bar{x}_{...})^2]^{-1} n_{i.} (\bar{x}_{i..} - \bar{x}_{...}) n^{-1} \sum_j n_{ij}^2 (\bar{x}_{i..} - \bar{x}_{...}). \end{aligned}$$

Using this fact, we form an unbiased estimator of σ_ξ^2 , denoted by σ_ξ^{*2} , as follows:

$$\begin{aligned}\sigma_\xi^{*2} = & (\sum_i \delta_{1i})^{-1} [\sum_i \sum_j \sum_k \xi_{ijk}^{*2} - \sum_i \tau_{1i}^{-1} \tau_{2i} \sum_j \sum_k u_{ijk}^{*2} \\ & - \sum_i (\tau_{1i} \nu_{1i})^{-1} (\nu_{3i} \tau_{1i} - \tau_{2i} \nu_{2i}) \sum_j \sum_k c_{ijk}^{*2}] \quad (4.4)\end{aligned}$$

where

$$\sum_i \delta_{1i} = n - n^{-1} \sum_i n_i^2 - [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{...})^2]^{-1} \sum_i n_i^2 (\bar{x}_{i...} - \bar{x}_{...})^2.$$

The jackknife procedure considered here is based on deleting the residuals $\{c_{ijk}^*, u_{ijk}^*, \xi_{ijk}^*; j = 1, \dots, m_i; k = 1, \dots, n_{ij}\}$ corresponding to the i^{th} small area (group), and then recomputing the estimators of c_e^2, σ_u^2 and σ_ξ^2 for $i = 1, \dots, t$. The resulting unbiased estimators are given by

$$\sigma_e^{*2}(-i) = (\sum_i \nu_{1i} - \nu_{1i})^{-1} [\sum_i \sum_j \sum_k c_{ijk}^{*2} - \sum_j \sum_k c_{ijk}^{*2}], \quad (4.5)$$

$$\begin{aligned}\sigma_u^{*2}(-i) = & (\sum_i \tau_{1i} - \tau_{1i})^{-1} [(\sum_i \sum_j \sum_k u_{ijk}^{*2} - \sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k c_{ijk}^{*2}) \\ & - (\sum_j \sum_k u_{ijk}^{*2} - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k c_{ijk}^{*2})] \quad (4.6)\end{aligned}$$

and

$$\begin{aligned}\sigma_\xi^{*2}(-i) = & (\sum_i \delta_{1i} - \delta_{1i})^{-1} \{[\sum_i \sum_j \sum_k \xi_{ijk}^{*2} - \sum_i \tau_{1i}^{-1} \tau_{2i} \sum_j \sum_k u_{ijk}^{*2} \\ & - \sum_i (\tau_{1i} \nu_{1i})^{-1} (\nu_{3i} \tau_{1i} - \tau_{2i} \nu_{2i}) \sum_j \sum_k c_{ijk}^{*2}] - [\sum_j \sum_k \xi_{ijk}^{*2} \\ & - \tau_{1i}^{-1} \tau_{2i} \sum_j \sum_k u_{ijk}^{*2} - (\tau_{1i} \nu_{1i})^{-1} (\nu_{3i} \tau_{1i} - \tau_{2i} \nu_{2i}) \sum_j \sum_k c_{ijk}^{*2}]\}. \quad (4.7)\end{aligned}$$

It is interesting to note that the estimators $\sigma_e^{*2}, \sigma_u^{*2}$ and σ_ξ^{*2} can be expressed as weighted sums of $\sigma_e^{*2}(-i), \sigma_u^{*2}(-i)$ and $\sigma_\xi^{*2}(-i)$ respectively, as follows:

$$\sigma_e^{*2} = (t-1)^{-1} \sum_i a_i \sigma_e^{*2}(-i), \quad (4.8)$$

$$\sigma_u^{*2} = (t-1)^{-1} \sum_i b_i \sigma_u^{*2}(-i) \quad (4.9)$$

and

$$\sigma_{\xi}^{*2} = (t-1)^{-1} \sum_i c_i \sigma_{\xi}^{*2}(-i) \quad (4.10)$$

where

$$a_i = 1 - (\sum_i \nu_{1i})^{-1} \nu_{1i},$$

$$b_i = 1 - (\sum_i \tau_{1i})^{-1} \tau_{1i}$$

and where

$$c_i = 1 - (\sum_i \delta_{1i})^{-1} \delta_{1i}.$$

We define the weighted pseudo-values along the lines of Hinkley (1977) as follows:

$$Q_i(c) = \sigma_c^{*2} + t a_i (\sigma_c^{*2} - \sigma_c^{*2}(-i)), \quad (4.11)$$

$$Q_i(u) = \sigma_u^{*2} + t b_i (\sigma_u^{*2} - \sigma_u^{*2}(-i)), \quad (4.12)$$

and

$$Q_i(\xi) = \sigma_{\xi}^{*2} + t c_i (\sigma_{\xi}^{*2} - \sigma_{\xi}^{*2}(-i)). \quad (4.13)$$

Jackknife estimators are formed by taking the mean of the weighted pseudo-values, i.e.,

$$\bar{Q}(c) = t^{-1} \sum_i Q_i(c), \quad (4.14)$$

$$\bar{Q}(u) = t^{-1} \sum_i Q_i(u), \quad (4.15)$$

and

$$\bar{Q}(\xi) = t^{-1} \sum_i Q_i(\xi). \quad (4.16)$$

The estimators (4.14)-(4.16) satisfy the reproducing property, i.e., $\bar{Q}(c) = \sigma_c^{*2}$, $\bar{Q}(u) = \sigma_u^{*2}$ and $\bar{Q}(\xi) = \sigma_{\xi}^{*2}$. Finally, the following are proposed as estimators of the variances of, and the covariances between, the variance component estimators:

$$v_J(\sigma_c^{*2}) = t^{-1}(t-1)^{-1} \sum_i [Q_i(c) - \bar{Q}(c)]^2 \quad (4.17)$$

$$= t(t-1)^{-1} \sum_i a_i^2 (\sigma_c^{*2} - \sigma_c^{*2}(-i))^2,$$

$$v_J(\sigma_u^{*2}) = t^{-1}(t-1)^{-1} \sum_i [Q_i(u) - \bar{Q}(u)]^2 \quad (4.18)$$

$$= t(t-1)^{-1} \sum_i b_i^2 (\sigma_u^{*2} - \sigma_u^{*2}(-i))^2,$$

$$v_J(\sigma_{\xi}^{*2}) = t^{-1}(t-1)^{-1} \sum_i [Q_i(\xi) - \bar{Q}(\xi)]^2 \quad (4.19)$$

$$= t(t-1)^{-1} \sum_i c_i^2 (\sigma_{\xi}^{*2} - \sigma_{\xi}^{*2}(-i))^2,$$

$$\begin{aligned}\text{cov}_J(\sigma_e^{*2}, \sigma_u^{*2}) &= t^{-1}(t-1)^{-1} \sum_i [Q_i(e) - \bar{Q}(e)][Q_i(u) - \bar{Q}(u)] \\ &= t(t-1)^{-1} \sum_i a_i b_i [\sigma_e^{*2} - \sigma_e^{*2}(-i)][\sigma_u^{*2} - \sigma_u^{*2}(-i)],\end{aligned}\quad (4.20)$$

$$\begin{aligned}\text{cov}_J(\sigma_e^{*2}, \sigma_\xi^{*2}) &= t^{-1}(t-1)^{-1} \sum_i [Q_i(e) - \bar{Q}(e)][Q_i(\xi) - \bar{Q}(\xi)] \\ &= t(t-1)^{-1} \sum_i a_i c_i [\sigma_e^{*2} - \sigma_e^{*2}(-i)][\sigma_\xi^{*2} - \sigma_\xi^{*2}(-i)],\end{aligned}\quad (4.21)$$

and

$$\begin{aligned}\text{cov}_J(\sigma_u^{*2}, \sigma_\xi^{*2}) &= t^{-1}(t-1)^{-1} \sum_i [Q_i(u) - \bar{Q}(u)][Q_i(\xi) - \bar{Q}(\xi)] \\ &= t(t-1)^{-1} \sum_i b_i c_i [\sigma_u^{*2} - \sigma_u^{*2}(-i)][\sigma_\xi^{*2} - \sigma_\xi^{*2}(-i)].\end{aligned}\quad (4.22)$$

Remark 4.1 The estimators (4.20)-(4.22) are motivated by the fact that

$$\text{cov}(\sigma_i^{*2}, \sigma_j^{*2}) = 1/2[\text{var}(\sigma_i^{*2} + \sigma_j^{*2}) - \text{var}(\sigma_i^{*2}) - \text{var}(\sigma_j^{*2})]$$

for $i \neq j$ and $i, j = 1, 2, 3$.

Remark 4.2 Since the Arvesen estimators are simply modified Henderson estimators, one possible application of the above jackknife method could be to replace $\hat{\sigma}_i^2$; $i = 1, 2, 3$ with σ_i^{*2} ; $i = 1, 2, 3$ in $g_{31}(\sigma^2)$, given by (3.18). The expressions for $\text{var}(\sigma_i^{*2})$; $i = 1, 2, 3$ and $\text{cov}(\sigma_i^{*2}, \sigma_j^{*2})$; $i \neq j$; $i, j = 1, 2, 3$ needed in (3.18) could then be approximated by the consistent jackknife estimators, given by (4.17)-(4.22). This application would avoid the use of the normality-based variance and covariance estimators, given by (3.21)-(3.26) which are computationally very cumbersome. Note, however, that the normality assumption has not been dispensed with since it is still required for the formation of the second order approximation, even though it is not needed for the development of the jackknife estimators.

The consistency of the jackknife estimators, given by (4.17) (4.19), is established by the following theorem.

Theorem 4.1 Assume the following regularity conditions on the model

$$y_{ijk} = \alpha + \beta x_{ijk} + \xi_i + u_{ij} + c_{ijk}; i = 1, \dots, t; j = 1, \dots, m_i; k = 1, \dots, n_{ij};$$

$$(R1) \quad \inf_{1 \leq i \leq t} m_i \geq 2, \\ \sup_{1 \leq i \leq t} m_i \leq m_u, \\ \inf_{1 \leq i \leq t} \inf_{1 \leq j \leq m_i} n_{ij} \geq 2$$

and

$$\sup_{1 \leq i \leq t} \sup_{1 \leq j \leq m_i} n_{ij} \leq n_u. \\ (R2) \quad \sup_{1 \leq i \leq t} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 < \infty. \\ (R3) \quad 0 < t^{-1} \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 < \infty. \\ (R4) \quad \sup_{1 \leq i \leq t} \sup_{1 \leq j \leq m_i} \sup_{1 \leq k \leq n_{ij}} E|c_{ijk}|^{4+\delta} < \infty \text{ for some } 0 < \delta < 1. \\ (R5) \quad \sup_{1 \leq i \leq t} \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 < \infty. \\ (R6) \quad 0 < t^{-1} \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 < \infty. \\ (R7) \quad \sup_{1 \leq i \leq t} \sum_j n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{i..})^2 < \infty. \\ (R8) \quad 0 < t^{-1} \sum_i \sum_j n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{i..})^2 < \infty. \\ (R9) \quad \sup_{1 \leq i \leq t} \sup_{1 \leq j \leq m_i} E|u_{ij}|^{4+\delta} < \infty \text{ for some } 0 < \delta < 1.$$

Then

$$(a) \quad t[v_J(\sigma_\epsilon^{*2}) - \text{var}(\sigma_\epsilon^{*2})] = o_p(1)$$

and

$$(b) \quad t[v_J(\sigma_u^{*2}) - \text{var}(\sigma_u^{*2})] = o_p(1).$$

It is also true that

$$(c) \quad t[v_J(\sigma_\xi^{*2}) - \text{var}(\sigma_\xi^{*2})] = o_p(1),$$

although additional regularity conditions are needed to prove (c). Since the proof of (c) is omitted due to its similarity in nature to the proofs of (a) and (b), the

statement of those additional regularity conditions is also omitted. The expressions for σ_e^{*2} , σ_u^{*2} and σ_ξ^{*2} are given by (4.2)-(4.4), whereas the expressions for $v_j(\sigma_e^{*2})$, $v_j(\sigma_u^{*2})$ and $v_j(\sigma_\xi^{*2})$ are given by (4.17)-(4.19).

Proof (a) First, we define the following terms:

$$\bar{\sigma}_e^2 = (\sum_i \nu_{1i})^{-1} \sum_i \sum_j \sum_k \bar{c}_{ijk}^2, \quad (4.23)$$

$$\bar{\sigma}_e^2(-i) = (\sum_i \nu_{1i} - \nu_{1i})^{-1} (\sum_i \sum_j \sum_k \bar{c}_{ijk}^2 - \sum_j \sum_k c_{ijk}^2) \quad (4.24)$$

and

$$v_j(\bar{\sigma}_e^2) = t(t-1)^{-1} \sum_i a_i^2 [\bar{\sigma}_e^2(-i) - \bar{\sigma}_e^2]^2 \quad (4.25)$$

where $\bar{c}_{ijk} = (y_{ijk} - \bar{y}_{ij}) - (x_{ijk} - \bar{x}_{ij})\beta = c_{ijk} - \bar{c}_{ij}$, and where $c_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} c_{ijk}$. 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 σ_e^{*2} , $\sigma_e^{*2}(-i)$ and $v_j(\sigma_e^{*2})$ by its expected value, β .

Now, we can write

$$\begin{aligned} t[v_j(\sigma_e^{*2}) - \text{var}(\sigma_e^{*2})] &= t[v_j(\sigma_e^{*2}) - v_j(\bar{\sigma}_e^2)] \\ &\quad + t[v_j(\bar{\sigma}_e^2) - \text{var}(\bar{\sigma}_e^2)] + t[\text{var}(\bar{\sigma}_e^2) - \text{var}(\sigma_e^{*2})] = o_p(1) \end{aligned}$$

if the three terms on the right hand side of the above equation each converge to zero in probability. We have that $t[v_j(\sigma_e^{*2}) - v_j(\bar{\sigma}_e^2)] = o_p(1)$ by Lemma 4.2(c) ahead. Addressing the second term, we have that $t[v_j(\bar{\sigma}_e^2) - \text{var}(\sigma_e^{*2})] = o_p(1)$ by Lemma 4.1 ahead. Finally $t[\text{var}(\bar{\sigma}_e^2) - \text{var}(\sigma_e^{*2})] = o_p(1)$ by Lemma 4.2(b) ahead.

(b) We require the following definitions:

$$\bar{\sigma}_u^2 = (\sum_i \tau_{1i})^{-1} [\sum_i \sum_j \sum_k \bar{u}_{ijk}^2 - \sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k \bar{c}_{ijk}^2], \quad (4.26)$$

$$\begin{aligned} \bar{\sigma}_u^2(-i) &= (\sum_i \tau_{1i} - \tau_{1i})^{-1} [(\sum_i \sum_j \sum_k \bar{u}_{ijk}^2 - \sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k c_{ijk}^2) \\ &\quad - (\sum_j \sum_k \bar{u}_{ijk}^2 - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k \bar{c}_{ijk}^2)] \end{aligned} \quad (4.27)$$

and

$$v_j(\bar{\sigma}_u^2) = t(t-1)^{-1} \sum_i b_i^2 [\bar{\sigma}_u^2(-i) - \bar{\sigma}_u^2]^2 \quad (4.28)$$

where $u_{ijk} = (y_{ijk} - \bar{y}_{i..}) - (x_{ijk} - \bar{x}_{i..})\beta = (u_{ij} - \bar{u}_{i..}) + (e_{ijk} - \bar{e}_{i..})$ and where $\bar{e}_{i..} = n_i^{-1} \sum_j \sum_k e_{ijk}$ whereas $\bar{u}_{i..} = n_i^{-1} \sum_j n_{ij} u_{ij}$. Note that $\bar{\sigma}_u^2, \bar{\sigma}_u^2(-i)$ and $v_j(\bar{\sigma}_u^2)$ are obtained respectively by replacing β^{**} in $\sigma_u^{*2}, \sigma_u^{*2}(-i)$ and $v_j(\sigma_u^{*2})$ by its expected value, β .

Now, we can write

$$\begin{aligned} t\{v_j(\sigma_u^{*2}) - \text{var}(\sigma_u^{*2})\} &= t\{v_j(\sigma_u^{*2}) - v_j(\bar{\sigma}_u^2)\} \\ &\quad + t\{v_j(\bar{\sigma}_u^2) - \text{var}(\bar{\sigma}_u^2)\} + t\{\text{var}(\bar{\sigma}_u^2) - \text{var}(\sigma_u^{*2})\} = o_p(1) \end{aligned}$$

if the three terms on the right hand side of the above equation each converge to zero in probability. We have that $t\{v_j(\sigma_u^{*2}) - v_j(\bar{\sigma}_u^2)\} = o_p(1)$ by Lemma 4.6(c) ahead. Addressing the second term, we have that $t\{v_j(\bar{\sigma}_u^2) - \text{var}(\bar{\sigma}_u^2)\} = o_p(1)$ by Lemma 4.5 ahead. Finally $t\{\text{var}(\bar{\sigma}_u^2) - \text{var}(\sigma_u^{*2})\} = o_p(1)$ by Lemma 4.6(b) ahead. ■

Lemma 4.1 Assuming the conditions of Theorem 4.1, we have that $t\{v_j(\sigma_e^2) - \text{var}(\sigma_e^2)\} = o_p(1)$.

Proof We can write $v_j(\bar{\sigma}_e^2)$ as

$$\begin{aligned} v_j(\sigma_e^2) &= t(t-1)^{-1} (\sum_i \nu_{1i})^{-2} \sum_i (\nu_{1i} \bar{\sigma}_e^2 - \sum_j \sum_k \bar{e}_{ijk}^2)^2 \\ &= t(t-1)^{-1} (\sum_i \nu_{1i})^{-2} [\bar{\sigma}_e^4 \sum_i \nu_{1i}^2 + \sum_i (\sum_j \sum_k \bar{e}_{ijk}^2)^2 - 2\bar{\sigma}_e^2 \sum_i \nu_{1i} \sum_j \sum_k \bar{e}_{ijk}^2]. \end{aligned}$$

Now, we establish that

$$(\sum_i \nu_{1i})^{-1} = O(t^{-1}) \quad (4.29)$$

as follows: $\sum_i \sum_j n_{ij} \geq 2 \sum_i m_i$ using regularity condition (R1) and so $\sum_i \nu_{1i} = \sum_i \sum_j n_{ij} - \sum_i m_i - 1 \geq 2 \sum_i m_i - \sum_i m_i - 1 = \sum_i m_i - 1 \geq 2t - 1$ using regularity condition (R1) again. Thus

$$t(\sum_i \nu_{1i})^{-1} = t(\sum_i \sum_j n_{ij} - \sum_i m_i - 1)^{-1} \leq t(2t - 1)^{-1} \leq 1.$$

Similarly, we can show that

$$\sum_i \nu_{1i} = O(t). \quad (4.30)$$

Upon multiplying the above expression for $v_j(\hat{\sigma}_e^2)$ by $(\sum_i \nu_{1i}/t)^2$, we see that to prove the lemma, it is sufficient to show that

$$\begin{aligned} & t^{-1} [\sum_i (\sum_j \sum_k \tilde{e}_{ijk}^2)^2 - 2\hat{\sigma}_e^2 \sum_i \nu_{1i} \sum_j \sum_k \tilde{e}_{ijk}^2 + \hat{\sigma}_e^4 \sum_i \nu_{1i}^2 \\ & - (\sum_i \nu_{1i})^2 \text{var}(\hat{\sigma}_e^2)] = o_p(1). \end{aligned} \quad (4.31)$$

To proceed further, we require the following three assertions:

$$t^{-1} [\sum_i (\sum_j \sum_k \tilde{e}_{ijk}^2)^2 - (\sum_i \nu_{1i})^2 \text{var}(\sigma_e^2) - \sigma_e^4 \sum_i \nu_{1i}^2] = o_p(1), \quad (4.32)$$

$$\hat{\sigma}_e^2 - \sigma_e^2 = o_p(1) \quad (4.33)$$

and

$$t^{-1} [\sum_i \nu_{1i} \sum_j \sum_k \tilde{e}_{ijk}^2 - \sigma_e^2 \sum_i \nu_{1i}^2] = o_p(1). \quad (4.34)$$

The last of these can be written equivalently as

$$t^{-1} [\sum_i \nu_{1i} \sum_j \sum_k \tilde{e}_{ijk}^2 - \sum_i \nu_{1i} E(\sum_j \sum_k \tilde{e}_{ijk}^2)] = o_p(1). \quad (4.35)$$

The proofs of these are relegated to Appendix B, Lemma B.1. Now, we can rewrite (4.31) as

$$\begin{aligned} & t^{-1} [\sum_i (\sum_j \sum_k \tilde{e}_{ijk}^2)^2 - (\sum_i \nu_{1i})^2 \text{var}(\sigma_e^2) - \sigma_e^4 \sum_i \nu_{1i}^2] \\ & + \sigma_e^4 t^{-1} \sum_i \nu_{1i}^2 - 2\hat{\sigma}_e^2 t^{-1} \sum_i \nu_{1i} [\sum_j \sum_k \tilde{e}_{ijk}^2 - \sigma_e^2 \nu_{1i}] \\ & - 2\hat{\sigma}_e^2 \sigma_e^2 t^{-1} \sum_i \nu_{1i}^2 + \hat{\sigma}_e^4 t^{-1} \sum_i \nu_{1i}^2 = o_p(1) \end{aligned}$$

using (4.32) (4.34). ■

Lemma 4.2 Assuming the conditions of Theorem 4.1, we have that

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

$$(b) \text{ var}(\sigma_\epsilon^{*2}) = \text{var}(\bar{\sigma}_\epsilon^2) + O(t^{-2}).$$

$$(c) \text{ } v_j(\sigma_\epsilon^{*2}) = v_j(\bar{\sigma}_\epsilon^2) + O_p(t^{-3/2}).$$

Proof (a) We can write

$$\begin{aligned} \sigma_\epsilon^{*2} - \bar{\sigma}_\epsilon^2 &= (\sum_i \nu_{1i})^{-1} [(\beta^{*2} - \beta^2) \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \\ &\quad - 2(\beta^* - \beta) \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})(x_{ijk} - \bar{x}_{ij.})] \\ &= -(\sum_i \nu_{1i})^{-1} (\beta^* - \beta)^2 \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2. \end{aligned}$$

Thus $\sigma_\epsilon^{*2} - \sigma_\epsilon^2 = O_p(t^{-1})$ since $(\sum_i \nu_{1i})^{-1} = O(t^{-1})$ by (4.29), and since $\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 = O(t)$ by regularity condition (R3). Finally,

$$(\beta^* - \beta)^2 = O_p(t^{-1}) \quad (4.36)$$

by applying Lemma 4.3 below to β^* , and by noting that

$$\text{var}(\beta^*) = \sigma_\epsilon^2 [\sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2]^{-1} = O(t^{-1}).$$

Lemma 4.3 (Bishop, Fienberg and Holland (1975), Theorem 14.4-1, pp. 476)

Let $\{X_n\}$ be a stochastic sequence such that $\mu_n = E(X_n)$ and $\sigma_n^2 = \text{var}(X_n) < \infty$.

Then $X_n - \mu_n = O_p(\sigma_n)$.

(b) From part (a), we can write

$$\begin{aligned} \text{var}(\sigma_\epsilon^{*2}) &= \text{var}(\bar{\sigma}_\epsilon^2) + (\sum_i \nu_{1i})^{-2} [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2]^2 \text{var}(\beta^* - \beta)^2 \\ &\quad - 2(\sum_i \nu_{1i})^{-1} \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \text{cov}(\bar{\sigma}_\epsilon^2, (\beta^* - \beta)^2). \end{aligned}$$

In view of the fact that $(\sum_i \nu_{1i})^{-1} = O(t^{-1})$ and $\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 = O(t)$ by regularity condition (R3), we need only show that

$$\text{var}(\beta^* - \beta)^2 = O(t^{-2}),$$

and that

$$\text{cov}(\bar{\sigma}_e^2, (\beta^* - \beta)^2) = O(t^{-2}).$$

Since the two proofs are similar in nature, we prove the latter assertion only.

First, define

$$\begin{aligned} d_i &= \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.}) \bar{c}_{ijk} \\ &= \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})(y_{ijk} - \bar{y}_{ij.}) - \beta \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \end{aligned}$$

so that $(\sum_i d_i) [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2]^{-1} = \beta^* - \beta$. Then, define

$$q_i = \sum_j \sum_k \bar{c}_{ijk}^2 - E[\sum_j \sum_k c_{ijk}^2]$$

so that $\sum_i q_i = \bar{\sigma}_e^2(\sum_i \nu_{1i}) - \sigma_e^2(\sum_i \nu_{1i} + O(1))$. Thus

$$\bar{\sigma}_e^2 = (\sum_i \nu_{1i})^{-1} [\sum_i q_i + \sigma_e^2 \sum_i \nu_{1i} + \sigma_e^2 O(1)].$$

In light of the above definitions, we can write

$$\text{cov}(\sigma_e^2, (\beta^* - \beta)^2) = (\sum_i \nu_{1i})^{-1} (\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2)^{-2} \text{cov}[(\sum_i d_i)^2, \sum_i q_i].$$

To prove (b), it suffices to show that $\text{cov}[(\sum_i d_i)^2, \sum_i q_i] = O(t)$. To do so, we require the well-known Cauchy-Schwarz inequality, stated below.

Lemma 4.4 Cauchy-Schwarz Inequality: *If $\{X_i\}$ and $\{Y_i\}$ are sequences of random variables then*

$$(a) \quad [\sum_i E(X_i Y_i)]^2 \leq [\sum_i E(X_i^2)] [\sum_i E(Y_i^2)].$$

(b) $(\sum_i X_i Y_i)^2 \leq (\sum_i X_i^2)(\sum_i Y_i^2)$. This is also true if $\{X_i\}$ and $\{Y_i\}$ are non-stochastic.

Now

$$|\text{cov}[(\sum_i d_i)^2, \sum_i q_i]| = |E((\sum_i d_i)^2 (\sum_i q_i))| = |E(\sum_i q_i d_i^2)|$$

since $(d_i, q_i; i = 1, \dots, t)$ are independent with $E(d_i) = E(q_i) = 0$. Now applying the Cauchy-Schwarz inequality (a), we have

$$|\text{cov}[(\sum_i d_i)^2, \sum_i q_i]| \leq [\sum_i E(q_i^2)]^{1/2} [\sum_i E(d_i^4)]^{1/2} = O(t)$$

if $E(q_i^2) = O(1)$ and $E(d_i^4) = O(1)$.

Now

$$\begin{aligned} E(q_i^2) &= E(\sum_j \sum_k \bar{c}_{ijk}^2 - E(\sum_j \sum_k \bar{c}_{ijk}^2))^2 \\ &= E(\sum_j \sum_k \bar{c}_{ijk}^2)^2 - [E(\sum_j \sum_k \bar{c}_{ijk}^2)]^2 < \infty \end{aligned}$$

using regularity condition (R4) in conjunction with Lemma B.2 from Appendix B, and noting that any lower order moment is also bounded. Finally,

$$|d_i| \leq (\sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2)^{1/2} (\sum_j \sum_k \bar{c}_{ijk}^2)^{1/2}$$

by invoking the Cauchy-Schwarz inequality (b), and thus

$$E(d_i^4) \leq (\sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2)^2 E(\sum_j \sum_k \bar{c}_{ijk}^2)^2 < \infty.$$

(c) We can write

$$v_j(\sigma_e^{*2}) = t(t-1)^{-1} (\sum_i \nu_{1i})^{-2} \sum_i [\sum_j \sum_k c_{ijk}^{*2} - \nu_{1i} \sigma_e^{*2}]^2$$

and

$$v_j(\sigma_e^2) = t(t-1)^{-1} (\sum_i \nu_{1i})^{-2} \sum_i [\sum_j \sum_k \bar{c}_{ijk}^2 - \nu_{1i} \bar{\sigma}_e^2]^2.$$

Now,

$$\begin{aligned} \sum_j \sum_k c_{ijk}^{*2} &= \sum_j \sum_k (y_{ijk} - y_{ij.})^2 + \beta^{*2} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 - 2\beta\beta^* \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \\ &\quad - 2\beta^* d_i \end{aligned}$$

and

$$\sum_j \sum_k \bar{e}_{ijk}^2 = \sum_j \sum_k (y_{ijk} - y_{ij})^2 - \beta^2 \sum_j \sum_k (x_{ijk} - x_{ij})^2 - 2\beta d_i.$$

Thus

$$\sum_j \sum_k e_{ijk}^{*2} = \sum_j \sum_k \bar{e}_{ijk}^2 - 2(\beta^* - \beta)d_i + (\beta^* - \beta)^2 \sum_j \sum_k (x_{ijk} - x_{ij})^2.$$

Substituting the last equation into the above expression for $v_j(\sigma_c^{*2})$, we get

$$\begin{aligned} v_j(\sigma_c^{*2}) &= t(t-1)^{-1} (\sum_i \nu_{1i})^{-2} \sum_i \{ \sum_j \sum_k e_{ijk}^2 - 2(\beta^* - \beta)d_i \\ &\quad + (\beta^* - \beta)^2 \sum_j \sum_k (x_{ijk} - x_{ij})^2 - \nu_{1i} \sigma_c^{*2} \}^2 \\ &= t(t-1)^{-1} (\sum_i \nu_{1i})^{-2} \sum_i \{ [\sum_j \sum_k e_{ijk}^2 - \nu_{1i} \sigma_c^2]^2 \\ &\quad + [(\beta^* - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2 - 2(\beta^* - \beta)d_i + \nu_{1i}(\sigma_c^2 - \sigma_c^{*2})]^2 \\ &\quad + 2[\sum_j \sum_k \bar{e}_{ijk}^2 - \nu_{1i} \bar{\sigma}_c^2][(\beta^* - \beta)^2 \sum_j \sum_k (x_{ijk} - x_{ij})^2 \\ &\quad - 2(\beta^* - \beta)d_i + \nu_{1i}(\bar{\sigma}_c^2 - \sigma_c^{*2})] \}. \end{aligned}$$

We apply the inequality $\sum_i (a_i + b_i)^2 \leq 2 \sum_i (a_i^2 + b_i^2)$ twice and obtain

$$\begin{aligned} v_j(\sigma_c^{*2}) &\leq v_j(\bar{\sigma}_c^2) + t(t-1)^{-1} (\sum_i \nu_{1i})^{-2} \{ 8(\beta^* - \beta)^2 \sum_i d_i^2 \\ &\quad + 4(\bar{\sigma}_c^2 - \sigma_c^{*2})^2 \sum_i \nu_{1i}^2 + 4(\beta^* - \beta)^4 \sum_i [\sum_j \sum_k (x_{ijk} - x_{ij})^2]^2 \\ &\quad + 2(\beta^* - \beta)^2 \sum_i (\sum_j \sum_k \bar{e}_{ijk}^2) (\sum_j \sum_k (x_{ijk} - x_{ij})^2) - 4(\beta^* - \beta) \sum_i d_i \sum_j \sum_k e_{ijk}^2 \\ &\quad + 2(\bar{\sigma}_c^2 - \sigma_c^{*2}) \sum_i \nu_{1i} \sum_j \sum_k \bar{e}_{ijk}^2 - 2(\beta^* - \beta)^2 \sigma_c^2 \sum_i \nu_{1i} \sum_j \sum_k (x_{ijk} - x_{ij})^2 \\ &\quad + 4(\beta^* - \beta) \bar{\sigma}_c^2 \sum_i \nu_{1i} d_i - 2\bar{\sigma}_c^2 (\bar{\sigma}_c^2 - \sigma_c^{*2}) \sum_i \nu_{1i}^2 \}. \end{aligned}$$

To prove part (c), we need only show that the 9 remainder terms on the right hand side of the inequality have orders no greater than $O_p(t^{-3/2})$. This is achieved using the following facts:

- (i) $(\sum_i \nu_{1i})^{-2} = O(t^{-2})$ via (4.29).
- (ii) $(\beta^* - \beta)^2 = O_p(t^{-1})$ via (4.36).

- (iii) $\sum_i d_i^2 = O(t)$ by applying Lemma 3.3 to d_i^2 and noting that $E(d_i^2) < \infty$. This is true since $E(d_i^4) < \infty$ from the proof of part (b), and since any lower order moment is also bounded.
- (iv) $(\sigma_e^2 - \sigma_e^{*2})^2 = O_p(t^{-2})$ by part (a).
- (v) $\sum_i \nu_{1i}^2 = O(t)$ since $\nu_{1i} = O(1)$.
- (vi) $\sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2 < \infty$ by regularity condition (R2).
- (vii) $\sum_j \sum_k \bar{e}_{ijk}^2 = O_p(1)$ using Lemma B.2 from Appendix B, where we show that $E[\sum_j \sum_k \bar{e}_{ijk}^2]^{2+\delta'} < \infty$ for $0 < \delta' < 1/2$. This implies that $E[\sum_j \sum_k \bar{e}_{ijk}^2] < \infty$ since any lower order moment is also bounded. We have the result upon applying Lemma 3.3.
- (viii) $\sigma_e^2 = O_p(1)$ since $\bar{\sigma}_e^2 = \bar{\sigma}_e^2 - \sigma_e^2 + \sigma_e^2 = O_p(1)$ by applying (4.33).
- (ix) $\sum_i d_i \sum_j \sum_k \bar{e}_{ijk}^2 = O_p(t)$ by applying Lemma 3.3 to $d_i \sum_j \sum_k \bar{e}_{ijk}^2$ and noting that $E[d_i \sum_j \sum_k \bar{e}_{ijk}^2] \leq (E(d_i^2)E(\sum_j \sum_k \bar{e}_{ijk}^2)^2)^{1/2} < \infty$ via Holder's inequality.

■

Lemma 4.5 Assuming the conditions of Theorem 4.1, we have that $t[r_J(\sigma_u^2) - \text{var}(\sigma_u^2)] = o_p(1)$.

Proof We can write $r_J(\sigma_u^2)$ as

$$\begin{aligned} r_J(\sigma_u^2) &= t(t-1)^{-1} (\sum_i \tau_{1i})^{-2} \sum_i \{ \sum_j \sum_k u_{ijk}^2 - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k \bar{e}_{ijk}^2 - \bar{\sigma}_u^2 \tau_{1i} \}^2 \\ &= t(t-1)^{-1} (\sum_i \tau_{1i})^{-2} \{ \bar{\sigma}_u^4 \sum_i \tau_{1i}^2 + \sum_i T_i^2 - 2\bar{\sigma}_u^2 \sum_i \tau_{1i} T_i \} \end{aligned}$$

by setting $T_i = \sum_j \sum_k u_{ijk}^2 - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k \bar{e}_{ijk}^2$.

Now, we establish that

$$\sum_i \tau_{1i} = O(t) \tag{4.37}$$

as follows:

$$\sum_i r_{1i} = n - \sum_i n_i^{-1} \sum_j n_{ij}^2 - [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2]^{-1} \sum_i \sum_j n_{ij}^2 (\bar{x}_{ij.} - \bar{x}_{i..})^2 = O(t)$$

using regularity conditions (R1), (R6) and (R8), and since $[\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2]^{-1}$ is bounded away from zero. Similarly, we can show that

$$(\sum_i r_{1i})^{-1} = O(t^{-1}). \quad (4.38)$$

Upon multiplying the above expression for $v_j(\bar{\sigma}_u^2)$ by $(\sum_i r_{1i}/t)^2$, we see that to prove the lemma, it is sufficient to show that

$$t^{-1} \{ \bar{\sigma}_u^4 \sum_i r_{1i}^2 + \sum_i T_i^2 - 2\bar{\sigma}_u^2 \sum_i r_{1i} T_i - (\sum_i r_{1i})^2 \text{var}(\sigma_u^2) \} = o_p(1). \quad (4.39)$$

To proceed further, we require the following three assertions:

$$\bar{\sigma}_u^2 - \sigma_u^2 = o_p(1), \quad (4.40)$$

$$t^{-1} [\sum_i r_{1i} T_i - \sigma_u^2 \sum_i r_{1i}^2] = o_p(1) \quad (4.41)$$

and

$$t^{-1} [\sum_i T_i^2 - (\sum_i r_{1i})^2 \text{var}(\sigma_u^2) - \sigma_u^4 \sum_i r_{1i}^2] = o_p(1). \quad (4.42)$$

The proofs of these are relegated to Appendix B, Lemma B.3. Now, we can rewrite (4.39) as

$$\begin{aligned} & t^{-1} [\sum_i T_i^2 - (\sum_i r_{1i})^2 \text{var}(\bar{\sigma}_u^2) - \bar{\sigma}_u^4 \sum_i r_{1i}^2] + \bar{\sigma}_u^4 t^{-1} \sum_i r_{1i}^2 \\ & - 2\bar{\sigma}_u^2 t^{-1} (\sum_i r_{1i} T_i - \bar{\sigma}_u^2 \sum_i r_{1i}^2) - 2\bar{\sigma}_u^2 \sigma_u^2 t^{-1} \sum_i r_{1i}^2 + \bar{\sigma}_u^4 t^{-1} \sum_i r_{1i}^2 = o_p(1) \end{aligned}$$

using (4.40) (4.42). ■

Lemma 4.6 *Assuming the conditions of Theorem 4.1, we have that*

$$(a) \quad \sigma_u^{*2} = \bar{\sigma}_u^2 + O_p(t^{-1}).$$

$$(b) \text{ var}(\sigma_u^{*2}) = \text{var}(\bar{\sigma}_u^2) + O(t^{-2}).$$

$$(c) \text{ } v_j(\sigma_u^{*2}) = v_j(\bar{\sigma}_u^2) + O_p(t^{-3/2}).$$

Proof (a) We can write

$$\begin{aligned} \sigma_u^{*2} - \bar{\sigma}_u^2 &= -(\sum_i \tau_{1i})^{-1} \{ (\beta^{**} - \beta)^2 \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \\ &\quad + (\beta^* - \beta)^2 \sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 - 2(\beta^* - \beta) \sum_i \nu_{1i}^{-1} \nu_{2i} d_i \}. \end{aligned}$$

Thus $\sigma_u^{*2} - \bar{\sigma}_u^2 = O_p(t^{-1})$ using regularity conditions (R2) and (R6), as well as (4.36), (4.38) and (B.3) from Appendix B. In addition, $\sum_i \nu_{1i}^{-1} \nu_{2i} d_i = O_p(t^{1/2})$ by applying Lemma 4.3 and noting that $\text{var}(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i) = \sum_i (\nu_{1i}^{-1} \nu_{2i})^2 E(d_i^2) = O(t)$. Finally, applying Lemma 4.3 to $\beta^{**} - \beta$, we have that

$$(\beta^{**} - \beta)^2 = O_p(t^{-1})$$

since

$$\begin{aligned} \text{var}(\beta^{**}) &= \sigma_e^2 [\sum_i \sum_j \sum_k (\tau_{ijk} - \bar{\tau}_{i..})^2]^{-1} \\ &\quad + \sigma_e^4 [\sum_i \sum_j \sum_k (\tau_{ijk} - \bar{\tau}_{i..})^2]^{-2} \sum_i \sum_j n_{ij}^2 (\bar{\tau}_{ij.} - \bar{\tau}_{i..})^2 = O(t^{-1}) \end{aligned}$$

using regularity conditions (R6) and (R8).

(b) From part (a), we can write

$$\begin{aligned} \text{var}(\sigma_u^{*2}) &= \text{var}(\bar{\sigma}_u^2) + (\sum_i \tau_{1i})^{-2} [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2]^2 \text{var}(\beta^{**} - \beta)^2 \\ &\quad + (\sum_i \tau_{1i})^{-2} [\sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2]^2 \text{var}(\beta^* - \beta)^2 \\ &\quad + 4(\sum_i \tau_{1i})^{-2} \sum_i (\nu_{1i}^{-1} \nu_{2i})^2 \text{var}(d_i (\beta^* - \beta)) \\ &\quad - 2(\sum_i \tau_{1i})^{-1} \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \text{cov}(\bar{\sigma}_u^2, (\beta^{**} - \beta)^2) \\ &\quad - 2(\sum_i \tau_{1i})^{-1} \sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \text{cov}(\bar{\sigma}_u^2, (\beta^* - \beta)^2) \\ &\quad + 4(\sum_i \tau_{1i})^{-1} \text{cov}(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i (\beta^* - \beta), \bar{\sigma}_u^2) \\ &\quad + 2(\sum_i \tau_{1i})^{-2} \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \text{cov}[(\beta^* - \beta)^2, (\beta^{**} - \beta)^2] \\ &\quad - 4(\sum_i \tau_{1i})^{-2} \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \text{cov}[\sum_i \nu_{1i}^{-1} \nu_{2i} d_i (\beta^* - \beta), (\beta^{**} - \beta)^2] \\ &\quad - 4(\sum_i \tau_{1i})^{-2} \sum_i \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \text{cov}[\sum_i \nu_{1i}^{-1} \nu_{2i} d_i (\beta^* - \beta), (\beta^* - \beta)^2]. \end{aligned}$$

We must show that the last 9 terms on the right hand side of the above equation have orders no greater than $O(t^{-2})$. To do so, we use regularity conditions (R2) and (R6), as well as (4.38) and (B.3) from Appendix B. We know that $\text{var}(\beta^* - \beta)^2 = O(t^{-2})$ from the proof of Lemma 4.2 (b). We have left to show the following:

- (i) $\text{var}(\beta^{**} - \beta)^2 = O(t^{-2})$.
- (ii) $\text{var}(d_i(\beta^* - \beta)) = O(t^{-1})$.
- (iii) $\text{cov}(\bar{\sigma}_u^2, (\beta^{**} - \beta)^2) = O(t^{-2})$.
- (iv) $\text{cov}(\bar{\sigma}_u^2, (\beta^* - \beta)^2) = O(t^{-2})$.
- (v) $\text{cov}(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i(\beta^* - \beta), \bar{\sigma}_u^2) = O(t^{-1})$.
- (vi) $\text{cov}((\beta^* - \beta)^2, (\beta^{**} - \beta)^2) = O(t^{-2})$.
- (vii) $\text{cov}(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i(\beta^* - \beta), (\beta^{**} - \beta)^2) = O(t^{-1})$.
- (viii) $\text{cov}(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i(\beta^* - \beta), (\beta^* - \beta)^2) = O(t^{-1})$.

The proofs of these are relegated to Appendix B, Lemma B.5.

(c) We can write

$$v_J(\sigma_u^{*2}) = t(t-1)^{-1}(\sum_i \tau_{1i})^{-2} \sum_i [\sum_j \sum_k u_{ijk}^{*2} - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k c_{ijk}^{*2} - \tau_{1i} \sigma_u^{*2}]^2$$

and

$$v_J(\bar{\sigma}_u^2) = t(t-1)^{-1}(\sum_i \tau_{1i})^{-2} \sum_i [\sum_j \sum_k u_{ijk}^2 - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k c_{ijk}^2 - \tau_{1i} \sigma_u^2]^2.$$

Now

$$\sum_j \sum_k u_{ijk}^{*2} = \sum_j \sum_k (y_{ijk} - \bar{y}_{i..})^2 + (\beta^{**2} - 2\beta^{**}\beta) \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 - 2\beta^{**} d_i^*$$

and

$$\sum_j \sum_k u_{ijk}^2 = \sum_j \sum_k (y_{ijk} - \bar{y}_{i..})^2 - \beta^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 - 2\beta d_i^*$$

where

$$d_i^* = \sum_j \sum_k (x_{ijk} - \bar{x}_{i..}) \bar{u}_{ijk}.$$

Thus,

$$\sum_j \sum_k u_{ijk}^{*2} = \sum_j \sum_k \bar{u}_{ijk}^2 + (\beta^{**} - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 - 2d_i^* (\beta^{**} - \beta).$$

In the proof of Theorem 4.2(c), we showed that

$$\sum_j \sum_k e_{ijk}^{*2} = \sum_j \sum_k \bar{e}_{ijk}^2 - 2(\beta^* - \beta)d_i + (\beta^* - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2.$$

Substituting these last two equations into the above expression for $v_j(\sigma_u^{*2})$, we get

$$\begin{aligned} v_j(\sigma_u^{*2}) &= t(t-1)^{-1} (\sum_i \tau_{1i})^{-2} \sum_i \{ \sum_j \sum_k \bar{u}_{ijk}^2 + (\beta^{**} - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \\ &\quad - 2d_i^* (\beta^{**} - \beta) - \nu_{1i}^{-1} \nu_{2i} [\sum_j \sum_k \bar{e}_{ijk}^2 - 2(\beta^* - \beta)d_i \\ &\quad + (\beta^* - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2] - \tau_{1i} \sigma_u^{*2} \}^2 \\ &= v_j(\bar{\sigma}_u^2) + t(t-1)^{-1} (\sum_i \tau_{1i})^{-2} \sum_i \{ (\beta^{**} - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \\ &\quad - 2d_i^* (\beta^{**} - \beta) + 2\nu_{1i}^{-1} \nu_{2i} (\beta^* - \beta)d_i - \nu_{1i}^{-1} \nu_{2i} (\beta^* - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \\ &\quad - \tau_{1i} (\sigma_u^{*2} - \bar{\sigma}_u^2) \}^2 - 2t(t-1)^{-1} (\sum_i \tau_{1i})^{-2} \sum_i \{ -(\beta^{**} - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \\ &\quad + 2d_i^* (\beta^{**} - \beta) - 2\nu_{1i}^{-1} \nu_{2i} (\beta^* - \beta)d_i + \nu_{1i}^{-1} \nu_{2i} (\beta^* - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \\ &\quad + \tau_{1i} (\sigma_u^{*2} - \bar{\sigma}_u^2) \} \cdot \{ \sum_j \sum_k \bar{u}_{ijk}^2 - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k \bar{e}_{ijk}^2 - \tau_{1i} \bar{\sigma}_u^2 \}. \end{aligned}$$

We apply the inequality $\sum_i (a_i + b_i)^2 \leq 2(\sum_i a_i^2 + \sum_i b_i^2)$ four times and obtain

$$\begin{aligned} v_j(\sigma_u^{*2}) &\leq v_j(\bar{\sigma}_u^2) + t(t-1)^{-1} (\sum_i \tau_{1i})^{-2} \{ 4(\beta^{**} - \beta)^4 \sum_i [\sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2]^2 \\ &\quad + 16(\beta^{**} - \beta)^2 \sum_i d_i^{*2} + 16(\beta^* - \beta)^2 \sum_i (\nu_{1i}^{-1} \nu_{2i})^2 d_i^2 \\ &\quad + 8(\beta^* - \beta)^4 \sum_i (\nu_{1i}^{-1} \nu_{2i})^2 [\sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2]^2 \\ &\quad + 8(\sigma_u^{*2} - \bar{\sigma}_u^2)^2 \sum_i \tau_{1i}^2 - 2 \sum_i \{ -(\beta^{**} - \beta)^2 \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 + 2(\beta^{**} - \beta)d_i^* \\ &\quad - 2(\beta^* - \beta)\nu_{1i}^{-1} \nu_{2i} d_i + (\beta^* - \beta)^2 \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 + (\sigma_u^{*2} - \bar{\sigma}_u^2) \tau_{1i} \} \\ &\quad \cdot \{ \sum_j \sum_k u_{ijk}^2 - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k \bar{e}_{ijk}^2 - \bar{\sigma}_u^2 \tau_{1i} \} \}. \end{aligned}$$

To prove part (c), we need only show that the remainder terms on the right hand side of the inequality have orders no greater than $O_p(t^{-3/2})$. This is achieved using the following facts:

- (i) $(\sum_i \tau_{1i})^{-2} = O(t^{-2})$ via (4.38).
- (ii) $(\beta^{**} - \beta)^4 = O_p(t^{-2})$ since $\text{var}(\beta^{**} - \beta) = \text{var}(\beta^{**}) = O(t^{-1})$ as demonstrated in the proof of part (a). Then apply Lemma 4.3.
- (iii) $\sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 < \infty$ by regularity condition (R5).
- (iv) $\sum_i d_i^{*2} = O_p(t)$ by applying Lemma 3.3 to d_i^{*2} and noting that $E(d_i^{*2}) < \infty$.
The proof of the latter statement is similar to showing that $E(d_i^2) < \infty$.
- (v) $(\nu_{1i}^{-1} \nu_{2i})^2 = O(1)$ via equation (B.3) from Appendix B.
- (vi) $(\beta^* - \beta)^2 = O_p(t^{-1})$ via (4.36).
- (vii) $\sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 < \infty$ by regularity condition (R2).
- (viii) $\sum_i \tau_{1i}^2 = O(t)$ since $\tau_{1i} = O(1)$.
- (ix) $(\sigma_u^{*2} - \bar{\sigma}_u^2)^2 = O_p(t^{-2})$ by part (a).
- (x) $\sum_j \sum_k u_{ijk}^2 = O_p(1)$ using Lemma B.4 from Appendix B, where we show that $E[\sum_j \sum_k \bar{u}_{ijk}^2]^{2+\delta'} < \infty$ for $0 < \delta' < 1/2$. This implies that $E[\sum_j \sum_k u_{ijk}^2] < \infty$ since any lower order moment is also bounded. We have the result upon applying Lemma 3.3.
- (xi) $\sum_j \sum_k \bar{c}_{ijk}^2 = O_p(1)$ using Lemma B.2 from Appendix B, and applying the same reasoning as in (x).
- (xii) $\bar{\sigma}_u^2 = O_p(1)$ since $\bar{\sigma}_u^2 = \bar{\sigma}_u^2 - \sigma_u^2 + \sigma_u^2 = O_p(1)$ by applying (4.40).

- (xiii) $\sum_i d_i^* \sum_j \sum_k \bar{u}_{ijk}^2 = O_p(t)$ by applying Lemma 3.3 to $d_i^* \sum_j \sum_k \bar{u}_{ijk}^2$ and noting that $E[d_i^* \sum_j \sum_k \bar{u}_{ijk}^2] \leq (E(d_i^{*2})E(\sum_j \sum_k \bar{u}_{ijk}^2))^1/2 < \infty$ via Holder's inequality.
- (xiv) $\sum_i \nu_{1i}^{-1} \nu_{2i} d_i^* \sum_j \sum_k \bar{e}_{ijk}^2 = O_p(t)$ using reasoning similar to (xiii).
- (xv) $\sum_i \nu_{1i}^{-1} \nu_{2i} d_i \sum_j \sum_k \bar{u}_{ijk}^2 = O_p(t)$ using reasoning similar to (xiii).
- (xvi) $\sum_i (\nu_{1i}^{-1} \nu_{2i})^2 d_i \sum_j \sum_k \bar{e}_{ijk}^2 = O_p(t)$ using reasoning similar to (xiii). ■

Remark 4.3 Assuming the conditions of Theorem 4.1, we have that the jackknife covariance estimators, given by (4.20)-(4.22) are consistent, i.e.,

$$t[\text{cov}_J(\sigma_i^{*2}, \sigma_j^{*2}) - \text{cov}(\sigma_i^{*2}, \sigma_j^{*2})] = o_p(1)$$

for $i, j = 1, 2, 3$. In view of Remark 4.1, it is sufficient to show that

$$t[v_J(\sigma_i^{*2} + \sigma_j^{*2}) - \text{var}(\sigma_i^{*2} + \sigma_j^{*2})] = o_p(1),$$

which follows along the lines of the proof of Theorem 4.1.

4.3 Robust Confidence Intervals On Nonlinear Functions of Variance Components

In many situations, it is of interest to obtain robust confidence intervals on bounded, nonlinear functions of the variance components $\sigma^2 = (\sigma_\xi^2, \sigma_u^2, \sigma_e^2)'$, say $g(\sigma^2)$. These confidence intervals are based on jackknife variance estimators which don't require underlying normality.

Assuming model (4.1), a weighted jackknife estimator of the nonlinear function $g(\sigma_\xi^2, \sigma_u^2, \sigma_e^2)$ is given by

$$g_J(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2}) = \bar{Q}(g) = t^{-1} \sum_{i=1}^t Q_i(g) \quad (4.43)$$

where

$$\begin{aligned}
Q_i(g) &= g(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2}) \\
&- t\{g(\sigma_\xi^{*2} + c_i(\sigma_\xi^{*2}(-i) - \sigma_\xi^{*2}), \sigma_u^{*2} + b_i(\sigma_u^{*2}(-i) - \sigma_u^{*2}), \sigma_e^{*2} + a_i(\sigma_e^{*2}(-i) - \sigma_e^{*2})) \\
&- g(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2})\}; \quad i = 1, \dots, t
\end{aligned} \tag{4.44}$$

are the weighted pseudo-values with weights a_i , b_i and c_i . The above choice of weighted pseudo-values is justified in that when $g(\sigma_\xi^2, \sigma_u^2, \sigma_e^2)$ is linear in the variance components, i.e., $g(\sigma_\xi^2, \sigma_u^2, \sigma_e^2) = \ell_1\sigma_\xi^2 + \ell_2\sigma_u^2 + \ell_3\sigma_e^2$ for known constants ℓ_1, ℓ_2 and ℓ_3 , then the jackknife estimator reduces to the unbiased estimator $g_J(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2}) = \ell_1\sigma_\xi^{*2} + \ell_2\sigma_u^{*2} + \ell_3\sigma_e^{*2}$. On the other hand, if $g(\sigma_\xi^2, \sigma_u^2, \sigma_e^2)$ is nonlinear in the variance components, then the bias of this estimator has an order lower than that of $g(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2})$. That is,

$$E[g_J(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2})] - g(\sigma_\xi^2, \sigma_u^2, \sigma_e^2) = o(t^{-1})$$

whereas

$$E[g(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2})] - g(\sigma_\xi^2, \sigma_u^2, \sigma_e^2) = O(t^{-1}).$$

The proofs of these are omitted here, but essentially involve finding the expectations of Taylor expansions. The corresponding weighted jackknife variance estimator is given by

$$v_{JW}(g) = t^{-1}(t-1)^{-1} \sum_i [Q_i(g) - \bar{Q}(g)]^2. \tag{4.45}$$

It can be shown that $v_{JW}(g)$ is consistent, but the proof is omitted here. These results can be used to form a robust confidence interval on $g(\sigma_\xi^2, \sigma_u^2, \sigma_e^2)$, given by

$$\begin{aligned}
g(\sigma_\xi^2, \sigma_u^2, \sigma_e^2) &\in [g_J(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2}) - t_{\frac{\alpha}{2}; t-1}(v_{JW}(g))^{1/2}, \\
&g_J(\sigma_\xi^{*2}, \sigma_u^{*2}, \sigma_e^{*2}) + t_{\frac{\alpha}{2}; t-1}(v_{JW}(g))^{1/2}]
\end{aligned} \tag{4.46}$$

where $t_{\frac{\alpha}{2}; t-1}$ is the upper $\frac{\alpha}{2}$ -point of a t -distribution with $t-1$ degrees of freedom.

Although weighted jackknifing is considered extensively in the literature, the evidence is not overwhelming that it has a clearcut advantage over unweighted jackknifing. One such possible unweighted jackknife variance estimator is given by

$$v_J(g) = t^{-1}(t-1) \sum_i [g(\sigma_{\xi}^{*2}(-i), \sigma_{\mathbf{u}}^{*2}(-i), \sigma_{\epsilon}^{*2}(-i)) - g(\sigma_{\xi}^{*2}, \sigma_{\mathbf{u}}^{*2}, \sigma_{\epsilon}^{*2})]^2 \quad (4.47)$$

and a corresponding robust confidence interval on $g(\sigma_{\xi}^2, \sigma_{\mathbf{u}}^2, \sigma_{\epsilon}^2)$ can be formed by taking

$$g(\sigma_{\xi}^2, \sigma_{\mathbf{u}}^2, \sigma_{\epsilon}^2) \in [g(\sigma_{\xi}^{*2}, \sigma_{\mathbf{u}}^{*2}, \sigma_{\epsilon}^{*2}) - t_{\frac{\alpha}{2}; t-1}(v_J(g))^{1/2}, \\ g(\sigma_{\xi}^{*2}, \sigma_{\mathbf{u}}^{*2}, \sigma_{\epsilon}^{*2}) + t_{\frac{\alpha}{2}; t-1}(v_J(g))^{1/2}]. \quad (4.48)$$

A useful application of (4.48) is the formation of confidence intervals on the ratios of variance components $\sigma_{\mathbf{u}}^2/\sigma_{\epsilon}^2$ and $\sigma_{\xi}^2/\sigma_{\epsilon}^2$ under the two-fold model. In the following section, the findings of a Monte Carlo study which investigates the finite sample performance of confidence intervals on these ratios, based on the above unweighted jackknife estimator, are reported.

4.4 Monte Carlo Study

In order to investigate the finite sample performance of confidence intervals based on jackknife estimators of nonlinear functions of σ^2 , a small Monte Carlo study was undertaken. The model used for this study was the balanced, two-fold nested error mean model with equal error variances, i.e.,

$$y_{ijk} = \beta + \xi_i + u_{ij} + e_{ijk}; \quad i = 1, \dots, t; \quad j = 1, \dots, m; \quad k = 1, \dots, n. \quad (4.49)$$

Note that this is a special case of model (4.1), which was used throughout this chapter. The following sections provide the objectives of the study, and a description of the experiment, as well as a discussion of the findings.

4.4.1 Objectives of the Study

The parameters of interest in this study are the ratios of the variance components $\rho_1 = \sigma_u^2/\sigma_e^2$ and $\rho_2 = \sigma_\xi^2/\sigma_e^2$. An exact $(1 - \alpha)\%$ confidence interval on ρ_1 exists for the balanced and unbalanced two-fold nested error mean model, assuming normality. However, an exact $(1 - \alpha)\%$ confidence interval on ρ_2 , independent of ρ_1 , is not possible. Many approximate solutions have been proposed in the past. Brœmeling (1969) finds a conservative simultaneous confidence region for ρ_1 and ρ_2 under the balanced two-fold nested error mean model, assuming normality. From the confidence region, it is possible to find one-sided confidence intervals on ρ_1 and ρ_2 separately. Assuming normality, Verdooren (1988) constructs a confidence interval on ρ_2 , given a specified value of ρ_1 , denoted by $[L(\rho_1), U(\rho_1)]$. An exact $(1 - \alpha)\%$ conservative confidence interval on ρ_2 , independent of ρ_1 , is then obtained by taking $\rho_2 \in [\min_{\rho_1}(L(\rho_1)), \max_{\rho_1}(U(\rho_1))]$, with ρ_1 varying in the range specified by the experimenter.

In this study, to find a robust confidence interval on $\rho_1 = \sigma_u^2/\sigma_e^2$, we exploit the invariance property and instead find a jackknife confidence interval on the nonlinear function

$$\theta_1 = \ln(1 + n\sigma_u^2/\sigma_e^2) = \ln(1 + n\rho_1), \quad (4.50)$$

as in Arvesen and Schmitz (1970), for the balanced one-fold nested error mean model. Note that θ_1 involves a log transformation on ρ_1 . This transformation stabilizes the variance, and therefore normality is approached more quickly. A robust confidence interval based on unweighted jackknifing is then given by

$$\theta_1 \in [\tilde{\theta}_1 - z_{\frac{\alpha}{2}}(v_J(\tilde{\theta}_1))^{1/2}, \tilde{\theta}_1 + z_{\frac{\alpha}{2}}(v_J(\tilde{\theta}_1))^{1/2}] = [\theta_1^I, \theta_1^{II}], \quad (4.51)$$

say, where

$$\hat{\theta}_1 = \ln[(\hat{\sigma}_e^2 + n\hat{\sigma}_u^2)/\hat{\sigma}_e^2] = \ln(MSU/MSE), \quad (4.52)$$

$$v_j(\hat{\theta}_1) = t^{-1}(t-1) \sum_{i=1}^t (\hat{\theta}_1 - \hat{\theta}_1(-i))^2 \quad (4.53)$$

and $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ -point of a $N(0, 1)$ distribution. In the balanced case, the Arvesen estimators reduce to the Henderson estimators, and so are denoted by $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$. The exact expressions for these are given by:

$$\hat{\sigma}_e^2 = [tm(n-1)]^{-1} \sum_i \sum_j \sum_k \hat{c}_{ijk}^2 \quad (4.54)$$

and

$$\hat{\sigma}_u^2 = [tn(m-1)]^{-1} [\sum_i \sum_j \sum_k \hat{u}_{ijk}^2 - t(nm-1)\hat{\sigma}_e^2] \quad (4.55)$$

where $\hat{c}_{ijk} = y_{ijk} - \bar{y}_{ij}$, and $\hat{u}_{ijk} = y_{ijk} - \bar{y}_{i..}$. The explicit expressions for MSU and MSE are given in Section 3.7.3. Also $\hat{\theta}_1(-i)$ is given by

$$\hat{\theta}_1(-i) = \ln[(\hat{\sigma}_e^2(-i) + n\hat{\sigma}_u^2(-i))/\hat{\sigma}_e^2(-i)] \quad (4.56)$$

where

$$\hat{\sigma}_e^2(-i) = [m(n-1)(t-1)]^{-1} [\sum_j \sum_k \hat{c}_{ijk}^2 - \sum_j \sum_k \hat{e}_{ijk}^2] \quad (4.57)$$

and

$$\begin{aligned} \hat{\sigma}_u^2(-i) = & [n(m-1)(t-1)]^{-1} \{ [\sum_j \sum_k \hat{u}_{ijk}^2 \\ & - m^{-1}(n-1)^{-1}(nm-1) \sum_j \sum_k \hat{c}_{ijk}^2] \\ & - [\sum_j \sum_k \hat{u}_{ijk}^2 - m^{-1}(n-1)^{-1}(nm-1) \sum_j \sum_k \hat{e}_{ijk}^2] \}. \end{aligned} \quad (4.58)$$

It can be shown that $(\hat{\sigma}_e^2(-i) + n\hat{\sigma}_u^2(-i))/\hat{\sigma}_e^2(-i) > 0$, and thus $\hat{\theta}_1(-i)$ is never undefined. Exploiting invariance, a $(1-\alpha)\%$ confidence interval on ρ_1 itself can be found by taking

$$\rho_1 \in [(\exp(\theta_1^L) - 1)/n, (\exp(\theta_1^U) - 1)/n]. \quad (4.59)$$

A traditional confidence interval on ρ_1 , based on underlying normality, is given by

$$\rho_1 \in \{n^{-1}[MSU/(MSE \cdot F_{\frac{\alpha}{2}}(t(m-1), tm(n-1))) - 1], \\ n^{-1}[MSU/(MSE \cdot F_{1-\frac{\alpha}{2}}(t(m-1), tm(n-1))) - 1]\} \quad (4.60)$$

where $F_{\frac{\alpha}{2}}(t(m-1), tm(n-1))$ represents the upper $\frac{\alpha}{2}$ -point of an F -distribution with $t(m-1)$ and $tm(n-1)$ degrees of freedom, and $F_{1-\frac{\alpha}{2}}(t(m-1), tm(n-1))$ represents the lower $\frac{\alpha}{2}$ -point of an F -distribution with $t(m-1)$ and $tm(n-1)$ degrees of freedom. Unfortunately, this confidence interval is highly nonrobust to nonnormality.

Our motivation for proposing the jackknife procedure is that it should perform as well as the normality based procedure when the data are normal, and that it should be robust to deviations from normality. The first objective of this study, then, is to compare the coverage probabilities of the jackknife confidence interval, given by (4.51), (or equivalently (4.59)), to that of the normality based confidence interval, given by (4.60), for nominal values of 90% and 95%, assuming both normal and nonnormal underlying distributions. It is also of interest to investigate the percentage relative bias of $v_J(\tilde{\theta}_1)$ with respect to the true variance of $\tilde{\theta}_1$, given by

$$R.B.(v_J(\tilde{\theta}_1)) = 100\{E_M(v_J(\tilde{\theta}_1)) - \text{var}_M(\tilde{\theta}_1)\}/\text{var}_M(\tilde{\theta}_1) \quad (4.61)$$

where E_M denotes Monte Carlo expectation, and var_M denotes Monte Carlo variance.

Finding a robust confidence interval on ρ_2 is more difficult, since it is not possible to find a log transformation on ρ_2 having a nonnegative domain. Thus, we take

$$\theta_2 = \rho_2 = \sigma_\xi^2/\sigma_\epsilon^2 \quad (4.62)$$

and form a robust confidence interval based on unweighted jackknifing, given by

$$\theta_2 \in [\tilde{\theta}_2 - z_{\frac{\alpha}{2}}(v_J(\tilde{\theta}_2))^{1/2}, \tilde{\theta}_2 + z_{\frac{\alpha}{2}}(v_J(\tilde{\theta}_2))^{1/2}] \quad (4.63)$$

where

$$\tilde{\theta}_2 = \tilde{\sigma}_\xi^2 / \tilde{\sigma}_e^2 \quad (4.64)$$

and

$$v_J(\tilde{\theta}_2) = t^{-1}(t-1) \sum_{i=1}^t (\tilde{\theta}_2 - \tilde{\theta}_2(-i))^2. \quad (4.65)$$

Here, $\tilde{\sigma}_\xi^2$ is given by

$$\begin{aligned} \tilde{\sigma}_\xi^2 = [nm(t-1)]^{-1} \{ \sum_i \sum_j \sum_k \tilde{\xi}_{ijk}^2 - t^{-1}(m-1)^{-1}(tm-1) \sum_i \sum_j \sum_k \tilde{u}_{ijk}^2 \\ - [tm(n-1)(m-1)]^{-1} [(m-1)(tnm-1) - (tm-1)(nm-1)] \sum_i \sum_j \sum_k \tilde{c}_{ijk}^2 \} \end{aligned} \quad (4.66)$$

where $\tilde{\xi}_{ijk} = y_{ijk} - \bar{y}_{...}$, and $\tilde{\theta}_2(-i)$ is given by

$$\tilde{\theta}_2(-i) = \tilde{\sigma}_\xi^2(-i) / \tilde{\sigma}_e^2(-i). \quad (4.67)$$

Now $\tilde{\sigma}_e^2(-i)$ is defined by (4.57) and $\tilde{\sigma}_\xi^2(-i)$ is defined as follows:

$$\begin{aligned} \tilde{\sigma}_\xi^2(-i) = (t-1)^{-2}(nm)^{-1} t \{ [\sum_i \sum_j \sum_k \tilde{\xi}_{ijk}^2 - t^{-1}(m-1)^{-1}(tm-1) \sum_i \sum_j \sum_k \tilde{u}_{ijk}^2 \\ - [tm(n-1)(m-1)]^{-1} [(m-1)(tnm-1) - (tm-1)(nm-1)] \sum_i \sum_j \sum_k \tilde{c}_{ijk}^2] \\ - [\sum_j \sum_k \tilde{\xi}_{ijk}^2 - t^{-1}(m-1)^{-1}(tm-1) \sum_j \sum_k \tilde{u}_{ijk}^2 \\ - [tm(n-1)(m-1)]^{-1} [(m-1)(tnm-1) - (tm-1)(nm-1)] \sum_j \sum_k \tilde{c}_{ijk}^2] \}. \end{aligned} \quad (4.68)$$

The second objective of this study is to investigate the coverage probabilities of the jackknife confidence interval, given by (4.63), for nominal values of 90% and 95%, assuming both normal and nonnormal underlying distributions. It is also of

interest to investigate the percentage relative bias of $v_J(\tilde{\theta}_2)$ with respect to the true variance of $\tilde{\theta}_2$, given by

$$R.B.(v_J(\tilde{\theta}_2)) = 100\{E_M(v_J(\tilde{\theta}_2)) - \text{var}_M(\tilde{\theta}_2)\}/\text{var}_M(\tilde{\theta}_2). \quad (4.69)$$

4.4.2 Description of the Experiment

For this study, model (4.49) was used, with $t = 30$ small areas, $m = 2$ PSUs within each small area, and $n = 2$ SSUs in each PSU within each small area. Without loss of generality, we set $\beta = 0$, since the confidence intervals used are invariant to the choice of β . The variance component σ_e^2 was fixed at $\sigma_e^2 = 300$, while σ_u^2 and σ_ξ^2 were each assigned the following values: 15, 30, 60, 150, 300 and 600, so that the ratios $\rho_1 = \sigma_u^2/\sigma_e^2$ and $\rho_2 = \sigma_\xi^2/\sigma_e^2$ took on all 36 possible combinations of the following values: 0.05, 0.1, 0.2, 0.5, 1.0 and 2.0. In addition, 2000 independent sets of pseudo random errors $\{\xi_i; i = 1, \dots, 30\}$, $\{u_{ij}; i = 1, \dots, 30; j = 1, 2\}$ and $\{e_{ijk}; i = 1, \dots, 30; j = 1, 2; k = 1, 2\}$ were generated according to the following three distributions: standard normal, standard double exponential and uniform (0,1). The random deviates $\{\xi_i\}$, $\{u_{ij}\}$ and $\{e_{ijk}\}$ were then transformed to deviates having means zero, and respective variances σ_ξ^2, σ_u^2 and σ_e^2 . From these generated sets, the observations $\{y_{ijk}; i = 1, \dots, 30; j = 1, 2; k = 1, 2\}$ were obtained using the model $y_{ijk} = \xi_i + u_{ij} + e_{ijk}$.

4.4.3 Results and Discussion

The coverage probabilities for both the jackknife and normality based confidence intervals on ρ_1 , given by (4.51), (or equivalently (4.59)), and (4.60) respectively, for nominal values of 90% and 95%, are reported in Table 4.1. The percentage relative biases of $v_J(\tilde{\theta}_1)$, with respect to the true variance of $\tilde{\theta}_1$, are reported as well. Since $\tilde{\theta}_1$ is independent of σ_ξ^2 , and thus is independent of $\rho_2 = \sigma_\xi^2/\sigma_e^2$

(for fixed σ_e^2), only varying values of $\rho_1 = \sigma_u^2/\sigma_e^2$ are considered here, having the following underlying distributions; normal, double exponential (long tail) and uniform (short tail). From this table, one can see that the normality based interval provides coverage probabilities close to the nominal values of 90% and 95% for normally distributed data only. Under the double exponential distribution, for a nominal value of 90%, the coverage probabilities monotonically decrease from 90.55% for $\rho_1 = .05$ to 82.65% for $\rho_1 = 2.0$. Under the same distribution, but for a nominal value of 95%, the coverage probabilities monotonically decrease again, this time from 95.20% for $\rho_1 = .05$ to 89.45% for $\rho_1 = 2.0$. On the other hand, under the uniform distribution, for a nominal value of 90%, the coverage probabilities of the normality based confidence interval monotonically increase from 90.35% for $\rho_1 = 0.1$ to 93.00% for $\rho_1 = 2.0$. Under the same distribution, but for a nominal value of 95%, the coverage probabilities monotonically increase again, this time from 94.65% for $\rho_1 = .05$ to 97.15% for $\rho_1 = 2.0$. Thus, there is a tendency towards undercoverage under the double exponential distribution, and towards overcoverage under the uniform distribution.

The coverage probabilities for the jackknife confidence interval behave much more evenly, regardless of the underlying distribution, although there is a slight tendency towards undercoverage. For a nominal value of 90%, across all three distributions, the coverage probabilities vary between 87.25% and 89.35%, whereas for a nominal value of 95%, the coverage probabilities vary between 93.05% and 94.30%.

The percentage relative bias of $v_r(\hat{\theta}_1)$ is small across all distributions, ranging between 2.92% to 8.04%.

The expected lengths of the intervals (4.59) and (4.60) were also computed for nominal values of 90% and 95%, under normality. For a nominal value of 90%,

the expected lengths of the jackknife confidence interval increased monotonically from 0.62 for $\rho_1 = .05$ to 2.83 for $\rho_1 = 2.0$, whereas the expected lengths of the normality based confidence interval increased monotonically from 0.64 for $\rho_1 = .05$ to 2.92 for $\rho_1 = 2.0$. For a nominal value of 95%, the expected lengths of the jackknife confidence interval increased monotonically from 0.76 for $\rho_1 = .05$ to 3.44 for $\rho_1 = 2.0$, whereas the expected lengths of the normality based confidence interval increased monotonically from 0.78 for $\rho_1 = .05$ to 3.57 for $\rho_1 = 2.0$. Thus, the increase in expected length, if any, due to using a jackknife confidence interval over a normality based confidence interval was negligible.

The coverage probabilities for the jackknife based confidence interval on ρ_2 , given by (4.63), for nominal values of 90% and 95% are reported in Table 4.2, for normally distributed data. The percentage relative biases of $v_J(\tilde{\theta}_2)$, with respect to the true variance of $\tilde{\theta}_2$, are reported as well. These values are reported for all 36 combinations of values of ρ_1 and ρ_2 . The results of Table 4.2 indicate that the coverage probabilities for the jackknife confidence interval are fairly close to the nominal values of 90% and 95%, over the entire range of ρ_1 and ρ_2 values, with a tendency towards increased undercoverage as ρ_2 increases. For a nominal value of 90%, the coverage probabilities range from a low of 86.50% for $\rho_1 = .05, \rho_2 = 1.0$ to a high of 91.45% for $\rho_1 = 0.1, \rho_2 = .05$. For a nominal value of 95%, the coverage probabilities range from a low of 90.30% for $\rho_1 = .05, \rho_2 = 1.0$ to a high of 96.30% for $\rho_1 = 1.0, \rho_2 = 0.1$.

The percentage relative biases of $v_J(\tilde{\theta}_2)$ are always less than 10%, ranging from a low of 3.61% for $\rho_1 = 2.0, \rho_2 = .05$ to a high of 7.28% for $\rho_1 = 0.2, \rho_2 = 0.5$.

In Table 4.3, similar coverage probabilities and percentage relative biases to that of Table 4.2 are displayed, but for double exponentially distributed data. Here, the jackknife confidence interval behaves reasonably well, but not as well

as under normality, again with a tendency towards increased undercoverage as ρ_2 increases. For a nominal value of 90%, the coverage probabilities range from a low of 83.05% for $\rho_1 = 0.1, \rho_2 = 2.0$ to a high of 92.00% for $\rho_1 = 2.0, \rho_2 = .05$. For a nominal value of 95%, the coverage probabilities range from a low of 87.15% for $\rho_1 = 0.1, \rho_2 = 2.0$ to a high of 96.45% for $\rho_1 = 2.0, \rho_2 = .05$. The percentage relative biases of $v_j(\tilde{\theta}_2)$ are small, ranging from a low of 1.81% for $\rho_1 = 2.0, \rho_2 = 0.2$ to a high of 9.38% for $\rho_1 = .05, \rho_2 = 0.5$.

In Table 4.4, similar coverage probabilities and percentage relative biases to that of Table 4.2 are displayed, but for uniformly distributed data. In this table, there is an improvement in the behaviour of the jackknife confidence interval over that of the last table. Once again, however, there is a tendency towards increased undercoverage as ρ_2 increases. For a nominal value of 90%, the coverage probabilities range from a low of 86.10% for $\rho_1 = 0.5, \rho_2 = 1.0$ to a high of 91.10% for $\rho_1 = 0.1, \rho_2 = .05$. For a nominal value of 95%, the coverage probabilities range from a low of 91.10% for $\rho_1 = 2.0, \rho_2 = 2.0$ to a high of 96.20% for $\rho_1 = 0.1, \rho_2 = .05$. The percentage relative biases of $v_j(\tilde{\theta}_2)$ are small with a slight tendency towards underestimation for large ρ_2 values, and overestimation for small ρ_2 values. The percentages vary from a low of -4.04% for $\rho_1 = .05, \rho_2 = 0.5$ to a high of 5.03% for $\rho_1 = 0.5, \rho_2 = .05$.

In conclusion, the jackknife confidence interval on ρ_1 , given by (4.59), behaves well when the data are normal, and is robust to deviations from normality. On the other hand, the jackknife confidence interval on ρ_2 , given by (4.63), behaves well for small values of ρ_2 , but experiences significant undercoverage for large values of ρ_2 . The situation could be improved if one were able to find a suitable log transformations on ρ_2 , having a nonnegative domain. However, the interval given by (4.63) is still of use, since no exact $(1 - \alpha)\%$ confidence interval on ρ_2 exists, even assuming normality.

**Table 4.1 Coverage Probabilities of Jackknife and Normality Based
Confidence Intervals on ρ_1 for Nominal Values of 90% and 95%,
and Percentage Relative Biases of $v_J(\hat{\theta}_1)$ with Respect to the True
Variance of $\hat{\theta}_1$**

	ρ_1	$CP_J(90\%)$	$CP_N(90\%)$	$CP_J(95\%)$	$CP_N(95\%)$	$R.B.(v_J(\hat{\theta}_1))$
Normal	.05	89.25	90.80	94.10	95.50	7.79
	0.1	89.35	90.90	94.20	95.55	8.04
	0.2	89.35	91.00	94.30	95.70	8.04
	0.5	89.25	90.90	94.05	95.80	7.41
	1.0	88.75	90.20	94.20	95.75	6.55
	2.0	88.40	90.10	93.80	95.60	5.63
Double Exponential	.05	87.25	90.55	93.05	95.20	3.92
	0.1	88.05	90.10	93.35	95.15	4.29
	0.2	88.35	89.60	93.75	94.30	4.88
	0.5	88.25	87.50	93.40	92.90	6.08
	1.0	87.85	85.10	93.40	91.00	7.11
	2.0	88.00	82.65	93.45	89.45	7.82
Uniform	.05	89.05	90.40	93.40	94.65	2.92
	0.1	88.90	90.35	93.75	94.85	3.63
	0.2	88.95	90.70	93.45	95.35	4.64
	0.5	88.30	91.40	94.00	95.70	5.71
	1.0	88.65	91.95	93.80	96.50	5.40
	2.0	88.80	93.00	93.90	97.15	4.35

Table 4.2 Coverage Probabilities of Jackknife Confidence Intervals on ρ_2 for Nominal Values of 90% and 95% (Denoted $CP(90\%)$ and $CP(95\%)$ Respectively), and Percentage Relative Biases of $v_j(\hat{\theta}_2)$ with Respect to the True Variance of $\hat{\theta}_2$, Under the Normal Distribution

		ρ_1					
		.05	0.1	0.2	0.5	1.0	2.0
ρ_2	.05	91.30	91.45	90.65	91.00	90.95	90.85
		95.80	95.80	96.05	96.20	96.25	96.00
		5.03	5.72	6.20	5.87	4.82	3.61
	0.1	90.85	90.75	90.20	90.65	90.30	91.00
		95.05	95.45	95.85	96.15	96.30	95.90
		5.51	6.25	6.74	6.31	5.09	3.71
	0.2	88.90	89.35	89.70	90.05	90.25	90.90
		93.75	94.00	93.90	95.00	95.35	95.65
		5.92	6.68	7.21	6.79	5.43	3.85
	0.5	86.90	86.95	87.80	88.60	88.80	89.20
		91.20	91.25	91.95	92.80	93.20	94.40
		6.04	6.71	7.28	7.19	5.97	4.15
	1.0	86.50	86.85	86.60	86.75	87.35	87.30
		90.30	90.30	90.95	91.45	91.10	92.70
		5.90	6.41	6.92	7.11	6.34	4.64
	2.0	86.70	86.90	86.95	87.00	86.60	87.15
		90.40	90.55	90.60	90.50	90.35	90.70
		5.78	6.13	6.52	6.84	6.54	5.34

Note: Each column of entries for each (ρ_1, ρ_2) combination represents:
 $CP(90\%)$
 $CP(95\%)$
 $R.B.(v_j(\hat{\theta}_2))$

Table 4.3 Coverage Probabilities of Jackknife Confidence Intervals on ρ_2 for Nominal Values of 90% and 95% (Denoted $CP(90\%)$ and $CP(95\%)$ Respectively), and Percentage Relative Biases of $v_r(\hat{\theta}_2)$ with Respect to the True Variance of $\hat{\theta}_2$, Under the Double Exponential Distribution

	ρ_1						
	.05	0.1	0.2	0.5	1.0	2.0	
ρ_2	.05	90.05	89.95	89.40	90.95	91.25	92.00
		95.00	95.05	94.95	95.20	96.00	96.45
		5.43	4.54	3.54	2.51	2.26	2.41
	0.1	89.75	89.40	89.35	89.80	91.15	91.95
		94.65	94.70	94.70	95.55	95.75	96.15
		6.60	5.64	4.40	2.79	2.13	2.09
	0.2	88.35	88.10	88.25	88.95	90.70	91.65
		92.80	93.35	93.30	94.35	95.00	95.85
		8.22	7.35	6.02	3.74	2.35	1.81
	0.5	85.20	85.80	86.50	86.90	88.65	89.85
		89.10	89.30	90.25	91.35	92.65	94.25
		9.38	8.99	8.19	6.14	3.96	2.16
	1.0	83.85	84.25	94.80	85.95	86.15	87.80
		88.00	87.90	88.35	89.55	91.00	92.45
		9.10	8.98	8.67	7.57	5.88	3.61
	2.0	83.10	83.05	83.80	84.05	84.75	85.80
		87.20	87.15	87.40	88.00	89.05	90.35
		8.45	8.46	8.39	8.00	7.16	5.54

Note: Each column of entries for each (ρ_1, ρ_2) combination represents:
 $CP(90\%)$
 $CP(95\%)$
 $R.B.(v_r(\hat{\theta}_2))$

Table 4.4 Coverage Probabilities of Jackknife Confidence Intervals on ρ_2 for Nominal Values of 90% and 95% (Denoted $CP(90\%)$ and $CP(95\%)$ Respectively), and Percentage Relative Biases of $v_J(\tilde{\theta}_2)$ with Respect to the True Variance of $\tilde{\theta}_2$, Under the Uniform Distribution

	ρ_1					
	.05	0.1	0.2	0.5	1.0	2.0
ρ_2	.05	90.95	91.10	90.85	90.15	90.10
		96.10	96.20	96.10	96.20	95.45
		1.17	2.51	4.08	5.03	3.92
	0.1	89.45	90.00	90.55	90.05	89.60
		94.90	95.05	95.65	95.55	95.35
		-0.39	0.76	2.23	3.43	2.75
	0.2	88.10	88.45	88.30	88.80	89.00
		93.50	93.40	94.05	94.30	94.40
		-2.31	-1.42	-0.16	1.27	1.16
	0.5	86.75	86.65	86.65	87.25	88.05
		91.70	91.35	91.30	92.30	92.95
		-4.04	-3.59	-2.82	-1.57	-1.14
	1.0	86.80	86.55	86.15	86.10	86.75
		91.40	91.40	91.40	91.65	91.70
		-3.78	-3.56	-3.16	-2.42	-2.07
	2.0	87.30	87.60	87.15	86.75	86.60
		91.85	91.70	91.70	91.10	91.40
		-2.42	-2.29	-2.09	-1.74	-1.62

Note: Each column of entries for each (ρ_1, ρ_2) combination represents:
 $CP(90\%)$
 $CP(95\%)$
 $R.B.(v_J(\tilde{\theta}_2))$

CHAPTER 5

A Hierarchical Bayes Approach to the Two-Fold Nested Error Regression Model

This chapter focuses on the Hierarchical Bayes (HB) approach to the two-fold nested error regression model. The main advantage to this approach is that exact measures of uncertainty may be found via posterior variances, whereas in previous chapters only second order approximations to the MSEs of the EBLUPs are possible. In Section 5.1, the HB approach is described in terms of a general mixed linear model and then applied to the special case of the two-fold nested error regression model, with equal errors variances. In Section 5.2, the main theorem of Datta and Ghosh (1990), which gives the posterior distribution of the nonsampled data, given the sampled data, is cited. In this theorem, the priors are modelled in stages, and one of the parameters, λ , is assumed to be known. Although Datta and Ghosh provide a proof of their theorem, a simpler alternative proof is offered, in this section. Then, the Bayes estimates of the small area means under the two-fold model are found, by invoking the Datta and Ghosh theorem. These are shown to be equivalent to the BLUPs found in Chapter 3. Finally, the corresponding posterior variances of the small area means are found. In Section 5.3, the main theorem of Datta and Ghosh (1990) is generalized to the case of unknown λ . The posterior distribution of the nonsampled data, given the sampled data, is found in stages, assuming λ to be unknown. Under this setup, the posterior expectations and variances of the small area means under the two-fold model are again found.

5.1 Description of the HB Model and Its Application to the Two-Fold Nested Error Regression Model

Consider the following HB model:

(A) Conditional on $\mathbf{B} = \boldsymbol{\beta}$, $\mathbf{A} = \boldsymbol{\lambda}$ and $L = \ell$, let $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \ell^{-1}(\mathbf{R}^* + \mathbf{Z}\mathbf{G}^*(\boldsymbol{\lambda})\mathbf{Z}') = \boldsymbol{\Sigma})$, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s)'$. Here $\mathbf{Y}$ is an N -vector of response variables for the $N = \sum_{i=1}^t \sum_{j=1}^{M_i} N_{ij}$ population units and $\mathbf{X}$ is a known $N \times p$ matrix of auxiliary variables.

(B) $\mathbf{B}$, $\mathbf{A}$ and L have a specified (proper or improper) joint prior distribution.

Stage (A) of the model can be reformulated in terms of a general mixed linear model (see Definition 2.4), i.e., we can write

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (5.1)$$

where $\mathbf{u}$ and $\mathbf{e}$ are independent $b \times 1$ and $N \times 1$ vectors respectively, with $\mathbf{u} \sim N(0, \ell^{-1}\mathbf{G}^*(\boldsymbol{\lambda}) = \mathbf{G}(\boldsymbol{\lambda}))$ and $\mathbf{e} \sim N(0, \ell^{-1}\mathbf{R}^* = \mathbf{R})$. In addition, $\mathbf{X}$ and $\mathbf{Z}$ are known $N \times p$ and $N \times b$ design matrices. Finally, $\mathbf{R}^*$ is a known, positive definite $N \times N$ matrix, whereas $\mathbf{G}^*(\boldsymbol{\lambda})$ is a positive definite $b \times b$ matrix, which is structurally known except possibly for some vector of parameters, $\boldsymbol{\lambda}$. In this chapter, $\boldsymbol{\lambda}$ is a vector consisting of ratios of variance components.

Now, we can partition the matrices $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{e}$ in terms of sampled and nonsampled units, and rewrite model (5.1) as

$$\begin{pmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{Z}^{(1)} \\ \mathbf{Z}^{(2)} \end{pmatrix} \mathbf{u} + \begin{pmatrix} \mathbf{e}^{(1)} \\ \mathbf{e}^{(2)} \end{pmatrix}$$

where $\mathbf{Y}^{(1)}$ and $\mathbf{e}^{(1)}$ have dimension $n \times 1$, $\mathbf{X}^{(1)}$ has dimension $n \times p$ and $\mathbf{Z}^{(1)}$ has dimension $n \times b$. Similarly, $\mathbf{Y}^{(2)}$ and $\mathbf{e}^{(2)}$ have dimension $(N - n) \times 1$, $\mathbf{X}^{(2)}$ has dimension $(N - n) \times p$ and $\mathbf{Z}^{(2)}$ has dimension $(N - n) \times b$. We shall assume for simplicity that the rank of $\mathbf{X}^{(1)}$ is p . Note that $\mathbf{Y}^{(1)}$ represents the vector of sampled units from the t small areas, whereas $\mathbf{Y}^{(2)}$ represents the vector of

nonsampled units. We partition the corresponding variance-covariance matrix $\Sigma = \ell^{-1} \Sigma^* = \ell^{-1}(\mathbf{R}^* + \mathbf{ZG}^*(\lambda)\mathbf{Z}')$ as follows:

$$\Sigma = \ell^{-1} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

In this chapter, we consider the two-fold nested error regression model

$$y_{ijk} = \mathbf{x}'_{ijk}\beta + \xi_i + u_{ij} + e_{ijk}; \quad i = 1, \dots, t; \quad j = 1, \dots, M_i; \quad k = 1, \dots, N_{ij} \quad (5.2)$$

where $\{\xi_i, u_{ij}, e_{ijk}; i = 1, \dots, t; j = 1, \dots, M_i; k = 1, \dots, N_{ij}\}$ are mutually independent with $\xi_i \sim N(0, \sigma_\xi^2 = (\lambda_1 \ell)^{-1})$, $u_{ij} \sim N(0, \sigma_u^2 = (\lambda_2 \ell)^{-1})$ and $e_{ijk} \sim N(0, \sigma_e^2 = \ell^{-1})$. Here $\lambda = (\lambda_1, \lambda_2)' = (\sigma_e^2/\sigma_\xi^2, \sigma_e^2/\sigma_u^2)'$. We can write model (5.2) in partitioned form where

$$\begin{aligned} \mathbf{Y}^{(1)} &= \begin{pmatrix} \text{col}_{1 \leq i \leq t} & \text{col}_{1 \leq j \leq m_i} & \text{col}_{1 \leq k \leq n_{ij}} \end{pmatrix} (y_{ijk}), \\ \mathbf{Y}^{(2)} &= \begin{pmatrix} \text{col}_{1 \leq i \leq t} & \text{col}_{1 \leq j \leq M_i} & \text{col}_{1+q_{ij} \leq k \leq N_{ij}} \end{pmatrix} (y_{ijk}), \\ q_{ij} &= n_{ij} I(j \leq m_i), \\ \mathbf{X}^{(1)} &= \begin{pmatrix} \text{col}_{1 \leq i \leq t} & \text{col}_{1 \leq j \leq m_i} & \text{col}_{1 \leq k \leq n_{ij}} \end{pmatrix} (\mathbf{x}'_{ijk}), \\ \mathbf{X}^{(2)} &= \begin{pmatrix} \text{col}_{1 \leq i \leq t} & \text{col}_{1 \leq j \leq M_i} & \text{col}_{1+q_{ij} \leq k \leq N_{ij}} \end{pmatrix} (\mathbf{x}'_{ijk}), \\ \mathbf{Z}^{(1)} &= \begin{pmatrix} \text{Diag} (1_{n_i}) & \text{Diag} & \text{Diag} (1_{n_{ij}}) & \mathbf{0}_{n \times (M-m)} \end{pmatrix}, \\ &\quad \begin{matrix} 1 \leq i \leq t & 1 \leq j \leq m_i \end{matrix} \\ \mathbf{Z}^{(2)} &= \begin{pmatrix} \text{Diag} (1_{N_i - n_i}) & \text{Diag} & \text{Diag} (1_{N_{ij} - q_{ij}}) \end{pmatrix}, \\ &\quad \begin{matrix} 1 \leq i \leq t & 1 \leq j \leq M_i \end{matrix} \\ \mathbf{u} &= \begin{pmatrix} \text{col}_{1 \leq i \leq t} (\xi_i) \\ \text{col}_{1 \leq i \leq t} & \text{col}_{1 \leq j \leq m_i} (u_{ij}) \\ \text{col}_{1 \leq i \leq t} & \text{col}_{m_i+1 \leq j \leq M_i} (u_{ij}) \end{pmatrix}, \\ \mathbf{e}^{(1)} &= \begin{pmatrix} \text{col}_{1 \leq i \leq t} & \text{col}_{1 \leq j \leq m_i} & \text{col}_{1 \leq k \leq n_{ij}} \end{pmatrix} (e_{ijk}), \\ \mathbf{e}^{(2)} &= \begin{pmatrix} \text{col}_{1 \leq i \leq t} & \text{col}_{1 \leq j \leq M_i} & \text{col}_{1+q_{ij} \leq k \leq N_{ij}} \end{pmatrix} (e_{ijk}), \\ \mathbf{R}^* &= \mathbf{I}_N \end{aligned}$$

and

$$G^*(\lambda) = \begin{pmatrix} \lambda_1^{-1} \mathbf{I}_l & \mathbf{0} \\ \mathbf{0} & \lambda_2^{-1} \mathbf{I}_M \end{pmatrix}.$$

In this chapter, $\mathbf{1}_u$ represents a vector of order u consisting entirely of ones, $\mathbf{I}_v$ represents the identity matrix of order v , and $\mathbf{0}_{u \times v}$ represents a $u \times v$ matrix consisting entirely of zeros. Also $I(j \leq m_i) = 1$ if $j \leq m_i$, and 0 otherwise, and $M. = \sum_{i=1}^l M_i$ while $m. = \sum_{i=1}^l m_i$.

5.2 An HB Analysis Assuming λ is Known

The focus of this section is to find the conditional distribution of $\mathbf{Y}^{(2)}$ given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$. In the theorem that follows, λ is assumed to be known, and the priors are modelled in stages. The desired conditional distribution is then given.

Theorem 5.1 (Datta and Ghosh (1990), Theorem 2) *Assume the model given by (A) and (B) in Section 5.1 with λ known, and $\mathbf{B}$ and L independent such that $\mathbf{B}$ has a uniform ($\mathbb{R}^p$) prior and L has a Gamma ($a_0/2, g_0/2$); $a_0 \geq 0, g_0 \geq 0$ prior. Then, assuming that $n + g_0 - p - 2 > 0$, the conditional distribution of $\mathbf{Y}^{(2)}$ given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$ is multivariate- t with location parameter $\mathbf{M}\mathbf{y}^{(1)}$, scale parameter $(n + g_0 - p)^{-1}(a_0 + \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)})\mathbf{Q}$ and degrees of freedom $\nu = n + g_0 - p$. Here*

$$\mathbf{K} = \Sigma_{11}^{-1} - \Sigma_{11}^{-1}\mathbf{X}^{(1)}(\mathbf{X}^{(1)'}\Sigma_{11}^{-1}\mathbf{X}^{(1)})^{-1}\mathbf{X}^{(1)'}\Sigma_{11}^{-1},$$

$$\mathbf{M} = \Sigma_{21}\mathbf{K} + \mathbf{X}^{(2)}(\mathbf{X}^{(1)'}\Sigma_{11}^{-1}\mathbf{X}^{(1)})^{-1}\mathbf{X}^{(1)'}\Sigma_{11}^{-1}$$

and

$$\begin{aligned} \mathbf{Q} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\ &+ (\mathbf{X}^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}^{(1)})(\mathbf{X}^{(1)'}\Sigma_{11}^{-1}\mathbf{X}^{(1)})^{-1}(\mathbf{X}^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}^{(1)})'. \end{aligned}$$

Proof Although Datta and Ghosh offer a proof for this theorem, the alternative proof presented here is simpler. To proceed, we require the following definitions:

Definition 5.1 A vector of random variables $\mathbf{T}$, given by $\mathbf{T} = (T_1, \dots, T_q)'$, is said to have a multivariate t -distribution with location parameter μ , scale parameter Φ and degrees of freedom ν , if it has p.d.f.

$$\pi(\mathbf{t}) \propto |\Phi|^{-1/2} [\nu + (\mathbf{t} - \mu)' \Phi^{-1} (\mathbf{t} - \mu)]^{-(\nu+q)/2}.$$

Assuming that $\nu > 2$, we have that $E(\mathbf{T}) = \mu$ and $\text{var}(\mathbf{T}) = \nu(\nu - 2)^{-1} \Phi$.

Definition 5.2 A random variable T is said to have a Gamma (a, g) distribution if it has p.d.f.

$$\pi(t) = [a^g t^{g-1} \exp(-at) / \Gamma(g)] I(t > 0); \quad a > 0, \quad g > 0.$$

Definition 5.3 An N -vector of random variables $\mathbf{T}$ is said to have a multivariate normal distribution with mean vector μ and variance-covariance matrix Φ if it has p.d.f.

$$\pi(\mathbf{t}) = (2\pi)^{-N/2} |\Phi|^{-1/2} \exp\{-1/2(\mathbf{t} - \mu)' \Phi^{-1} (\mathbf{t} - \mu)\}.$$

As an added fact, if $\mathbf{T}$ has a multivariate normal distribution and can be partitioned such that

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^{(1)} \\ \mathbf{T}^{(2)} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} \right),$$

then the conditional distribution of $\mathbf{T}^{(2)}$ given $\mathbf{T}^{(1)} = \mathbf{t}^{(1)}$ is as follows:

$$\mathbf{T}^{(2)} | \mathbf{t}^{(1)} \sim N(\mu_2 + \Phi_{21} \Phi_{11}^{-1} (\mathbf{t}^{(1)} - \mu_1), \Phi_{22} - \Phi_{21} \Phi_{11}^{-1} \Phi_{12}).$$

We commence by noting that the conditional distribution of $\mathbf{Y}^{(2)}$ given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$ has a p.d.f. that can be written as

$$\pi(\mathbf{y}^{(2)} | \mathbf{y}^{(1)}) = \int_{\ell} \int_{\beta} \pi(\mathbf{y}^{(2)} | \mathbf{y}^{(1)}, \beta, \ell) \pi(\beta, \ell | \mathbf{y}^{(1)}) d\beta d\ell \quad (5.3)$$

where $\pi(\mathbf{y}^{(2)}|\mathbf{y}^{(1)}, \boldsymbol{\beta}, \ell)$ is the p.d.f. of $\mathbf{Y}^{(2)}$ given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$, $\mathbf{B} = \boldsymbol{\beta}$ and $L = \ell$, and where $\pi(\boldsymbol{\beta}, \ell|\mathbf{y}^{(1)})$ is the p.d.f. of $\mathbf{B}$ and L given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$. However, $\pi(\boldsymbol{\beta}, \ell|\mathbf{y}^{(1)})$ can be written as

$$\pi(\boldsymbol{\beta}, \ell|\mathbf{y}^{(1)}) \propto \pi(\mathbf{y}^{(1)}|\boldsymbol{\beta}, \ell)\pi(\boldsymbol{\beta})\pi(\ell) \quad (5.4)$$

where $\pi(\mathbf{y}^{(1)}|\boldsymbol{\beta}, \ell)$ is the p.d.f. of $\mathbf{Y}^{(1)}$ given $\mathbf{B} = \boldsymbol{\beta}$ and $L = \ell$, $\pi(\boldsymbol{\beta})$ is the p.d.f. of $\mathbf{B}$ and where $\pi(\ell)$ is the p.d.f. of L .

Now, noting that

$$\begin{pmatrix} \mathbf{Y}^{(1)}|\boldsymbol{\beta}, \ell \\ \mathbf{Y}^{(2)}|\boldsymbol{\beta}, \ell \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} \boldsymbol{\beta}, \ell^{-1} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right),$$

and, using Definition 5.3, we have that

$$\mathbf{Y}^{(2)}|\mathbf{y}^{(1)}, \boldsymbol{\beta}, \ell \sim N(\mathbf{X}^{(2)}\boldsymbol{\beta} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\boldsymbol{\beta}), \ell^{-1}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}) = \ell^{-1}\boldsymbol{\Sigma}_{22.1}).$$

Now since $\mathbf{M}\mathbf{y}^{(1)}$ can be written as

$$\mathbf{M}\mathbf{y}^{(1)} = \mathbf{X}^{(2)}\tilde{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\tilde{\boldsymbol{\beta}})$$

where

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}^{(1)})^{-1}\mathbf{X}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}^{(1)},$$

we have that

$$\mathbf{X}^{(2)}\boldsymbol{\beta} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\boldsymbol{\beta}) = \mathbf{M}\mathbf{y}^{(1)} + \boldsymbol{\Delta}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})$$

where

$$\boldsymbol{\Delta} = \mathbf{X}^{(2)} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}^{(1)}.$$

Thus, using Definition 5.3,

$$\begin{aligned} \pi(\mathbf{y}^{(2)}|\mathbf{y}^{(1)}, \boldsymbol{\beta}, \ell) &\propto \ell^{(N-n)/2} \exp\{-\ell/2[\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)} \\ &\quad - \boldsymbol{\Delta}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})]'\boldsymbol{\Sigma}_{22.1}^{-1}[\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)} - \boldsymbol{\Delta}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})]\}. \end{aligned} \quad (5.5)$$

Using (5.4), as well as Definitions 5.2 and 5.3, $\pi(\beta, \ell | y^{(1)})$ can be written as

$$\begin{aligned}\pi(\beta, \ell | y^{(1)}) &\propto \ell^{(n_0/2)-1} \exp(-a_0 \ell/2) \ell^{n/2} \exp\{-\ell/2(y^{(1)} - X^{(1)}\beta)' \Sigma_{11}^{-1}(y^{(1)} - X^{(1)}\beta)\} \\ &= \ell^{(n+n_0)/2-1} \exp\{-\ell/2[a_0 + (y^{(1)} - X^{(1)}\tilde{\beta})' \Sigma_{11}^{-1}(y^{(1)} - X^{(1)}\tilde{\beta}) \\ &\quad + (\beta - \tilde{\beta})' X^{(1)'} \Sigma_{11}^{-1} X^{(1)}(\beta - \tilde{\beta})]\}.\end{aligned}$$

Substituting the above expression and (5.5) into (5.3), we have

$$\begin{aligned}\pi(y^{(2)} | y^{(1)}) &\propto \int_{\ell} \ell^{(N+n_0)/2-1} \exp\{-\ell/2[a_0 + (y^{(1)} - X^{(1)}\tilde{\beta})' \Sigma_{11}^{-1}(y^{(1)} - X^{(1)}\tilde{\beta})]\} \\ &\quad \cdot \int_{\beta} \exp\{-\ell/2[(\beta - \tilde{\beta})' X^{(1)'} \Sigma_{11}^{-1} X^{(1)}(\beta - \tilde{\beta}) + (y^{(2)} - My^{(1)} \\ &\quad - \Delta(\beta - \tilde{\beta}))' \Sigma_{22.1}^{-1}(y^{(2)} - My^{(1)} - \Delta(\beta - \tilde{\beta}))]\} d\beta \, d\ell \\ &= \int_{\ell} \ell^{(N+n_0)/2-1} \exp\{-\ell/2[a_0 + (y^{(1)} - X^{(1)}\tilde{\beta})' \Sigma_{11}^{-1}(y^{(1)} - X^{(1)}\tilde{\beta}) \\ &\quad + (y^{(2)} - My^{(1)})' \Sigma_{22.1}^{-1}(y^{(2)} - My^{(1)})]\} \\ &\quad \cdot \int_{\beta} \exp\{-\ell/2[(\beta - \tilde{\beta})' F(\beta - \tilde{\beta}) - 2(\beta - \tilde{\beta})' H + H' F^{-1} H - H' F^{-1} H]\} d\beta \, d\ell\end{aligned}$$

where

$$F = \Delta' \Sigma_{22.1}^{-1} \Delta + X^{(1)'} \Sigma_{11}^{-1} X^{(1)}$$

and

$$H = \Delta' \Sigma_{22.1}^{-1}(y^{(2)} - My^{(1)}).$$

Thus,

$$\begin{aligned}\pi(y^{(2)} | y^{(1)}) &\propto \int_{\ell} \ell^{(N+n_0)/2-1} \exp\{-\ell/2[a_0 + (y^{(1)} - X^{(1)}\tilde{\beta})' \Sigma_{11}^{-1}(y^{(1)} - X^{(1)}\tilde{\beta}) \\ &\quad + (y^{(2)} - My^{(1)})' \Sigma_{22.1}^{-1}(y^{(2)} - My^{(1)}) - H' F^{-1} H]\} \\ &\quad \cdot \int_{\beta} \exp\{-\ell/2[(\beta - \tilde{\beta}) - F^{-1} H]' F[(\beta - \tilde{\beta}) - F^{-1} H]\} d\beta \, d\ell.\end{aligned}\tag{5.6}$$

To perform the inner integral with respect to β , make the transformation $\mathbf{v} = \ell^{1/2} \mathbf{T}(\beta - \tilde{\beta} - \mathbf{F}^{-1} \mathbf{H})$ where $\mathbf{F} = \mathbf{T}'\mathbf{T}$ for some $\mathbf{T}$, so that $d\mathbf{v} = \ell^{r/2} |\mathbf{T}| d\beta$. Thus, the inner integral of (5.6) becomes

$$\begin{aligned} & \int_{\beta} \exp\{-\ell/2[(\beta - \tilde{\beta}) - \mathbf{F}^{-1} \mathbf{H}]' \mathbf{F}[(\beta - \tilde{\beta}) - \mathbf{F}^{-1} \mathbf{H}]\} d\beta \\ &= \ell^{-r/2} |\mathbf{T}|^{-1} \int_{\mathbf{v}} \exp\{-\mathbf{v}' \mathbf{v}/2\} d\mathbf{v} \propto \ell^{-r/2} \end{aligned}$$

noting that the p.d.f. of a multivariate $N(0, 1)$ random p -vector integrated over the real line p times, equals one. Consequently,

$$\begin{aligned} \pi(\mathbf{y}^{(2)} | \mathbf{y}^{(1)}) &\propto \int_{\ell} \ell^{(N+g-r)/2-1} \exp\{-\ell/2[a_0 + (\mathbf{y}^{(1)} - \mathbf{X}^{(1)} \tilde{\beta})' \Sigma_{11}^{-1} (\mathbf{y}^{(1)} - \mathbf{X}^{(1)} \tilde{\beta}) \\ &+ (\mathbf{y}^{(2)} - \mathbf{M} \mathbf{y}^{(1)})' (\Sigma_{22.1}^{-1} - \Sigma_{22.1}^{-1} \Delta \mathbf{F}^{-1} \Delta' \Sigma_{22.1}^{-1}) (\mathbf{y}^{(2)} - \mathbf{M} \mathbf{y}^{(1)})]\} d\ell. \end{aligned}$$

Now, we can write $\mathbf{y}^{(1)'} \mathbf{K} \mathbf{y}^{(1)}$ as

$$\mathbf{y}^{(1)'} \mathbf{K} \mathbf{y}^{(1)} = \mathbf{y}^{(1)'} \Sigma_{11}^{-1} \mathbf{y}^{(1)} - \mathbf{y}^{(1)'} \Sigma_{11}^{-1} \mathbf{X}^{(1)} \tilde{\beta} = (\mathbf{y}^{(1)} - \mathbf{X}^{(1)} \tilde{\beta})' \Sigma_{11}^{-1} (\mathbf{y}^{(1)} - \mathbf{X}^{(1)} \tilde{\beta}). \quad (5.7)$$

Next, we invert $\mathbf{Q} = \Sigma_{22.1} + \Delta(\mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{X}^{(1)})^{-1} \Delta'$ using the following lemma:

Lemma 5.1 (Rao (1965), pp. 33, #2.9) *Let $\mathbf{A}$ and $\mathbf{D}$ be nonsingular matrices of order m and n respectively, and $\mathbf{B}$ be an $m \times n$ matrix. Then*

$$(\mathbf{A} + \mathbf{B} \mathbf{D} \mathbf{B}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{B}' \mathbf{A}^{-1} \mathbf{B} + \mathbf{D}^{-1})^{-1} \mathbf{B}' \mathbf{A}^{-1}.$$

Applying Lemma 5.1 to $\mathbf{Q}$ with $\mathbf{A} = \Sigma_{22.1}$, $\mathbf{B} = \Delta$ and $\mathbf{D} = (\mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{X}^{(1)})^{-1}$, we get

$$\mathbf{Q}^{-1} = \Sigma_{22.1}^{-1} - \Sigma_{22.1}^{-1} \Delta \mathbf{F}^{-1} \Delta' \Sigma_{22.1}^{-1}. \quad (5.8)$$

Now substituting expressions (5.7) and (5.8) into the expression for $\pi(\mathbf{y}^{(2)}|\mathbf{y}^{(1)})$, we have

$$\begin{aligned}\pi(\mathbf{y}^{(2)}|\mathbf{y}^{(1)}) &\propto \int_{\ell} \ell^{(N+g_0-p)/2-1} \exp\{-\ell/2[a_0 + \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)} \\ &\quad + (\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)})'\mathbf{Q}^{-1}(\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)})]\} d\ell \\ &= \int_{\ell} \ell^{g-1} \exp\{-a\ell\} d\ell\end{aligned}$$

where

$$a = \frac{1}{2}[a_0 + \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)} + (\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)})'\mathbf{Q}^{-1}(\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)})]$$

and

$$g = (N + g_0 - p)/2 = (N - n)/2 + (n + g_0 - p)/2 = (q + \nu)/2,$$

and where

$$q = N - n$$

and

$$\nu = n + g_0 - p.$$

Making the transformation $v = a\ell$, we have

$$\pi(\mathbf{y}^{(2)}|\mathbf{y}^{(1)}) \propto a^{-1} \int_v (v/a)^{g-1} e^{-v} dv = a^{-g} \int_v v^{g-1} e^{-v} dv = a^{-g} \Gamma(g) \propto a^{-g}.$$

Thus,

$$\begin{aligned}\pi(\mathbf{y}^{(2)}|\mathbf{y}^{(1)}) &\propto [a_0 + \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)} + (\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)})'\mathbf{Q}^{-1}(\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)})]^{-(N+g_0-p)/2} \\ &= [n + g_0 - p + (a_0 + \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)})^{-1}(n + g_0 - p) \\ &\quad \cdot (\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)})'\mathbf{Q}^{-1}(\mathbf{y}^{(2)} - \mathbf{M}\mathbf{y}^{(1)})]^{-(N+g_0-p)/2} \\ &\quad \cdot [(n + g_0 - p)^{-1}(a_0 + \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)})]^{-(N+g_0-p)/2} \\ &\propto [\nu + (\mathbf{y}^{(2)} - \boldsymbol{\mu})'\boldsymbol{\Phi}^{-1}(\mathbf{y}^{(2)} - \boldsymbol{\mu})]^{-(\nu+g)/2} |\boldsymbol{\Phi}|^{-1/2}\end{aligned}$$

where

$$\mu = \mathbf{M}\mathbf{y}^{(1)}$$

and

$$\Phi = (n + g_0 - p)^{-1}(a_0 + \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)})\mathbf{Q}.$$

In view of Definition 5.1, the proof is complete. ■

Now, we are interested in predicting linear functions of the form $\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)}$ where $\mathbf{A}$ and $\mathbf{C}$ are known matrices. As always, our specific interest lies in estimating the i^{th} small area mean for $i = 1, \dots, t$. That is, we are interested in estimating $(\bar{Y}_{1..}, \dots, \bar{Y}_{t..})'$ where

$$Y_{i..} = N_i^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk} = N_i^{-1} \left\{ \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} y_{ijk} + \sum_{j=1}^{m_i} \sum_{k=n_{ij}+1}^{N_{ij}} y_{ijk} + \sum_{j=m_i+1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk} \right\}.$$

The vector of small area means can be written in the following form:

$$(\bar{Y}_{1..}, \dots, \bar{Y}_{t..})' = \mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)}$$

where

$$\mathbf{A} = \text{Diag} (N_i^{-1} \mathbf{1}'_{n_i})_{1 \leq i \leq t}$$

and

$$\mathbf{C} = \text{Diag} (N_i^{-1} \mathbf{1}'_{N_i - n_i})_{1 \leq i \leq t}.$$

Note that $\mathbf{A}$ has dimension $t \times n$ whereas $\mathbf{C}$ has dimension $t \times (N - n)$. The Bayes estimator of $(Y_{1..}, \dots, Y_{t..})'$ conditional on $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$ is given by the posterior expectation of $\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)}$, i.e.,

$$\begin{aligned} E[(Y_{1..}, \dots, Y_{t..})' | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] &= E[\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] \\ &= \mathbf{A}\mathbf{y}^{(1)} + \mathbf{C}E[\mathbf{Y}^{(2)} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] \\ &= (\mathbf{A} + \mathbf{C}\mathbf{M})\mathbf{y}^{(1)}. \end{aligned}$$

Now since

$$\mathbf{M}\mathbf{y}^{(1)} = \mathbf{X}^{(2)}\tilde{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\tilde{\boldsymbol{\beta}}), \quad (5.9)$$

we can write

$$E[(Y_{1..}, \dots, Y_{t..})' | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] = \mathbf{A}\mathbf{y}^{(1)} + \mathbf{C}\mathbf{X}^{(2)}\tilde{\boldsymbol{\beta}} + \mathbf{C}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}^{(2)} - \mathbf{X}^{(1)}\tilde{\boldsymbol{\beta}}). \quad (5.10)$$

We require explicit expressions for $\boldsymbol{\Sigma}_{11}^{-1}$, $\boldsymbol{\Sigma}_{21}$ and $\tilde{\boldsymbol{\beta}}$. We use the expression for $\mathbf{V}_i$, given by (3.4) in Chapter 3 with $f_{ijk} = 1$, and write

$$\boldsymbol{\Sigma}_{11} = \text{Diag} (\boldsymbol{\Sigma}_{11}^{(i)})_{1 \leq i \leq t}$$

where

$$\boldsymbol{\Sigma}_{11}^{(i)} = \mathbf{I}_{n_i} + \lambda_2^{-1} \text{Diag} (\mathbf{J}_{n_{ij}}) + \lambda_1^{-1} \mathbf{J}_{n_i} = \sigma_e^{-2} \mathbf{V}_i,$$

and where $\mathbf{J}_u$ represents a $u \times u$ matrix, consisting entirely of ones. Now, to find $\boldsymbol{\Sigma}_{11}^{-1}$, we use expression (3.7) from Chapter 3 and write

$$\boldsymbol{\Sigma}_{11}^{-1} = \text{Diag} (\boldsymbol{\Sigma}_{11}^{(i)-1})_{1 \leq i \leq t} \quad (5.11)$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{11}^{(i)-1} &= \sigma_e^2 \mathbf{V}_i^{-1} = \mathbf{I}_{n_i} - \text{Diag} (n_{ij}^{-1} \gamma_{ij} \mathbf{J}_{n_{ij}}) \\ &\quad - \lambda_2 \gamma_i \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \left[\text{col}_{1 \leq j \leq m_i} (n_{ij}^{-1} \gamma_{ij} \mathbf{1}_{n_{ij}}) \right] \left[\text{col}_{1 \leq j \leq m_i} (n_{ij}^{-1} \gamma_{ij} \mathbf{1}_{n_{ij}}) \right]', \end{aligned}$$

and where γ_{ij} and γ_i , defined by (3.6) in Chapter 3, can also be written as

$$\gamma_{ij} = n_{ij}(n_{ij} + \lambda_2)^{-1}$$

and

$$\gamma_i = \lambda_2 \left(\sum_{j=1}^{m_i} \gamma_{ij} \right) (\lambda_1 + \lambda_2 \sum_{j=1}^{m_i} \gamma_{ij})^{-1}.$$

Next,

$$\Sigma_{21} = \text{Diag} (\Sigma_{21}^{(i)}) \quad (5.12)$$

where

$$\begin{aligned} \Sigma_{21}^{(i)} &= \begin{pmatrix} \lambda_1^{-1} \mathbf{J}_{(N_i^* - n_i) \times n_i} + \lambda_2^{-1} \text{Diag} (\mathbf{J}_{(N_{ij} - n_{ij}) \times n_{ij}})_{1 \leq j \leq m_i} \\ \lambda_1^{-1} \mathbf{J}_{(N_i - N_i^*) \times n_i} \end{pmatrix} \\ &= \lambda_1^{-1} \mathbf{J}_{(N_i - q_i) \times n_i} + \lambda_2^{-1} \text{Diag} (\mathbf{B}_{ij})_{1 \leq j \leq M_i}, \end{aligned}$$

and where

$$\mathbf{B}_{ij} = \begin{cases} \mathbf{J}_{(N_{ij} - n_{ij}) \times n_{ij}} & \text{if } j \leq m_i \\ \mathbf{0}_{N_{ij} \times n_{ij}} & \text{if } j > m_i \end{cases}.$$

Note that $q_i = \sum_{j=1}^{M_i} q_{ij}$ and $N_i^* = \sum_{j=1}^{m_i} N_{ij}$, and that Σ_{21} has dimension $(N - n) \times n$. Finally, we have that $\tilde{\beta} = (\mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{X}^{(1)})^{-1} \mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{y}^{(1)}$, and in view of Remark 3.4 from Chapter 3 with $f_{ijk} = 1$, we have explicit expressions for the two component parts of $\tilde{\beta}$, as follows:

$$\begin{aligned} \mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{X}^{(1)} &= \sum_i \sum_j \sum_k \mathbf{x}_{ijk} \mathbf{x}_{ijk}' - \sum_i \sum_j n_{ij} \gamma_{ij} \bar{\mathbf{x}}_{ij} \bar{\mathbf{x}}_{ij}' \\ &\quad - \lambda_2 \sum_i [\gamma_i (\sum_j \gamma_{ij})^{-1} (\sum_j \gamma_{ij} \bar{\mathbf{x}}_{ij}) (\sum_j \gamma_{ij} \bar{\mathbf{x}}_{ij}')] \end{aligned} \quad (5.13)$$

and

$$\begin{aligned} \mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{y}^{(1)} &= \sum_i \sum_j \sum_k y_{ijk} \mathbf{x}_{ijk} - \sum_i \sum_j n_{ij} \gamma_{ij} \bar{y}_{ij} \bar{\mathbf{x}}_{ij} \\ &\quad - \lambda_2 \sum_i [\gamma_i (\sum_j \gamma_{ij})^{-1} (\sum_j \gamma_{ij} \bar{y}_{ij}) (\sum_j \gamma_{ij} \bar{\mathbf{x}}_{ij})]. \end{aligned} \quad (5.14)$$

Now, evaluating (5.10), we find the i^{th} element to be:

$$\begin{aligned} E[N_i^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] &= N_i^{-1} \{ \sum_{j=1}^{m_i} n_{ij} \bar{y}_{ij} + \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \bar{\mathbf{x}}_{ij}' \tilde{\beta} \\ &\quad + \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \gamma_{ij} (\bar{y}_{ij} - \bar{\mathbf{x}}_{ij}' \tilde{\beta}) \\ &\quad + \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} (\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (1 - \gamma_{ij}) \sum_{j=1}^{m_i} \gamma_{ij} (\bar{y}_{ij} - \bar{\mathbf{x}}_{ij}' \tilde{\beta}) \\ &\quad + (\sum_{j=1}^{m_i} N_{ij} \bar{\mathbf{x}}_{ij} - \sum_{j=1}^{m_i} N_{ij} \bar{\mathbf{x}}_{ij})' \tilde{\beta} \\ &\quad + (\sum_{j=1}^{m_i} N_{ij} - \sum_{j=1}^{m_i} N_{ij}) \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} (\bar{y}_{ij} - \bar{\mathbf{x}}_{ij}' \tilde{\beta}) \} \end{aligned} \quad (5.15)$$

which is equal to expression (3.9) in Chapter 3, with $f_{ijk} = 1$. Thus, the Bayes estimator of $Y_{i..}$ using HB modelling is identical to the BLUP of $Y_{i..}$ for the finite population case! This is not entirely surprising, however, since if we write $\mathbf{M}\mathbf{y}^{(1)}$ as it is in (5.9), then Royall (1976) shows that the BLUP of $\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)}$ is given by $\mathbf{A}\mathbf{y}^{(1)} + \mathbf{C}\mathbf{M}\mathbf{y}^{(1)}$. Datta and Ghosh (1990) also provide a proof of this, but the result of Royall makes their proof unnecessary.

Finally, we are interested in finding the posterior variances of the small area means, since they are a natural measure of the risk associated with the posterior expectations. Using the results of Theorem 5.1, the posterior variance of $\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)}$ is given by

$$\begin{aligned}\text{var}[\mathbf{A}\mathbf{Y}^{(1)} + \mathbf{C}\mathbf{Y}^{(2)} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] &= \mathbf{C}\text{var}(\mathbf{Y}^{(2)} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)})\mathbf{C}' \\ &= (n + g_0 - p - 2)^{-1} (a_0 + \mathbf{y}^{(1)'} \mathbf{K} \mathbf{y}^{(1)}) \mathbf{C} \mathbf{Q} \mathbf{C}'\end{aligned}\quad (5.16)$$

where

$$\mathbf{Q} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} + \Delta (\mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{X}^{(1)})^{-1} \Delta'$$

and where $\Delta = \mathbf{X}^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} \mathbf{X}^{(1)}$.

We require an explicit expression for Σ_{22} . this is given by

$$\Sigma_{22} = \text{Diag} (\Sigma_{22}^{(i)})_{1 \leq i \leq t} \quad (5.17)$$

where

$$\Sigma_{22}^{(i)} = \mathbf{I}_{N_i - n_i} + \lambda_2^{-1} \text{Diag} (\mathbf{J}_{(N_{ij} - q_{ij}) \times (N_{ij} - q_{ij})}) + \lambda_1^{-1} \mathbf{J}_{(N_i - n_i) \times (N_i - n_i)}$$

and where q_{ij} is as before.

Using (5.7), we can write $\mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)}$ as

$$\mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)} = \mathbf{y}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}^{(1)} - (\mathbf{X}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}^{(1)})'(\mathbf{X}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}^{(1)})^{-1}(\mathbf{X}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}^{(1)})$$

which, in view of Remark 3.4 in Chapter 3 with $f_{ijk} = 1$, can be written as

$$\begin{aligned} \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)} &= \left\{ \sum_i \sum_j \sum_k y_{ijk}^2 - \sum_i \sum_j n_{ij} \gamma_{ij} \bar{y}_{ij}^2 - \lambda_2 \sum_i \gamma_i (\sum_j \gamma_{ij})^{-1} (\sum_j \gamma_{ij} \bar{y}_{ij})^2 \right\} \\ &- \left\{ \sum_i \sum_j \sum_k y_{ijk} \mathbf{x}_{ijk} - \sum_i \sum_j n_{ij} \gamma_{ij} \bar{y}_{ij} \bar{\mathbf{x}}_{ij} - \lambda_2 \sum_i \gamma_i (\sum_j \gamma_{ij})^{-1} (\sum_j \gamma_{ij} \bar{y}_{ij}) (\sum_j \gamma_{ij} \bar{\mathbf{x}}_{ij}) \right\}' \\ &\cdot (\mathbf{X}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}^{(1)})^{-1} \\ &\cdot \left\{ \sum_i \sum_j \sum_k y_{ijk} \mathbf{x}_{ijk} - \sum_i \sum_j n_{ij} \gamma_{ij} \bar{y}_{ij} \bar{\mathbf{x}}_{ij} - \lambda_2 \sum_i \gamma_i (\sum_j \gamma_{ij})^{-1} (\sum_j \gamma_{ij} \bar{y}_{ij}) (\sum_j \gamma_{ij} \bar{\mathbf{x}}_{ij}) \right\}. \end{aligned} \quad (5.18)$$

Now the posterior variance of the i^{th} small area mean is given by the i^{th} diagonal element of (5.16), i.e.,

$$\begin{aligned} \text{var}(N_i^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}) &= (n + g_0 - p - 2)^{-1} (a_0 + \mathbf{y}^{(1)'}\mathbf{K}\mathbf{y}^{(1)}) \\ &\cdot \{ N_i^{-2} [\lambda_1^{-1} (1 - \gamma_i) (N_i - n_i) - \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \gamma_{ij}]^2 + \lambda_2^{-1} \sum_{j=1}^{m_i} (N_{ij} - n_{ij})^2 (1 - \gamma_{ij}) \\ &+ \lambda_2^{-1} \sum_{j=1}^{M_i} N_{ij}^2 - \lambda_2^{-1} \sum_{j=1}^{m_i} N_{ij}^2 + (N_i - n_i) \} \\ &+ N_i^{-2} \left[\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (\bar{\mathbf{x}}_{ij}^* - \gamma_{ij} \bar{\mathbf{x}}_{ij}) + \sum_{j=1}^{M_i} N_{ij} (\bar{\mathbf{X}}_{ij} - \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}) \right. \\ &- \sum_{j=1}^{m_i} N_{ij} (\mathbf{X}_{ij} - \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}) \\ &- \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} (\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (1 - \gamma_{ij})) \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij} \left. \right]' \\ &\cdot (\mathbf{X}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}^{(1)})^{-1} \\ &\cdot \left[\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (\bar{\mathbf{x}}_{ij}^* - \gamma_{ij} \bar{\mathbf{x}}_{ij}) + \sum_{j=1}^{M_i} N_{ij} (\bar{\mathbf{X}}_{ij} - \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}) \right. \\ &- \sum_{j=1}^{m_i} N_{ij} (\mathbf{X}_{ij} - \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}) \\ &- \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} (\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (1 - \gamma_{ij})) \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij} \left. \right\} \end{aligned} \quad (5.19)$$

where $\mathbf{y}^{(1)'} \mathbf{K} \mathbf{y}^{(1)}$ is given by (5.18).

Remark 5.1 Note that the off-diagonal elements of (5.16) with our choice of $\mathbf{A}$ and $\mathbf{C}$ give the posterior covariances between the small area means.

5.3 An HB Analysis Assuming λ is Unknown

In Section 5.2, we assumed λ to be known. In general, however, λ is unknown and must have a prior distribution assigned to it, to reflect the uncertainty. We now consider the model given by (A) and (B) in Section 5.1 where λ is unknown, and attempt to find the conditional distribution of $\mathbf{Y}^{(2)}$ given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$ in several steps, via the following theorem due to Datta and Ghosh (1990).

Theorem 5.2 (Datta and Ghosh (1990), Theorem 1) Assume the model given by (A) and (B) with λ unknown. Assume $\mathbf{B}, L, \Lambda_1 L, \dots, \Lambda_s L$ are distributed independently having the following priors: $\mathbf{B} \sim U(\mathbb{R}^p)$, $L \sim \text{Gamma}(a_0/2, g_0/2)$; $a_0 \geq 0$, $g_0 \geq 0$, and $\Lambda_i L \sim \text{Gamma}(a_i/2, g_i/2)$; $a_i > 0$, $g_i \geq 0$ for $i = 1, \dots, s$. Then, assuming that $n + \sum_{i=0}^s g_i - p - 2 > 0$, the conditional distribution of $\mathbf{Y}^{(2)}$ given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$ and $\Lambda = \lambda$ is multivariate- t with location parameter $\mathbf{M} \mathbf{y}^{(1)}$, scale parameter $(n + \sum_{i=0}^s g_i - p)^{-1} [a_0 + \sum_{i=1}^s a_i \lambda_i + \mathbf{y}^{(1)'} \mathbf{K} \mathbf{y}^{(1)}] \mathbf{Q}$, and degrees of freedom $\nu = n + \sum_{i=0}^s g_i - p$, where $\mathbf{K}, \mathbf{M}$ and $\mathbf{Q}$ are specified in Theorem 5.1. Also, conditional on $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$, Λ has p.d.f.

$$\pi(\lambda | \mathbf{y}^{(1)}) \propto |\Sigma_{11}|^{-1/2} |\mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{X}^{(1)}|^{-1/2} \prod_{i=1}^s \lambda_i^{g_i/2-1} \\ \cdot [a_0 + \sum_{i=1}^s a_i \lambda_i + \mathbf{y}^{(1)'} \mathbf{K} \mathbf{y}^{(1)}]^{-(n + \sum_{i=0}^s g_i - p)/2}.$$

Proof (Sketch) We proceed along the lines of the proof of Theorem 5.1, noting that

$$\pi(\lambda | \mathbf{y}^{(1)}) \propto \int_{\ell} \int_{\beta} \pi(\mathbf{y}^{(1)} | \beta, \lambda, \ell) \pi(\beta) \pi(\lambda) \pi(\ell) d\beta d\ell$$

where $\pi(\lambda)$, the p.d.f. of Λ , can be found from $\pi(\lambda\ell)$, the p.d.f. of ΛL , via a transformation. Also

$$\pi(y^{(2)}|y^{(1)}, \lambda) \propto \int_{\ell} \int_{\beta} \pi(y^{(2)}|y^{(1)}, \beta, \lambda, \ell) \pi(y^{(1)}|\beta, \lambda, \ell) \pi(\beta) \pi(\lambda) \pi(\ell) d\beta d\ell.$$

Finally, one can find the conditional distribution of $Y^{(2)}$ given $Y^{(1)} = y^{(1)}$ by taking

$$\pi(y^{(2)}|y^{(1)}) = \int_{\lambda} \pi(y^{(2)}|y^{(1)}, \lambda) \pi(\lambda|y^{(1)}) d\lambda.$$

However, to evaluate this last integral requires numerical methods which become increasingly difficult as the dimension of λ increases. ■

Now, for model (5.2), we can simplify the expression for $\pi(\lambda|y^{(1)})$ given in Theorem 5.2. To do so, we require an explicit expression for the determinant of Σ_{11} , which we find via the following result:

Lemma 5.2 (Kleffe and Rao (1986)) *If a matrix has the form $V = I + \alpha_1 U U' + \alpha_0 11'$ where α_0 and α_1 are known constants and $U'U$ has the form $U'U = \text{Diag}(r_j)$, then the determinant of V is given by*

$$|V| = (1 + \alpha_0 \sum_{j=1}^m r_j (1 + \alpha_1 r_j)^{-1}) \prod_{j=1}^m (1 + \alpha_1 r_j).$$

Now, we note that $|\Sigma_{11}| = \prod_{i=1}^t |\Sigma_{11}^{(i)}|$ and that $\Sigma_{11}^{(i)}$ can be written as

$$\Sigma_{11}^{(i)} = I_{n_i} + \lambda_2^{-1} (\text{Diag}(1_{n_{ij}})) (\text{Diag}(1_{n_{ij}}))' + \lambda_1^{-1} 1_{n_i} 1_{n_i}'.$$

We apply Lemma 5.2 to $\Sigma_{11}^{(i)}$ with $\alpha_0 = \lambda_1^{-1}$, $\alpha_1 = \lambda_2^{-1}$ and $U = \text{Diag}(1_{n_{ij}})_{1 \leq j \leq m_i}$ and obtain $U'U = \text{Diag}(n_{ij})_{1 \leq j \leq m_i}$ and thus

$$|\Sigma_{11}| = [\prod_{i=1}^t (1 - \gamma_i)^{-1}] [\prod_{i=1}^t \prod_{j=1}^{m_i} (1 - \gamma_{ij})^{-1}].$$

In view of the above, for model (5.2), $\pi(\lambda|y^{(1)})$ simplifies to

$$\begin{aligned} \pi(\lambda|y^{(1)}) &\propto \left[\prod_{i=1}^I (1 - \gamma_i)^{1/2} \right] \left[\prod_{i=1}^I \prod_{j=1}^{m_i} (1 - \gamma_{ij})^{1/2} \right] \lambda_1^{g_1/2-1} \lambda_2^{g_2/2-1} |X^{(1)'} \Sigma_{11}^{-1} X^{(1)}|^{-1/2} \\ &\cdot [a_0 + a_1 \lambda_1 + a_2 \lambda_2 + y^{(1)'} K y^{(1)}]^{-(n+g_0+g_1+g_2-p)/2} \end{aligned} \quad (5.20)$$

where $y^{(1)'} K y^{(1)}$ is given by (5.18), and $X^{(1)'} \Sigma_{11}^{-1} X^{(1)}$ is given by (5.13).

Again, we find the posterior expectation $E[(Y_{1..}, \dots, Y_{t..})' | Y^{(1)} = y^{(1)}]$, this time assuming λ to be unknown, by invoking Theorem 5.2 and noting that

$$\begin{aligned} E[(Y_{1..}, \dots, Y_{t..})' | Y^{(1)} = y^{(1)}] &= A y^{(1)} + C E(Y^{(2)} | y^{(1)}) \\ &= A y^{(1)} + C E_{\lambda|y^{(1)}} E(Y^{(2)} | y^{(1)}, \lambda) \\ &= A y^{(1)} + C E_{\lambda|y^{(1)}} (M y^{(1)}) \\ &= A y^{(1)} + E(C M y^{(1)} | y^{(1)}). \end{aligned}$$

Note that we have the expressions for the i^{th} elements of $A y^{(1)}$ and $C M y^{(1)}$ from (5.15) in Section 5.2, and thus we can write

$$\begin{aligned} E[N_i^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk} | Y^{(1)} = y^{(1)}] &= N_i^{-1} \left\{ \sum_{j=1}^{m_i} n_{ij} y_{ij.} + \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \mathbf{x}_{ij}' E(\tilde{\beta} | y^{(1)}) \right. \\ &+ \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) E[\gamma_{ij}(\bar{y}_{ij.} - \bar{\mathbf{x}}_{ij}' \tilde{\beta}) | y^{(1)}] \\ &+ E[\gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} (\sum_{j=1}^{m_i} (N_{ij} - n_{ij})(1 - \gamma_{ij})) \sum_{j=1}^{m_i} \gamma_{ij}(\bar{y}_{ij.} - \mathbf{x}_{ij}' \tilde{\beta}) | y^{(1)}] \\ &+ (\sum_{j=1}^{m_i} N_{ij} \bar{\mathbf{X}}_{ij.} - \sum_{j=1}^{m_i} N_{ij} \bar{\mathbf{X}}_{ij.})' E(\tilde{\beta} | y^{(1)}) \\ &\left. + (\sum_{j=1}^{m_i} N_{ij} - \sum_{j=1}^{m_i} N_{ij}) E[\gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij}(\bar{y}_{ij.} - \mathbf{x}_{ij}' \tilde{\beta}) | y^{(1)}] \right\}. \end{aligned} \quad (5.21)$$

Similarly, we can find the posterior variance $\text{var}[(Y_{1..}, \dots, Y_{t..})' | Y^{(1)} = y^{(1)}]$ by invoking Theorem 5.2 and noting that

$$\begin{aligned} \text{var}[(Y_{1..}, \dots, Y_{t..})' | Y^{(1)} = y^{(1)}] &= C \text{var}(Y^{(2)} | y^{(1)}) C' \\ &= C \text{var}_{\lambda|y^{(1)}} [E(Y^{(2)} | y^{(1)}, \lambda)] C' + C E_{\lambda|y^{(1)}} [\text{var}(Y^{(2)} | y^{(1)}, \lambda)] C' \\ &= C \text{var}_{\lambda|y^{(1)}} [M y^{(1)}] C' + (n + \sum_{i=0}^I g_i - p - 2)^{-1} C E_{\lambda|y^{(1)}} [(a_0 + \sum_{i=1}^I a_i \lambda_i + y^{(1)'} K y^{(1)}) Q] C' \\ &= \text{var}[C M y^{(1)} | y^{(1)}] + (n + \sum_{i=0}^I g_i - p - 2)^{-1} E[(a_0 + \sum_{i=1}^I a_i \lambda_i + y^{(1)'} K y^{(1)}) C Q C' | y^{(1)}]. \end{aligned}$$

We have the expressions for the i^{th} elements of $\mathbf{CMy}^{(1)}$ and $\mathbf{CQC}'$ from (5.15) and (5.19) in Section 5.2, and thus we can write

$$\begin{aligned}
\text{var}[N_i^{-1} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}] &= N_i^{-2} \{ \text{var}[\{ \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \bar{\mathbf{x}}_{ij}^* \tilde{\beta} \\
&+ \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \gamma_{ij} (\bar{y}_{ij} - \bar{\mathbf{x}}_{ij}' \tilde{\beta}) + \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} (\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (1 - \gamma_{ij})) \sum_{j=1}^{m_i} \gamma_{ij} (\bar{y}_{ij} - \bar{\mathbf{x}}_{ij}' \tilde{\beta}) \\
&+ (\sum_{j=1}^{M_i} N_{ij} \mathbf{X}_{ij} - \sum_{j=1}^{m_i} N_{ij} \bar{\mathbf{X}}_{ij})' \tilde{\beta} + (\sum_{j=1}^{M_i} N_{ij} - \sum_{j=1}^{m_i} N_{ij}) \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} (\bar{y}_{ij} - \bar{\mathbf{x}}_{ij}' \tilde{\beta}) | \mathbf{y}^{(1)}] \} \\
&+ (n + g_0 + g_1 + g_2 - p - 2)^{-1} E\{(a_0 + a_1 \lambda_1 + a_2 \lambda_2 + \mathbf{y}^{(1)'} \mathbf{K} \mathbf{y}^{(1)}) \\
&\cdot \{ N_i^{-2} [\lambda_1^{-1} (1 - \gamma_i) ((N_i - n_i) - \sum_{j=1}^{m_i} (N_{ij} - n_{ij}) \gamma_{ij})^2 + \lambda_2^{-1} \sum_{j=1}^{m_i} (N_{ij} - n_{ij})^2 (1 - \gamma_{ij}) \\
&+ \lambda_2^{-1} \sum_{j=1}^{M_i} N_{ij}^2 - \lambda_2^{-1} \sum_{j=1}^{m_i} N_{ij}^2 + (N_i - n_i)] + N_i^{-2} [\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (\bar{\mathbf{x}}_{ij}^* - \gamma_{ij} \bar{\mathbf{x}}_{ij}) \\
&+ \sum_{j=1}^{M_i} N_{ij} (\mathbf{X}_{ij} - \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}) - \sum_{j=1}^{m_i} N_{ij} (\bar{\mathbf{X}}_{ij} - \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}) \\
&- \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} (\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (1 - \gamma_{ij})) \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}]' (\mathbf{X}^{(1)'} \Sigma_{11}^{-1} \mathbf{X}^{(1)})^{-1} \\
&\cdot [\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (\bar{\mathbf{x}}_{ij}^* - \gamma_{ij} \bar{\mathbf{x}}_{ij}) + \sum_{j=1}^{M_i} N_{ij} (\bar{\mathbf{X}}_{ij} - \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}) \\
&- \sum_{j=1}^{m_i} N_{ij} (\mathbf{X}_{ij} - \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}) \\
&- \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} (\sum_{j=1}^{m_i} (N_{ij} - n_{ij}) (1 - \gamma_{ij})) \sum_{j=1}^{m_i} \gamma_{ij} \bar{\mathbf{x}}_{ij}] | \mathbf{y}^{(1)} \}. \tag{5.22}
\end{aligned}$$

The evaluation of (5.21) and (5.22) involves two-dimensional numerical integration, which can be implemented using one of the available packages. In the case where λ has a higher dimension, Datta and Ghosh (1990) suggest the use of the Gibbs sampler, as proposed by Gelfand and Smith (1990).

As a final note, in the absence of prior knowledge, we can assign diffuse prior Gamma distributions to L , $\Lambda_1 L$ and $\Lambda_2 L$. For instance, Datta and Ghosh (1990) use as a numerical illustration, the two-fold nested error mean model. They set $a_0 = g_0 = g_1 = g_2 = 0$, $a_1 = a_2 = .0005$, and perform a two-dimensional numerical integration to obtain the posterior expectation of the i^{th} small area mean. (They omit the evaluation of the posterior variance.)

CHAPTER 6

Summary and Proposed Future Research

This chapter is broken into two parts. In Section 6.1, the results of this thesis are discussed, while in Section 6.2, proposals are made for further research.

6.1 Summary

This thesis primarily deals with the prediction of small area means under one-fold and two-fold nested error regression models, with unequal error variances. Chapter 2 devotes itself entirely to the former model, while Chapter 3 devotes itself to the latter model. Both models are expressed as general mixed linear models, and best linear unbiased predictors (BLUPs) of the small area means are explicitly found, under the assumption of known variance components. Since, in practice, the variance components are unknown, they are estimated using Henderson's well-known method of fitting constants. Empirical best linear unbiased predictors (EBLUPs) are formed by substituting these estimators back into the expressions for the BLUPs. Since an explicit expression for the mean square error (MSE) of the EBLUPs is in most cases intractable, a second order approximation to the MSE is found, assuming normality. Using this approximation, an estimator of the MSE, having a bias of order $o(t^{-1})$ for large t , where t is the number of small areas, is also derived. As well in Chapter 3, for the two-fold model, a Monte Carlo study is undertaken, which investigates the finite sample accuracy of the second order approximation, and the bias of the estimator of the MSE. This study seems to indicate that the second order approximation is accurate when $Pr(\hat{\sigma}_\epsilon^2 < 0)$ is small, where $\hat{\sigma}_\epsilon^2$ is the unbiased estimator of the variance of the random small area effects in the two-fold model. However, the approximation can lead to a substantial overstatement of the true MSE when this probability is large.

This study also indicates that the estimator of the MSE, $\text{mse}[t_i(\hat{\sigma}^2, y)]$, slightly overestimates the true MSE when $Pr(\hat{\sigma}_\xi^2 < 0)$ is small, and grossly overestimates the true MSE for large such probabilities. On the other hand, $\text{mse}_1[t_i(\hat{\sigma}^2, y)]$, the commonly used estimator which ignores the uncertainty in estimating the variance components, almost always underestimates the true MSE. A compromise is struck by forming an adaptive estimator, which employs $\text{mse}_1[t_i(\hat{\sigma}^2, y)]$ when $\hat{\sigma}_\xi^2/\hat{\sigma}_e^2$ is small, and $\text{mse}[t_i(\hat{\sigma}^2, y)]$ when $\hat{\sigma}_\xi^2/\hat{\sigma}_e^2$ is large, where $\hat{\sigma}_e^2$ is the estimator of the error variance. Not surprisingly, the adaptive estimator performs well in tracking the true MSE, particularly for $\sigma_\xi^2/\sigma_e^2 \geq \sigma_u^2/\sigma_e^2$, which is the region of interest practically speaking, where σ_u^2 is the variance of the PSU effects in the two-fold model.

In Chapter 4 of this thesis, jackknife estimators of the variances of, and the covariances between, estimators of the variance components are found, under the two-fold model with one auxiliary variable and constant error variances. These jackknife estimators are shown to be asymptotically consistent. In addition, jackknife estimators of nonlinear functions of $\sigma^2 = (\sigma_\xi^2, \sigma_u^2, \sigma_e^2)'$, the vector of variance components, are found, as well as corresponding jackknife variance estimators. A Monte Carlo study investigates the robustness of jackknife confidence intervals on the nonlinear functions, $\rho_1 = \sigma_u^2/\sigma_e^2$ and $\rho_2 = \sigma_\xi^2/\sigma_e^2$. Exploiting the invariance property of confidence intervals, a jackknife confidence interval on the variance-stabilizing log transformation given by $\theta_1 = \ln(1 + n\rho_1)$ is found. This interval is then converted into an interval on ρ_1 . The study indicates that the jackknife confidence interval behaves well when the data are normally distributed, and are fairly robust when the data are uniformly or double exponentially distributed, for nominal values of both 90% and 95%. On the other hand, using the traditional normality based confidence interval on ρ_1 , there is a tendency towards overcoverage

when the data are distributed uniformly, and a tendency towards undercoverage when the data are distributed double exponentially. For the second ratio, ρ_2 , a robust confidence interval is formed by jackknifing ρ_2 directly, since no suitable transformation on ρ_2 can be found. Here, the confidence interval seems to behave well under normality, but its performance is not so satisfactory under nonnormal distributions, particularly for large ρ_2 values, under the double exponential distribution. It may be noted in this context that no exact $(1 - \alpha)\%$ confidence interval on ρ_2 , independent of ρ_1 , exists, even under normality.

In Chapter 5, the two-fold nested-error regression model with equal error variances, is considered from a Hierarchical Bayes standpoint. Assuming the vector of hyperparameters, λ , to be known, an explicit expression for the posterior expectation of the small area means, given the data, is found, and is shown to be equivalent to the BLUP found in Chapter 3. The corresponding explicit expression for the posterior variance is also found. Then, assuming λ to be unknown, a prior distribution is assigned to it and the posterior expectation and variance of the small area means, given the data, are found. The expressions for these are not in closed-form, and thus require numerical methods to be evaluated.

6.2 Proposed Future Research

While this thesis addresses several problems associated with the two-fold nested error regression model, many areas remain to be considered for future research. Three such areas are discussed below.

First, it would be useful to extend the theory in this thesis to nested error nonlinear regression models of the form

$$y_{ijk} = f(x_{ijk}, \theta) + \xi_i + u_{ij} + e_{ijk}; i = 1, \dots, t; j = 1, \dots, m_i; k = 1, \dots, n_{ij}$$

where $f(x_{ijk}, \theta)$ is a nonlinear function of the auxiliary variable x_{ijk} , and some parameter θ . Results in nonlinear regression theory and variance component analysis exist, upon which could be drawn to aid in the development of this theory.

A second problem is motivated by Laird and Louis' (1987) use of bootstrap methods to construct empirical Bayes (EB) confidence intervals, for some simple cases. These methods could be extended to the two-fold nested error regression model.

A final important area to be considered for further research is the estimation of small area means under two-fold nested error models, when measurement errors are present in the x -variables. Extensive literature on measurement error models exists (see, for example, Fuller (1987)), which could be employed in the development of this area.

Appendix A

Lemma A.1 *Assuming the conditions of Theorem 3.2, we have that*

$$E[(\hat{\sigma}^2 - \sigma^2)' d_i(\sigma^2)]^2 = E[(\hat{\sigma}^2 - \sigma^2)' d_i(\sigma^2)]^2 + o(t^{-1}).$$

Proof We can write

$$\begin{aligned} E[(\hat{\sigma}^2 - \sigma^2)' d_i(\sigma^2)]^2 &= E\left[\sum_{m=1}^3 t'_{im} (\hat{\sigma}_m^2 - \sigma_m^2)\right]^2 \\ &= E\left\{\sum_{m=1}^3 t'_{im} (\hat{\sigma}_m^2 - \sigma_m^2) \right. \\ &\quad \left. - \left[\sum_{m=1}^3 t'_{im} (\hat{\sigma}_m^2 - \sigma_m^2) + \sum_{m=1}^3 t'_{im} \sigma_m^2 I(\hat{\sigma}_m^2 \leq 0)\right] \right\}^2 \\ &= E\left\{\sum_{m=1}^3 t'_{im} (\hat{\sigma}_m^2 - \sigma_m^2) - \sum_{m=1}^3 t'_{im} \hat{\sigma}_m^2 I(\hat{\sigma}_m^2 \leq 0)\right\}^2 \end{aligned}$$

where t'_{im} is the m^{th} element of $d_i(\sigma^2)$ and $I(\hat{\sigma}_m^2 \leq 0) = 1$ if $\hat{\sigma}_m^2 \leq 0$, and zero otherwise. Expanding the square, we have

$$\begin{aligned} E[(\hat{\sigma}^2 - \sigma^2)' d_i(\sigma^2)]^2 &= E[(\hat{\sigma}^2 - \sigma^2)' d_i(\sigma^2)]^2 \\ &\quad + E\left\{\sum_{m=1}^3 t'_{im} \hat{\sigma}_m^2 I(\hat{\sigma}_m^2 \leq 0)\right\}^2 \\ &\quad - 2 \sum_{m=1}^3 \sum_{m'=1}^3 E\{t'_{im} t'_{im'} (\hat{\sigma}_{m'}^2 - \sigma_{m'}^2) \hat{\sigma}_m^2 I(\hat{\sigma}_m^2 \leq 0)\}. \end{aligned}$$

It suffices to show that the last two terms have orders $o(t^{-1})$.

Applying Minkowski's inequality followed by Holder's inequality twice to the first neglected term with $r = s = 2$ each time (see Lemma 3.4), we have

$$\begin{aligned} &E\left\{\sum_{m=1}^3 t'_{im} I(\hat{\sigma}_m^2 \leq 0) \hat{\sigma}_m^2\right\}^2 \\ &\leq \left\{\sum_{m=1}^3 [E|t'_{im}|^8 E|I(\hat{\sigma}_m^2 \leq 0)|^8]^{1/8} [E|\hat{\sigma}_m^2|^4]^{1/4}\right\}^2 = o(t^{-1}). \end{aligned}$$

We argue this as follows: $E|t'_{im}|^8 = O(1)$ from the body of the proof of Theorem 3.2. In addition,

$$E|\hat{\sigma}_m^2|^4 = E|\hat{\sigma}_m^2 - \sigma_m^2 + \sigma_m^2|^4 \leq 2^3 \{E|\hat{\sigma}_m^2 - \sigma_m^2|^4 + E|\sigma_m^2|^4\} = O(1)$$

upon applying the C_r inequality and noting that $E|\sigma_m^2|^4 = O(1)$ while $E|\tilde{\sigma}_m^2 - \sigma_m^2|^4 = O(t^{-2})$ by Lemma A.3 with $s = 2$. Finally, $E|I(\tilde{\sigma}_m^2 \leq 0)|^6 = E|I(\tilde{\sigma}_m^2 \leq 0)| = O(t^{-6})$ by Lemma A.2(b) with $s = 8$.

Now, addressing the second neglected term, we have

$$\begin{aligned} & E\{t'_{im}t'_{im'}(\tilde{\sigma}_{m'}^2 - \sigma_{m'}^2)\tilde{\sigma}_m^2 I(\tilde{\sigma}_m^2 \leq 0)\} \\ & \leq \{E|t'_{im}|^8 E|t'_{im'}|^8\}^{1/8} \{E|\tilde{\sigma}_m^2|^4 E|I(\tilde{\sigma}_m^2 \leq 0)|^4 E|\tilde{\sigma}_{m'}^2 - \sigma_{m'}^2|^4\}^{1/4} = o(t^{-1}) \end{aligned}$$

using Holder's inequality successively four times with $r = s = 2$ each time, and upon noting that $E|\tilde{\sigma}_m^2 - \sigma_m^2|^4 = O(t^{-2})$ by Lemma A.3 with $s = 2$. In addition $E|I(\tilde{\sigma}_m^2 \leq 0)|^4 = E|I(\tilde{\sigma}_m^2 \leq 0)| = O(t^{-6})$ by Lemma A.2(b) with $s = 6$. ■

Lemma A.2 *Assuming the conditions of Theorem 3.2, we have that*

- (a) $\tilde{\sigma}_m^2 - \sigma_m^2 = o_p(t^{-\epsilon})$ for $0 < \epsilon < \frac{1}{2}$ and $m = 1, 2, 3$.
- (b) $E|I(\tilde{\sigma}_m^2 \leq 0)| = \Pr(\tilde{\sigma}_m^2 \leq 0) = O(t^{-s})$ for $m = 1, 2, 3$ and for any $s \in \mathbf{Z}^+$ such that $s > 1$.
- (c) $\tilde{\sigma}_m^2 - \sigma_m^2 = o_p(t^{-\epsilon})$ for $0 < \epsilon < \frac{1}{2}$ and $m = 1, 2, 3$.

Proof (a) Applying Markov's inequality and Holder's inequality successively, we get

$$\Pr[t'|\tilde{\sigma}_m^2 - \sigma_m^2| > \gamma] \leq t' E|\tilde{\sigma}_m^2 - \sigma_m^2|/\gamma \leq t^\epsilon (\text{var}(\tilde{\sigma}_m^2))^{1/2}/\gamma.$$

Using the explicit expressions from Chapter 11 of Scarle (1971), it is easy to verify that $\text{var}(\tilde{\sigma}_m^2) = O(t^{-1})$ for $m = 1, 2, 3$. Therefore $\Pr[t'|\tilde{\sigma}_m^2 - \sigma_m^2| > \gamma] \rightarrow 0$ for $0 < \epsilon < \frac{1}{2}$ as $t \rightarrow \infty$.

(b) We have

$$\begin{aligned} \Pr(\tilde{\sigma}_m^2 \leq 0) &= \Pr(\tilde{\sigma}_m^2 - \sigma_m^2 \leq -\sigma_m^2) \leq \Pr(|\tilde{\sigma}_m^2 - \sigma_m^2| \geq \sigma_m^2) \\ &= \Pr(|\tilde{\sigma}_m^2 - \sigma_m^2|^{2s} \geq (\sigma_m^2)^{2s}) \leq E|\tilde{\sigma}_m^2 - \sigma_m^2|^{2s}/(\sigma_m^2)^{2s}, \end{aligned}$$

for $s > 1$ by Markov's inequality. However, $E|\hat{\sigma}_m^2 - \sigma_m^2|^{2s} = O(t^{-s})$ by Lemma A.3 and thus we have the result.

(c) Writing $\hat{\sigma}_m^2 - \sigma_m^2 = (\hat{\sigma}_m^2 - \tilde{\sigma}_m^2) + (\tilde{\sigma}_m^2 - \sigma_m^2)$ and using part (a), it is sufficient to show that $\hat{\sigma}_m^2 - \tilde{\sigma}_m^2 = o_p(t^{-\epsilon})$. Now by Markov's inequality for $\gamma > 0$ we have,

$$Pr\{t^\epsilon |\hat{\sigma}_m^2 - \tilde{\sigma}_m^2| > \gamma\} \leq t^\epsilon E|\hat{\sigma}_m^2 - \tilde{\sigma}_m^2|/\gamma \rightarrow 0$$

since

$$E|\hat{\sigma}_m^2 - \tilde{\sigma}_m^2| = o(t^{-1}).$$

This latter result is obtained by observing that

$$E|\hat{\sigma}_m^2 - \tilde{\sigma}_m^2| = E\{|\hat{\sigma}_m^2|I(\hat{\sigma}_m^2 \leq 0)\} \leq (E|\hat{\sigma}_m^2|^2)^{\frac{1}{2}}(E|I(\hat{\sigma}_m^2 \leq 0)|^2)^{\frac{1}{2}}$$

via Holder's inequality. Now

$$E|\hat{\sigma}_m^2|^2 = E|\hat{\sigma}_m^2 - \sigma_m^2 + \sigma_m^2|^2 \leq 2\{E|\hat{\sigma}_m^2 - \sigma_m^2|^2 + E|\sigma_m^2|^2\} = O(1)$$

upon applying the C_r inequality and noting that $E|\sigma_m^2|^2 = O(1)$ while $E|\hat{\sigma}_m^2 - \sigma_m^2|^2 = \text{var}(\hat{\sigma}_m^2) = O(t^{-1})$ using the expressions for $\text{var}(\hat{\sigma}_m^2)$; $m = 1, 2, 3$ from Chapter 11 of Searle (1971). Finally,

$$E|I(\hat{\sigma}_m^2 \leq 0)|^2 = E|I(\hat{\sigma}_m^2 \leq 0)| = O(t^{-4})$$

by part (b) with $s = 4$. Thus $(E|I(\hat{\sigma}_m^2 \leq 0)|^2)^{\frac{1}{2}} = o(t^{-1})$. ■

Lemma A.3 *Assuming the conditions of Theorem 3.2, we have that $E|\hat{\sigma}_m^2 - \sigma_m^2|^{2s} = O(t^{-s})$ for any $s \in \mathbf{Z}^+$ such that $s > 1$, and for $m = 1, 2, 3$ where $\hat{\sigma}_1^2 = \hat{\sigma}_\xi^2, \hat{\sigma}_2^2 = \hat{\sigma}_\mathbf{u}^2$ and $\hat{\sigma}_3^2 = \hat{\sigma}_\epsilon^2$.*

Proof To proceed, we require the following two lemmas:

Lemma A.3.1 (Rao and Kleffe (1988), pp. 39, #2.3.10) Suppose $\epsilon_i = U_i \eta_i$; $i = 1, \dots, t$ are independently distributed random vectors such that $E(\epsilon_i) = 0$, $E(\epsilon_i \epsilon_i') = \Sigma_i = U_i U_i'$, $E(\eta_i \eta_i') = I$ and $E(\eta_i' \eta_i)^{2s} < \infty$ for some $s > 1$. Then,

$$E|\epsilon' Q \epsilon - E(\epsilon' Q \epsilon)|^{2s} \leq K(s) \left\{ \max_{1 \leq i \leq t} E(\eta_i' \eta_i)^{2s} \right\} [\text{trace}(Q \Sigma Q \Sigma)]^s \\ \propto [\text{trace}(Q \Sigma Q \Sigma)]^s$$

where $K(s)$ is a known function of s and $Q = (Q_{kj})$ is a block matrix such that $\epsilon' Q \epsilon = \sum_k \sum_j \epsilon_k' Q_{kj} \epsilon_j$ and $\Sigma = \text{Diag}(\Sigma_i)$.

Lemma A.3.2 (Rao and Kleffe (1988), pp. 10, #1.5.1) If A and B are matrices, then

- (a) $\text{trace}(AB') = (\text{vec } A)'(\text{vec } B)$
- (b) $\text{vec}(AB') = (A \otimes I)(\text{vec } B') = (I \otimes B)(\text{vec } A)$

where $\otimes$ represents the Kronecker product operation on matrices. In addition, if A is an $m \times n$ matrix, then $\text{vec } A$ is the $mn \times 1$ vector formed by stacking the columns of A one below another, beginning with the leftmost column.

Now, appealing to Lemma A.3.1, and noting that all moments exist and are finite under the normality assumption of Theorem 3.2, and that each of the three variance component estimators, $\hat{\sigma}_\epsilon^2$, $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_\beta^2$, can be written in quadratic form as $\epsilon' Q \epsilon = y' Q y$ where $\epsilon = y - \beta$ and $\beta = \beta 1_n$, we need only show that $\text{trace}(Q \Sigma Q \Sigma) = O(t^{-1})$.

Invoking both parts of Lemma A.3.2, we see that

$$\begin{aligned} \text{trace}(Q \Sigma Q \Sigma) &= (\text{vec } Q \Sigma)'(\text{vec } \Sigma Q) \\ &= [(I \otimes \Sigma)(\text{vec } Q)]'[(\Sigma \otimes I)(\text{vec } Q)] \\ &= (\text{vec } Q)'(I \otimes \Sigma)'(\Sigma \otimes I)(\text{vec } Q) \\ &= (\text{vec } Q)'(I \otimes \Sigma)(\Sigma \otimes I)(\text{vec } Q) \\ &= (\text{vec } Q)'(\Sigma \otimes \Sigma)(\text{vec } Q), \end{aligned}$$

noting that for matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$, $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$ and that $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$. (See Theorem 8.8.4 and Theorem 8.8.6; Graybill (1969)).

Now, for the model in question, $\Sigma = \text{var}(\epsilon) = \text{var}(\mathbf{y} - \beta)$ is given by (3.4) in Chapter 3 with $f_{ijk} = 1 \forall (i, j, k)$, and can be written in the form $\Sigma = \text{Diag}(\Sigma_i)_{1 \leq i \leq t}$. Let λ_{ik} ; $k = 1, \dots, n_i$ be the (not necessarily distinct) eigenvalues of Σ_i . Then $\lambda_{i\max} = \sup_{1 \leq k \leq n_i} \lambda_{ik} < \infty$ since $\sum_{k=1}^{n_i} \lambda_{ik} = \text{trace}(\Sigma_i) = n_i(\sigma_e^2 + \sigma_u^2 + \sigma_\xi^2) < \infty$. Finally, $\lambda_{\max} = \sup_{1 \leq i \leq t} \lambda_{i\max} < \infty$ since $\lambda_{i\max} < \infty \forall i$. Note that if $\lambda_{\max}$ is the largest eigenvalue of Σ , then $\lambda_{\max}^2$ is the largest eigenvalue of $\Sigma \odot \Sigma$. (See Theorem 8.8.13; Graybill (1969)), and $\lambda_{\max}^2 < \infty$. Thus

$$\begin{aligned} \text{trace}(\mathbf{Q}\Sigma\mathbf{Q}\Sigma) &= (\text{vec } \mathbf{Q})'(\Sigma \odot \Sigma)(\text{vec } \mathbf{Q}) \\ &\leq \lambda_{\max}^2 (\text{vec } \mathbf{Q})'(\text{vec } \mathbf{Q}) \\ &= \lambda_{\max}^2 \text{trace}(\mathbf{Q}^2). \end{aligned}$$

We need only show that $\text{trace}(\mathbf{Q}^2) = O(t^{-1})$ for each of the three variance component estimators. We prove the results separately for each case.

(i) For the model defined in Theorem 3.2, we have the following expression for $\tilde{\sigma}_e^2$ from Chapter 11 of Searle (1971):

$$\tilde{\sigma}_e^2 = (n - \sum_{i=1}^t m_i)^{-1} (\sum_{i=1}^t \sum_{j=1}^{m_i} \sum_{k=1}^n y_{ijk}^2 - \sum_{i=1}^t \sum_{j=1}^{m_i} n_{ij} y_{ij.}^2)$$

where $y_{ij.} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} y_{ijk}$. We can write $\tilde{\sigma}_e^2$ in quadratic form, as follows:

$$\tilde{\sigma}_e^2 = \mathbf{y}' \mathbf{Q}_1 \mathbf{y}$$

where

$$\mathbf{Q}_1 = (n - \sum_i m_i)^{-1} [\mathbf{I}_n - \text{Diag}_{1 \leq i \leq t} \text{Diag}_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}_{n_{ij}}')].$$

Now

$$\begin{aligned}
\text{trace}(\mathbf{Q}_1^2) &= (n - \sum_i m_i)^{-2} \text{trace}[\mathbf{I}_n - \text{Diag}_{1 \leq i \leq t} \text{Diag}_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}})]^2 \\
&= (n - \sum_i m_i)^{-2} \text{trace} [\mathbf{I}_n - \text{Diag}_{1 \leq i \leq t} \text{Diag}_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}})] \\
&= (n - \sum_i m_i)^{-2} (n - \sum_i m_i) = (n - \sum_i m_i)^{-1} \\
&\leq (2 \sum_i m_i - \sum_i m_i)^{-1} = (\sum_i m_i)^{-1} \leq (2t)^{-1}
\end{aligned}$$

since $n = \sum_i \sum_j n_{ij} \geq 2 \sum_i m_i$ and $\sum_i m_i \geq 2t \forall (i, j)$. Thus $\text{trace}(\mathbf{Q}_1^2) = O(t^{-1})$.

(ii) For the model defined in Theorem 3.2, we have the following expression for $\hat{\sigma}_u^2$ from Chapter 11 of Searle (1971):

$$\hat{\sigma}_u^2 = (n - \sum_{i=1}^t \sum_{j=1}^{m_i} n_{ij}^2 n_i^{-1})^{-1} [(\sum_{i=1}^t \sum_{j=1}^{m_i} n_{ij} \bar{y}_{ij}^2 - \sum_{i=1}^t n_i \bar{y}_{i..}^2) - (\sum_{i=1}^t m_i - t) \hat{\sigma}_e^2]$$

where $\bar{y}_{i..} = n_i^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} y_{ijk}$. We can write $\hat{\sigma}_u^2$ in quadratic form, as follows:

$$\hat{\sigma}_u^2 = \mathbf{y}' \mathbf{Q}_2 \mathbf{y}$$

where

$$\begin{aligned}
\mathbf{Q}_2 &= (n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-1} \{ \text{Diag}_{1 \leq i \leq t} \text{Diag}_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}}) - \text{Diag}_{1 \leq i \leq t} (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) \\
&\quad - (\sum_i m_i - t) \mathbf{Q}_1 \}.
\end{aligned}$$

Now

$$\begin{aligned}
\text{trace}(\mathbf{Q}_2^2) &= (n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-2} \{ \text{trace}[\text{Diag}_{1 \leq i \leq t} \text{Diag}_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}}) \\
&\quad - \text{Diag}_{1 \leq i \leq t} (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i})]^2 + (\sum_i m_i - t)^2 \text{trace}(\mathbf{Q}_1^2) \}
\end{aligned}$$

since the crossproduct term is zero, i.e.,

$$[\text{Diag}_{1 \leq i \leq t} \text{Diag}_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}}) - \text{Diag}_{1 \leq i \leq t} (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i})] \mathbf{Q}_1 = 0.$$

Now

$$\begin{aligned} & \text{trace} [\text{Diag}_{1 \leq i \leq t} \text{Diag}_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}}) - \text{Diag}_{1 \leq i \leq t} (n_i^{-1} \mathbf{1}_n \mathbf{1}'_n)]^2 \\ &= \text{trace} [\text{Diag}_{1 \leq i \leq t} \text{Diag}_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}}) - \text{Diag}_{1 \leq i \leq t} (n_i^{-1} \mathbf{1}_n \mathbf{1}'_n)] \\ &= \sum_i m_i - t. \end{aligned}$$

Thus $\text{trace} (\mathbf{Q}_2^2) = O(t^{-1})$ if

$$(n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-2} (\sum_i m_i - t)^2 = O(1)$$

or equivalently

$$(n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-2} (\sum_i m_i - t) = O(t^{-1}).$$

We show this as follows:

$$\begin{aligned} n - \sum_i \sum_j n_{ij}^2 n_i^{-1} &= \sum_i \sum_j n_{ij} (1 - n_{ij} n_i^{-1}) \geq 2 \sum_i \sum_j (1 - n_{ij} n_i^{-1}) \\ &= 2[\sum_i m_i - \sum_i n_i^{-1} \sum_j n_{ij}] = 2[\sum_i m_i - t] \end{aligned}$$

and so

$$(n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-2} (\sum_i m_i - t)^2 \leq 1/4$$

or

$$(n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-2} (\sum_i m_i - t)^2 = O(1).$$

(iii) For the model defined in Theorem 3.2, we have the following expression for $\hat{\sigma}_\xi^2$ from Chapter 11 of Searle (1971):

$$\begin{aligned} \hat{\sigma}_\xi^2 &= (n - \sum_{i=1}^t n_i^2 n^{-1})^{-1} \{ (\sum_{i=1}^t n_i y_{i...}^2 - n y_{...}^2) - (\sum_{i=1}^t \sum_{j=1}^{m_i} n_{ij}^2 n_i^{-1} - \sum_{i=1}^t \sum_{j=1}^{m_i} n_{ij}^2 n^{-1}) \hat{\sigma}_u^2 \\ &\quad - (t-1) \hat{\sigma}_c^2 \} \end{aligned}$$

where $y_{...} = n^{-1} \sum_{i=1}^t \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} y_{ijk}$. We can write $\hat{\sigma}_\xi^2$ in quadratic form, as follows:

$$\hat{\sigma}_\xi^2 = \mathbf{y}' \mathbf{Q}_3 \mathbf{y}$$

where

$$\begin{aligned} \mathbf{Q}_3 &= (n - \sum_i n_i^2 n^{-1})^{-1} \{ [\text{Diag } (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) - n^{-1} \mathbf{1}_n \mathbf{1}'_n] \\ &\quad - (\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1}) \mathbf{Q}_2 - (t-1) \mathbf{Q}_1 \}. \end{aligned}$$

Now

$$\begin{aligned} \text{trace } (\mathbf{Q}_3^2) &= (n - \sum_i n_i^2 n^{-1})^{-2} \{ \text{trace } [\text{Diag } (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) - n^{-1} \mathbf{1}_n \mathbf{1}'_n]^2 \\ &\quad + (\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1})^2 \text{trace } (\mathbf{Q}_2^2) + (t-1)^2 \text{trace } (\mathbf{Q}_1^2) \\ &\quad + 2(t-1)(\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1}) \text{trace } (\mathbf{Q}_1 \mathbf{Q}_2) \} \end{aligned}$$

since two of the crossproduct terms are zero, i.e.,

$$[\text{Diag } (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) - n^{-1} \mathbf{1}_n \mathbf{1}'_n] \mathbf{Q}_1 = \mathbf{0}$$

and

$$[\text{Diag } (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) - n^{-1} \mathbf{1}_n \mathbf{1}'_n] \mathbf{Q}_2 = \mathbf{0}.$$

Now

$$\begin{aligned} &\text{trace } [\text{Diag } (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) - n^{-1} \mathbf{1}_n \mathbf{1}'_n]^2 \\ &= \text{trace } [\text{Diag } (n_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) - n^{-1} \mathbf{1}_n \mathbf{1}'_n] \\ &= t - 1. \end{aligned}$$

Also

$$\begin{aligned} \text{trace } (\mathbf{Q}_1 \mathbf{Q}_2) &= -(n - \sum_i m_i)^{-2} (n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-1} (\sum_i m_i - t) \\ &\quad \cdot \text{trace } [\mathbf{I}_n - \text{Diag }_{1 \leq i \leq t} \text{Diag }_{1 \leq j \leq m_i} (n_{ij}^{-1} \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}})] \\ &= -(n - \sum_i m_i)^{-1} (n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-1} (\sum_i m_i - t). \end{aligned}$$

Thus $\text{trace } (\mathbf{Q}_3^2) = O(t^{-1})$ if

$$(n - \sum_i n_i^2 n^{-1})^{-2} (t-1)^2 = O(1)$$

or equivalently

$$(n - \sum_i n_i^2 n^{-1})^{-2} (t-1) = O(t^{-1}),$$

if

$$(n - \sum_i n_i^2 n^{-1})^{-2} (\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1})^2 = O(1)$$

and if

$$(n - \sum_i m_i)^{-1} (n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-1} (\sum_i m_i - t)(t - 1) (\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1}) \\ \cdot (n - \sum_i n_i^2 n^{-1})^{-2} = O(t^{-1}).$$

We show these as follows:

First,

$$n - \sum_i n_i^2 n^{-1} = n^{-1} \sum_i n_i (n - n_i) \geq 4n^{-1} (nt - n) = 4(t - 1),$$

and so

$$(n - \sum_i n_i^2 n^{-1})^{-2} (t - 1)^2 \leq 1/16$$

or

$$(n - \sum_i n_i^2 n^{-1})^{-2} (t - 1)^2 = O(1).$$

Next, we have that $\sum_i \sum_j n_{ij}^2 n_i^{-1} \leq \sum_i \sum_j n_{ij}^2 n_{ij}^{-1} = n$ since $n_{ij} \leq n_i \quad \forall (i, j)$, and that $\sum_i \sum_j n_{ij}^2 n^{-1} \geq 2 \sum_i \sum_j n_{ij} n^{-1} = 2$ since $n_{ij} \geq 2 \quad \forall (i, j)$. Thus,

$$(n - \sum_i n_i^2 n^{-1})^{-2} (\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1})^2 \leq [(n - 2)/4(t - 1)]^2 \\ \leq [n/4(t - 1)]^2 \leq [tm_{\mathbf{u}} n_{\mathbf{u}}/4(t - 1)]^2 \leq [n_{\mathbf{u}} m_{\mathbf{u}}/2]^2$$

since $t/(t - 1) \leq 2$, or $(n - \sum_i n_i^2 n^{-1})^{-2} (\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1})^2 = O(1)$.

Finally,

$$(n - \sum_i m_i)^{-1} (n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-1} (n - \sum_i n_i^2 n^{-1})^{-2} (\sum_i m_i - t)(t - 1) \\ \cdot (\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1}) = O(t^{-1})$$

using the following facts:

$$(n - \sum_i n_i^2 n^{-1})^{-2} (t - 1)^2 = O(1),$$

$$(n - \sum_i n_i^2 n^{-1})^{-2} (\sum_i \sum_j n_{ij}^2 n_i^{-1} - \sum_i \sum_j n_{ij}^2 n^{-1})^2 = O(1),$$

$$(n - \sum_i \sum_j n_{ij}^2 n_i^{-1})^{-2} (\sum_i m_i - t)^2 = O(1)$$

and

$$(n - \sum_i m_i)^{-1} = O(t^{-1}).$$

■

Lemma A.4 Assuming the conditions of Theorem 3.2, we have that

$$|t'_{im}| = |\partial t_i(\hat{\sigma}^2, y) / \partial \hat{\sigma}_m^2|_{\hat{\sigma}_m^2 = \sigma_m^2} \leq a_m |y_{i\gamma}| + b_m t^{-1} \sum_{\gamma=1}^i |y_{i\gamma}|$$

where a_m and b_m are known constants for $m = 1, 2, 3$.

Proof For simplicity of notation in this proof, we take all derivatives with respect to σ_m^2 , but mean this to represent the derivatives taken with respect to $\hat{\sigma}_m^2$ and evaluated at $\hat{\sigma}_m^2 = \sigma_m^2$ for $m = 1, 2, 3$.

Now given that

$$t_i(\sigma^2, y) = (1 - \gamma_i) \tilde{\beta} + \gamma_i \tilde{y}_{i\gamma}$$

where

$$\tilde{\beta} = (\sum_{\gamma=1}^i \gamma_i)^{-1} \sum_{\gamma=1}^i \gamma_i \tilde{y}_{i\gamma}$$

and

$$\tilde{y}_{i\gamma} = (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} \sum_{j=1}^{m_i} \gamma_{ij} y_{ij},$$

we have that

$$t'_{im} = (1 - \gamma_i) \partial \tilde{\beta} / \partial \sigma_m^2 + (\tilde{y}_{i\gamma} - \tilde{\beta}) \partial \gamma_i / \partial \sigma_m^2 + \gamma_i \partial \tilde{y}_{i\gamma} / \partial \sigma_m^2.$$

Using the assumption that the m_i 's and n_{ij} 's are bounded below $\forall (i, j)$, we have that γ_{ij} and γ_i are bounded above and below $\forall (i, j)$, i.e.,

$$\begin{aligned} \inf_{1 \leq i \leq t} \inf_{1 \leq j \leq m_i} \gamma_{ij} &\geq \gamma_* = \sigma_u^2(\sigma_u^2 + \sigma_\epsilon^2/2)^{-1}, \\ \sup_{1 \leq i \leq t} \sup_{1 \leq j \leq m_i} \gamma_{ij} &\leq 1, \\ \inf_{1 \leq i \leq t} \gamma_i &\geq \gamma_{**} = \sigma_\xi^2(\sigma_\xi^2 + \sigma_u^2(2\gamma_*)^{-1})^{-1}, \end{aligned}$$

and

$$\sup_{1 \leq i \leq t} \gamma_i \leq 1.$$

We bound the component parts of $|l'_{im}|$ using the bounds on γ_i , γ_{ij} , m_i and n_{ij} as follows:

$$\begin{aligned} |\hat{\beta}| &= |(\sum_i \gamma_i)^{-1} (\sum_i \gamma_i y_{i\gamma})| \leq |(t\gamma_{**})^{-1} \sum_i y_{i\gamma}| \leq c_0 t^{-1} \sum_i |y_{i\gamma}|, \\ |\partial \gamma_i / \partial \sigma_\xi^2| &= |\sigma_u^2(\sigma_\xi^4)^{-1} \gamma_i^2 (\sum_j \gamma_{ij})^{-1}| \leq \sigma_u^2(\sigma_\xi^4)^{-1} (2\gamma_*)^{-1} = c_1, \\ |\partial \gamma_i / \partial \sigma_u^2| &= |-(\sigma_\xi^2)^{-1} \gamma_i^2 (\sum_j \gamma_{ij}^2) (\sum_j \gamma_{ij})^{-2}| \leq (\sigma_\xi^2)^{-1} m_u (2\gamma_*)^{-2} = c_2, \\ |\partial \gamma_i / \partial \sigma_\epsilon^2| &= |-(\sigma_\xi^2)^{-1} \gamma_i^2 (\sum_j n_{ij}^{-1} \gamma_{ij}^2) (\sum_j \gamma_{ij})^{-2}| \\ &\leq (\sigma_\xi^2)^{-1} m_u 2^{-1} (2\gamma_*)^{-2} = c_3, \end{aligned}$$

$$|\partial y_{i\gamma} / \partial \sigma_\xi^2| = 0 = d_1 |y_{i\gamma}| \text{ where } d_1 = 0,$$

$$\begin{aligned} |\partial y_{i\gamma} / \partial \sigma_u^2| &= |\sigma_\epsilon^2(\sigma_u^4)^{-1} (\sum_j \gamma_{ij})^{-1} \sum_j n_{ij}^{-1} \gamma_{ij}^2 (y_{ij} - y_{i\gamma})| \\ &\leq \sigma_\epsilon^2(\sigma_u^4)^{-1} 2^{-1} |(\sum_j \gamma_{ij})^{-1} \sum_j \gamma_{ij} y_{ij}| \\ &\quad + \sigma_\epsilon^2(\sigma_u^4)^{-1} (2\gamma_*)^{-1} 2^{-1} m_u |\bar{y}_{i\gamma}| = d_2 |y_{i\gamma}|, \end{aligned}$$

and

$$\begin{aligned} |\partial \bar{y}_{i\gamma} / \partial \sigma_\epsilon^2| &= |-\sigma_u^2(\sigma_\epsilon^2)^{-1} \partial \bar{y}_{i\gamma} / \partial \sigma_u^2| \leq \sigma_u^2(\sigma_\epsilon^2)^{-1} |\partial y_{i\gamma} / \partial \sigma_u^2| \leq \sigma_u^2(\sigma_\epsilon^2)^{-1} d_2 |y_{i\gamma}| \\ &= d_3 |y_{i\gamma}|. \end{aligned}$$

Therefore

$$|\dot{\beta}| \leq c_0 t^{-1} \sum_i |\dot{y}_{i\gamma}|.$$

$$|\partial \gamma_i / \partial \sigma_m^2| \leq c_m$$

and

$$|\partial y_{i\gamma} / \partial \sigma_m^2| \leq d_m |y_{i\gamma}|$$

for $m = 1, 2, 3$ where $d_1 = 0$. Finally,

$$\begin{aligned} |\partial \dot{\beta} / \partial \sigma_m^2| &= |(\sum_i \gamma_i)^{-2} \{ (\sum_i \dot{y}_{i\gamma} \partial \gamma_i / \partial \sigma_m^2) (\sum_i \gamma_i) \\ &\quad + (\sum_i \gamma_i \partial \dot{y}_{i\gamma} / \partial \sigma_m^2) (\sum_i \gamma_i) \\ &\quad - (\sum_i \partial \gamma_i / \partial \sigma_m^2) (\sum_i \gamma_i \dot{y}_{i\gamma}) \}| \\ &\leq (\sum_i \gamma_i)^{-2} \{ (\sum_i |\dot{y}_{i\gamma}| |\partial \gamma_i / \partial \sigma_m^2|) (\sum_i \gamma_i) \\ &\quad + (\sum_i \gamma_i |\partial y_{i\gamma} / \partial \sigma_m^2|) (\sum_i \gamma_i) + (\sum_i |\partial \gamma_i / \partial \sigma_m^2|) (\sum_i \gamma_i |\dot{y}_{i\gamma}|) \} \\ &\leq (t \gamma_{**})^{-2} \{ t c_m \sum_i |\dot{y}_{i\gamma}| + t d_m \sum_i |\dot{y}_{i\gamma}| + t c_m \sum_i |\dot{y}_{i\gamma}| \} = f_m t^{-1} \sum_i |\dot{y}_{i\gamma}| \end{aligned}$$

say, or

$$|\partial \dot{\beta} / \partial \sigma_m^2| \leq f_m t^{-1} \sum_i |\dot{y}_{i\gamma}| \text{ for } m = 1, 2, 3.$$

We conclude the proof by noting that

$$\begin{aligned} |t'_{im}| &\leq (1 - \gamma_i) |\partial \dot{\beta} / \partial \sigma_m^2| + |\partial \gamma_i / \partial \sigma_m^2| |\dot{y}_{i\gamma}| + |\partial \gamma_i / \partial \sigma_m^2| |\dot{\beta}| + |\gamma_i| |\partial \dot{y}_{i\gamma} / \partial \sigma_m^2| \\ &\leq f_m t^{-1} \sum_i |\dot{y}_{i\gamma}| + c_m |\dot{y}_{i\gamma}| + c_0 c_m t^{-1} \sum_i |\dot{y}_{i\gamma}| + d_m |\dot{y}_{i\gamma}| \\ &= a_m |\dot{y}_{i\gamma}| + b_m t^{-1} \sum_i |\dot{y}_{i\gamma}| \end{aligned}$$

as required. ■

Lemma A.5 Assuming the conditions of Theorem 3.2, we have that

$$|t''_{i\alpha}| = |\partial^2 t_i(\hat{\sigma}^2, \mathbf{y}) / \partial \hat{\sigma}_i^2 \partial \hat{\sigma}_\alpha^2|_{\substack{\hat{\sigma}_i^2 = \sigma_i^2 \\ \hat{\sigma}_\alpha^2 = \sigma_\alpha^2}} \leq a_{i\alpha} |\dot{y}_{i\gamma}| + b_{i\alpha} t^{-1} \sum_{j=1}^I |\dot{y}_{j\gamma}|$$

where $a_{\ell s}$ and $b_{\ell s}$ are known constants for $\ell, s = 1, 2, 3$.

Proof For simplicity of notation in this proof, we take all derivatives with respect to σ_ℓ^2 , but mean this to represent the derivatives taken with respect to $\hat{\sigma}_\ell^2$ and evaluated at $\hat{\sigma}_\ell^2 = \sigma_\ell^2$ for $\ell = 1, 2, 3$. Equivalent notation is used for second order derivatives.

Now,

$$\begin{aligned} t''_{\ell s} = & (1 - \gamma_i)(\partial^2 \tilde{\beta} / \partial \sigma_\ell^2 \partial \sigma_s^2) - (\partial \tilde{\beta} / \partial \sigma_\ell^2)(\partial \gamma_i / \partial \sigma_s^2) + (y_{i\gamma} - \tilde{\beta})(\partial^2 \gamma_i / \partial \sigma_\ell^2 \partial \sigma_s^2) \\ & + (\partial y_{i\gamma} / \partial \sigma_s^2 - \partial \tilde{\beta} / \partial \sigma_s^2) \partial \gamma_i / \partial \sigma_\ell^2 + (\partial \gamma_i / \partial \sigma_s^2)(\partial y_{i\gamma} / \partial \sigma_\ell^2) + \gamma_i \partial^2 y_{i\gamma} / \partial \sigma_\ell^2 \partial \sigma_s^2. \end{aligned}$$

From Lemma A.4, we have that $|\tilde{\beta}| \leq c_\sigma t^{-1} \sum |\dot{y}_{i\gamma}|$, $|\partial \gamma_i / \partial \sigma_m^2| \leq c_m$, $|\partial y_{i\gamma} / \partial \sigma_m^2| \leq d_m |\dot{y}_{i\gamma}|$ and $|\partial \tilde{\beta} / \partial \sigma_m^2| \leq f_m t^{-1} \sum |\dot{y}_{i\gamma}|$ for $m = 1, 2, 3$. We bound the component parts of $t''_{\ell s}$ using the bounds on $\gamma_i, \gamma_{ij}, m_i$ and n_{ij} as follows:

$$\begin{aligned} |\partial \gamma_{ij} / \partial \sigma_\xi^2| &= 0 = g_1, \\ |\partial \gamma_{ij} / \partial \sigma_u^2| &= |\sigma_\epsilon^2 (\sigma_u^4)^{-1} \gamma_{ij}^2 n_{ij}^{-1}| \leq \sigma_\epsilon^2 (\sigma_u^4)^{-1} 2^{-1} = g_2, \\ |\partial \gamma_{ij} / \partial \sigma_\epsilon^2| &= |-\sigma_u^2 (\sigma_\epsilon^2)^{-1} \partial \gamma_{ij} / \partial \sigma_u^2| \leq \sigma_u^2 (\sigma_\epsilon^2)^{-1} |\partial \gamma_{ij} / \partial \sigma_u^2| \leq \sigma_u^2 (\sigma_\epsilon^2)^{-1} g_2 = g_3, \\ |\partial^2 y_{i\gamma} / \partial (\sigma_\xi^2)^2| &= 0 = c_{11} |\dot{y}_{i\gamma}| \text{ where } c_{11} = 0, \\ |\partial^2 y_{i\gamma} / \partial \sigma_\xi^2 \partial \sigma_u^2| &= 0 = c_{12} |\dot{y}_{i\gamma}| \text{ where } c_{12} = 0, \\ |\partial^2 y_{i\gamma} / \partial \sigma_\xi^2 \partial \sigma_\epsilon^2| &= 0 = c_{13} |\dot{y}_{i\gamma}| \text{ where } c_{13} = 0, \end{aligned}$$

and

$$\begin{aligned} |\partial^2 y_{i\gamma} / \partial (\sigma_u^2)^2| &= |\sigma_\epsilon^2 (\sum_j \gamma_{ij})^{-1} \{-2(\sigma_u^6)^{-1} \sum_j n_{ij}^{-1} \gamma_{ij} (y_{ij} - y_{i\gamma}) \\ &\quad + 2(\sigma_u^4)^{-1} \sum_j n_{ij}^{-1} \gamma_{ij} (y_{ij} - y_{i\gamma}) \partial \gamma_{ij} / \partial \sigma_u^2 - (\sigma_u^4)^{-1} \sum_j n_{ij}^{-1} \gamma_{ij}^2 \partial y_{i\gamma} / \partial \sigma_u^2 \\ &\quad - (\sigma_u^4)^{-1} (\sum_j \gamma_{ij})^{-1} (\sum \partial \gamma_{ij} / \partial \sigma_u^2) (\sum_j n_{ij}^{-1} \gamma_{ij}^2 (y_{ij} - y_{i\gamma}))\}| \\ &\leq \sigma_\epsilon^2 (\sigma_u^6)^{-1} |(\sum_j \gamma_{ij})^{-1} \sum_j \gamma_{ij} y_{ij}| + \sigma_\epsilon^2 (\sigma_u^6)^{-1} (2\gamma_\bullet)^{-1} m_u |\dot{y}_{i\gamma}| \\ &\quad + \sigma_\epsilon^2 (\sigma_u^4)^{-1} g_2 |(\sum_j \gamma_{ij})^{-1} \sum_j \gamma_{ij} y_{ij}| + \sigma_\epsilon^2 (\sigma_u^4)^{-1} (2\gamma_\bullet)^{-1} g_2 m_u |\dot{y}_{i\gamma}| \\ &\quad + \sigma_\epsilon^2 (\sigma_u^4)^{-1} (2\gamma_\bullet)^{-1} 2^{-1} m_u d_2 |\dot{y}_{i\gamma}| \\ &\quad + \sigma_\epsilon^2 (\sigma_u^4)^{-1} (2\gamma_\bullet)^{-1} m_u g_2 2^{-1} |(\sum_j \gamma_{ij})^{-1} (\sum_j \gamma_{ij} y_{ij})| \\ &\quad + \sigma_\epsilon^2 (\sigma_u^4)^{-1} (2\gamma_\bullet)^{-2} m_u^2 g_2 2^{-1} |\dot{y}_{i\gamma}| = c_{22} |\dot{y}_{i\gamma}|. \end{aligned}$$

Similarly, we obtain the following two results:

$$\begin{aligned}
|\partial^2 y_{i\gamma} / \partial \sigma_u^2 \partial \sigma_c^2| &= |(\sigma_u^4)^{-1} (\sum_j \gamma_{ij})^{-1} \{ \sum_j n_{ij}^{-1} \gamma_{ij}^2 (\bar{y}_{ij} - \bar{y}_{i\gamma}) \\
&\quad + 2\sigma_c^2 \sum_j n_{ij}^{-1} \gamma_{ij} (\bar{y}_{ij} - \bar{y}_{i\gamma}) \partial \gamma_{ij} / \partial \sigma_c^2 - \sigma_c^2 \sum_j n_{ij}^{-1} \gamma_{ij}^2 \partial \bar{y}_{i\gamma} / \partial \sigma_c^2 \\
&\quad - \sigma_c^2 (\sum_j \gamma_{ij})^{-1} (\sum_j n_{ij}^{-1} \gamma_{ij}^2 (\bar{y}_{ij} - \bar{y}_{i\gamma})) (\sum_j \partial \gamma_{ij} / \partial \sigma_c^2) \}| \\
&\leq c_{23} |y_{i\gamma}|
\end{aligned}$$

and

$$\begin{aligned}
|\partial^2 y_{i\gamma} / \partial (\sigma_c^2)^2| &= |(\sigma_u^2)^{-1} (\sum_j \gamma_{ij})^{-1} \{ 2 \sum_j n_{ij}^{-1} \gamma_{ij} (\bar{y}_{ij} - \bar{y}_{i\gamma}) \partial \gamma_{ij} / \partial \sigma_c^2 \\
&\quad + \sum_j n_{ij}^{-1} \gamma_{ij}^2 \partial y_{i\gamma} / \partial \sigma_c^2 \\
&\quad - (\sum_j \gamma_{ij})^{-1} (\sum_j \partial \gamma_{ij} / \partial \sigma_c^2) (\sum_j n_{ij}^{-1} \gamma_{ij}^2 (\bar{y}_{ij} - \bar{y}_{i\gamma})) \}| \\
&\leq c_{33} |y_{i\gamma}|.
\end{aligned}$$

Continuing in the same manner, we have,

$$\begin{aligned}
|\partial^2 \gamma_i / \partial (\sigma_\xi^2)^2| &= |\sigma_u^2 (\sum_j \gamma_{ij})^{-1} \{ -2(\sigma_\xi^6)^{-1} \gamma_i^2 + 2(\sigma_\xi^4)^{-1} \partial \gamma_i / \partial \sigma_\xi^2 \\
&\quad - (\sigma_\xi^4)^{-1} \gamma_i^2 (\sum_j \gamma_{ij})^{-1} \sum_j \partial \gamma_{ij} / \partial \sigma_\xi^2 \}| \leq 2\sigma_u^2 (\sigma_\xi^6)^{-1} (2\gamma_*)^{-1} \\
&\quad + 2\sigma_u^2 (\sigma_\xi^4)^{-1} c_1 (2\gamma_*)^{-1} + \sigma_u^2 (\sigma_\xi^4)^{-1} (2\gamma_*)^{-2} m_{ug1} = d_{11}.
\end{aligned}$$

In a similar fashion, we can show that

$$\begin{aligned}
|\partial^2 \gamma_i / \partial \sigma_\xi^2 \partial \sigma_u^2| &= |(\sigma_\xi^4)^{-1} (\sum_j \gamma_{ij})^{-1} \{ \gamma_i^2 + 2\sigma_u^2 \gamma_i \partial \gamma_i / \partial \sigma_u^2 \\
&\quad - (\sum_j \gamma_{ij})^{-1} \sigma_u^2 \gamma_i^2 \sum_j \partial \gamma_{ij} / \partial \sigma_u^2 \}| \\
&\leq d_{12}
\end{aligned}$$

and

$$\begin{aligned}
|\partial^2 \gamma_i / \partial \sigma_\xi^2 \partial \sigma_c^2| &= |\sigma_u^2 (\sigma_\xi^4)^{-1} (\sum_j \gamma_{ij})^{-1} \{ 2\gamma_i \partial \gamma_i / \partial \sigma_c^2 \\
&\quad - (\sum_j \gamma_{ij})^{-1} \gamma_i^2 \sum_j \partial \gamma_{ij} / \partial \sigma_c^2 \}| \\
&\leq d_{13}.
\end{aligned}$$

Finally, we have,

$$\begin{aligned}
|\partial^2 \gamma_i / \partial (\sigma_u^2)^2| &= | - (\sigma_\xi^2)^{-1} (\sum_j \gamma_{ij})^{-2} \{ 2 \gamma_i \partial \gamma_i / \partial \sigma_u^2 (\sum_j \gamma_{ij}^2) \\
&\quad + 2 \gamma_i^2 \sum_j \gamma_{ij} \partial \gamma_{ij} / \partial \sigma_u^2 - 2 \gamma_i^2 (\sum_j \gamma_{ij})^{-1} (\sum_j \partial \gamma_{ij} / \partial \sigma_u^2) (\sum_j \gamma_{ij}^2) \} | \\
&\leq 2 (\sigma_\xi^2)^{-1} (2 \gamma_*)^{-2} c_2 m_u + 2 (\sigma_\xi^2)^{-1} (2 \gamma_*)^{-2} g_2 m_u \\
&\quad + 2 (\sigma_\xi^2)^{-1} (2 \gamma_*)^{-3} m_u^2 g_2 = d_{22},
\end{aligned}$$

and again, in a similar fashion, we obtain

$$\begin{aligned}
|\partial^2 \gamma_i / \partial \sigma_u^2 \partial \sigma_\epsilon^2| &= | - (\sigma_\xi^2)^{-1} (\sum_j \gamma_{ij})^{-2} \{ 2 \gamma_i \partial \gamma_i / \partial \sigma_\epsilon^2 (\sum_j \gamma_{ij}^2) \\
&\quad + 2 \gamma_i^2 \sum_j \gamma_{ij} \partial \gamma_{ij} / \partial \sigma_\epsilon^2 - 2 \gamma_i^2 (\sum_j \gamma_{ij})^{-1} (\sum_j \partial \gamma_{ij} / \partial \sigma_\epsilon^2) (\sum_j \gamma_{ij}^2) \} | \leq d_{23}
\end{aligned}$$

and

$$\begin{aligned}
|\partial^2 \gamma_i / \partial (\sigma_\epsilon^2)^2| &= | - (\sigma_\xi^2)^{-1} (\sum_j \gamma_{ij})^{-2} \{ 2 \gamma_i \partial \gamma_i / \partial \sigma_\epsilon^2 (\sum_j n_{ij}^{-1} \gamma_{ij}^2) \\
&\quad + 2 \gamma_i^2 \sum_j n_{ij}^{-1} \gamma_{ij} \partial \gamma_{ij} / \partial \sigma_\epsilon^2 \\
&\quad - 2 \gamma_i^2 (\sum_j n_{ij}^{-1} \gamma_{ij}^2) (\sum_j \partial \gamma_{ij} / \partial \sigma_\epsilon^2) (\sum_j \gamma_{ij})^{-1} \} | \\
&\leq d_{33}.
\end{aligned}$$

Therefore,

$$|\partial \bar{y}_{i\gamma} / \partial \sigma_\ell^2 \partial \sigma_s^2| \leq c_{\ell s} |\bar{y}_{i\gamma}|$$

and

$$|\partial^2 \gamma_i / \partial \sigma_\ell^2 \partial \sigma_s^2| \leq d_{\ell s}$$

where $c_{11} = c_{12} = c_{13} = 0$ and $\ell, s = 1, 2, 3$. Note that $c_{\ell s} = c_{s\ell}$ and $d_{\ell s} = d_{s\ell} \forall (\ell, s)$.

Now, applying these results, we have

$$\begin{aligned}
|\partial^2 \tilde{\beta} / \partial \sigma_\ell^2 \partial \sigma_s^2| &= (\sum_i \gamma_i)^{-2} \{ (\sum_i |\partial \bar{y}_{i\gamma} / \partial \sigma_s^2| |\partial \gamma_i / \partial \sigma_\ell^2|) (\sum_i \gamma_i) \\
&\quad + (\sum_i |\bar{y}_{i\gamma}| |\partial^2 \gamma_i / \partial \sigma_\ell^2 \partial \sigma_s^2|) (\sum_i \gamma_i)
\end{aligned}$$

$$\begin{aligned}
& + (\sum_i |\partial \gamma_i / \partial \sigma_s^2| |\partial \bar{y}_{i\gamma} / \partial \sigma_\ell^2|) (\sum_i \gamma_i) \\
& + (\sum_i \gamma_i |\partial^2 \bar{y}_{i\gamma} / \partial \sigma_\ell^2 \partial \sigma_s^2|) (\sum_i \gamma_i) \\
& + (\sum_i |\partial \gamma_i / \partial \sigma_s^2|) (\sum_i |\bar{y}_{i\gamma}| |\partial \gamma_i / \partial \sigma_\ell^2|) \\
& + (\sum_i |\partial \gamma_i / \partial \sigma_s^2|) (\sum_i \gamma_i |\partial \bar{y}_{i\gamma} / \partial \sigma_\ell^2|) \\
& + (\sum_i \gamma_i)^{-4} \{ (\sum_i |\partial^2 \gamma_i / \partial \sigma_\ell^2 \partial \sigma_s^2|) (\sum_i \gamma_i |\bar{y}_{i\gamma}|) (\sum_i \gamma_i)^2 \\
& + (\sum_i |\partial \gamma_i / \partial \sigma_\ell^2|) (\sum_i |\partial \gamma_i / \partial \sigma_s^2| |\bar{y}_{i\gamma}|) (\sum_i \gamma_i)^2 \\
& + (\sum_i |\partial \gamma_i \partial \sigma_\ell^2|) (\sum_i \gamma_i |\partial \bar{y}_{i\gamma} / \partial \sigma_s^2|) (\sum_i \gamma_i)^2 \\
& + 2 (\sum_i \gamma_i) (\sum_i |\partial \gamma_i / \partial \sigma_s^2|) (\sum_i |\partial \gamma_i / \partial \sigma_\ell^2|) (\sum_i \gamma_i |\bar{y}_{i\gamma}|) \} \\
& \leq (t\gamma_{**})^{-2} \{ c_\ell d_s t \sum_i |\bar{y}_{i\gamma}| + d_\ell s t \sum_i |\bar{y}_{i\gamma}| + d_\ell c_s t \sum_i |\bar{y}_{i\gamma}| \\
& + c_\ell s t \sum_i |\bar{y}_{i\gamma}| + c_\ell c_s t \sum_i |\bar{y}_{i\gamma}| + d_\ell c_s t \sum_i |\bar{y}_{i\gamma}| \} \\
& + (t\gamma_{**})^{-4} \{ d_\ell s t^3 \sum_i |\bar{y}_{i\gamma}| + c_\ell c_s t^3 \sum_i |\bar{y}_{i\gamma}| + c_\ell d_s t^3 \sum_i |\bar{y}_{i\gamma}| \\
& + 2 c_\ell c_s t^3 \sum_i |\bar{y}_{i\gamma}| \} = f_\ell s t^{-1} \sum_i |\bar{y}_{i\gamma}|
\end{aligned}$$

or

$$|\partial^2 \tilde{\beta} / \partial \sigma_\ell^2 \partial \sigma_s^2| \leq f_\ell s t^{-1} \sum_i |\bar{y}_{i\gamma}| \text{ for } \ell, s = 1, 2, 3.$$

We conclude the proof by noting that

$$\begin{aligned}
|t''_{\ell s}| & \leq (1 - \gamma_i) |\partial^2 \tilde{\beta} / \partial \sigma_\ell^2 \partial \sigma_s^2| + |\partial \tilde{\beta} / \partial \sigma_\ell^2| |\partial \gamma_i / \partial \sigma_s^2| \\
& + |\bar{y}_{i\gamma}| |\partial^2 \gamma_i / \partial \sigma_\ell^2 \partial \sigma_s^2| + |\tilde{\beta}| |\partial^2 \gamma_i / \partial \sigma_\ell^2 \partial \sigma_s^2| + |\partial \bar{y}_{i\gamma} / \partial \sigma_s^2| |\partial \gamma_i / \partial \sigma_\ell^2| \\
& + |\partial \tilde{\beta} / \partial \sigma_s^2| |\partial \gamma_i / \partial \sigma_\ell^2| + |\partial \gamma_i / \partial \sigma_s^2| |\partial \bar{y}_{i\gamma} / \partial \sigma_\ell^2| + \gamma_i |\partial^2 \bar{y}_{i\gamma} / \partial \sigma_\ell^2 \partial \sigma_s^2| \\
& \leq f_\ell s t^{-1} \sum_i |\bar{y}_{i\gamma}| + c_s f_\ell t^{-1} \sum_i |\bar{y}_{i\gamma}| + d_\ell s |\bar{y}_{i\gamma}| \\
& + d_\ell c_s t^{-1} \sum_i |\bar{y}_{i\gamma}| + c_\ell d_s |\bar{y}_{i\gamma}| + c_\ell f_s t^{-1} \sum_i |\bar{y}_{i\gamma}| \\
& + c_s d_\ell |\bar{y}_{i\gamma}| + c_\ell s |\bar{y}_{i\gamma}| = a_\ell s |\bar{y}_{i\gamma}| + b_\ell s t^{-1} \sum_i |\bar{y}_{i\gamma}|
\end{aligned}$$

or

$$|t''_{\ell s}| \leq a_\ell s |\bar{y}_{i\gamma}| + b_\ell s t^{-1} \sum_i |\bar{y}_{i\gamma}|$$

for $\ell, s = 1, 2, 3$, as desired. ■

Appendix B

Lemma B.1 Assuming the conditions of Theorem 4.1, we have that

$$(a) \quad t^{-1}[\sum_i (\sum_j \sum_k \bar{c}_{ijk}^2)^2 - (\sum_i \nu_{1i})^2 \text{var}(\bar{\sigma}_c^2) - \sigma_c^4 \sum_i \nu_{1i}^2] = o_p(1).$$

$$(b) \quad \bar{\sigma}_c^2 - \sigma_c^2 = o_p(1).$$

$$(c) \quad t^{-1}[\sum_i \nu_{1i} \sum_j \sum_k \bar{c}_{ijk}^2 - \sigma_c^2 \sum_i \nu_{1i}^2] = o_p(1).$$

Proof To proceed, we require the following lemma:

Lemma B.1.1 (Loève (1963), Theorem B(i), pp. 275) Let $\{X_i; i = 1, \dots, t\}$ be a sequence of independent random variables such that $E(X_i) = 0 \forall_i$. If $t^{-(1+\delta)} \sum_{i=1}^t E|X_i|^{1+\delta} = o(1)$ for $0 < \delta < 1$ then $t^{-1} \sum_{i=1}^t X_i = o_p(1)$.

(a) We apply Lemma B.1.1 to the sequence $(\sum_j \sum_k \bar{c}_{ijk}^2)^2 - E(\sum_j \sum_k \bar{c}_{ijk}^2)^2$, but first show that the condition

$$t^{-(1+\delta)} \sum_{i=1}^t E|(\sum_j \sum_k \bar{c}_{ijk}^2)^2 - E(\sum_j \sum_k \bar{c}_{ijk}^2)^2|^{1+\delta} = o(1) \quad (B.1)$$

is satisfied. Now by regularity condition (R4), $E|\bar{c}_{ijk}|^{1+\delta} < \infty$. By Lemma B.2 ahead, this implies that $E[\sum_j \sum_k \bar{c}_{ijk}^2]^{2+\delta'} < \infty$ for $0 < \delta < 1$ and $\delta' = \delta$. Thus $E[\sum_j \sum_k \bar{c}_{ijk}^2]^2 < \infty$ since any lower order moment is also bounded. Therefore, using the C_r inequality, we need only show that $t^{-(1+\delta)} \sum_i E|(\sum_j \sum_k \bar{c}_{ijk}^2)^2|^{1+\delta} = o(1)$ to prove (B.1). But $E|(\sum_j \sum_k \bar{c}_{ijk}^2)^2|^{1+\delta} < \infty$ since $E[\sum_j \sum_k \bar{c}_{ijk}^2]^{2+\delta'} < \infty$. Therefore

$$t^{-(1+\delta)} \sum_i E|(\sum_j \sum_k \bar{c}_{ijk}^2)^2 - E(\sum_j \sum_k \bar{c}_{ijk}^2)^2|^{1+\delta} = O(t^{-\delta}),$$

which in turn implies that (B.1) is satisfied.

Now applying Lemma B.1.1 to the sequence $(\sum_j \sum_k \bar{c}_{ijk}^2)^2 - E(\sum_j \sum_k \bar{c}_{ijk}^2)^2$, we have

$$t^{-1} \sum_i (\sum_j \sum_k \bar{c}_{ijk}^2)^2 - t^{-1} \text{var}(\bar{\sigma}_c^2 \sum_i \nu_{1i}) - t^{-1} \sum_i (E(\sum_j \sum_k \bar{c}_{ijk}^2))^2 = o_p(1)$$

using the independence of $\sum_j \sum_k \bar{c}_{ijk}^2$ over i . However,

$$E(\sum_j \sum_k \bar{c}_{ijk}^2) = \sigma_c^2(n_{1\cdot} - m_1) = \sigma_c^2(\nu_{11} + O(t^{-1})).$$

Thus

$$\begin{aligned} t^{-1} \sum_i (\sum_j \sum_k \bar{c}_{ijk}^2)^2 - t^{-1} (\sum_i \nu_{1i})^2 \text{var}(\bar{\sigma}_c^2) - t^{-1} \sigma_c^4 \sum_i \nu_{1i}^2 - 2t^{-1} \sigma_c^4 (\sum_i \nu_{1i}) O(t^{-1}) \\ - \sigma_c^4 O(t^{-2}) = o_p(1), \end{aligned}$$

or equivalently,

$$t^{-1} [\sum_i (\sum_j \sum_k \bar{c}_{ijk}^2)^2 - (\sum_i \nu_{1i})^2 \text{var}(\sigma_c^2) - \sigma_c^4 \sum_i \nu_{1i}^2] = o_p(1)$$

since $(\sum_i \nu_{1i})/t$ is bounded via (4.30).

(b) We apply Lemma B.1.1 to the sequence $\sum_j \sum_k \bar{c}_{ijk}^2 - E(\sum_j \sum_k \bar{c}_{ijk}^2)$. First, we need the following condition to be satisfied:

$$t^{-(1+\delta)} \sum_i E |\sum_j \sum_k \bar{c}_{ijk}^2 - E(\sum_j \sum_k \bar{c}_{ijk}^2)|^{1+\delta} = o(1).$$

The proof of this is similar in nature to that of (B.1).

Now, applying Lemma B.1.1, we get that

$$t^{-1} \sum_i \sum_j \sum_k \bar{c}_{ijk}^2 - t^{-1} E(\sum_i \sum_j \sum_k \bar{c}_{ijk}^2) = o_p(1),$$

or equivalently, $\hat{\sigma}_c^2 - \sigma_c^2 = o_p(1)$.

(c) Finally, we apply Lemma B.1.1 to the sequence $\nu_{1i} \sum_j \sum_k \bar{c}_{ijk}^2 - \nu_{1i} E(\sum_j \sum_k \bar{c}_{ijk}^2)$. First, we need the following condition to be satisfied

$$t^{-(1+\delta)} \sum_i E |\nu_{1i} \sum_j \sum_k \bar{c}_{ijk}^2 - \nu_{1i} E(\sum_j \sum_k \bar{c}_{ijk}^2)|^{1+\delta} = o(1).$$

The proof of this is similar in nature to that of (B.1).

Now applying Lemma B.1.1, we get that

$$t^{-1}[\sum_i \nu_{1i} \sum_j \sum_k \bar{c}_{ijk}^2 - \sum_i \nu_{1i} E[\sum_j \sum_k c_{ijk}^2]] = o_p(1),$$

or equivalently,

$$t^{-1}[\sum_i \nu_{1i} \sum_j \sum_k c_{ijk}^2 - \sigma_c^2 \sum_i \nu_{1i}^2] = o_p(1). \quad \blacksquare$$

Lemma B.2 *Assuming the conditions of Theorem 4.1, we have that if regularity condition (R4) holds, i.e., if $\sup_{1 \leq i \leq t} \sup_{1 \leq j \leq m} \sup_{1 \leq k \leq n_{ij}} E|c_{ijk}|^{4+\delta} < \infty$, then $\sup_{1 \leq i \leq t} E[\sum_j \sum_k \bar{c}_{ijk}^2]^{2+\delta'} < \infty$ for some $0 < \delta < 1$ and $2\delta' = \delta$.*

Proof Now

$$\begin{aligned} \sum_j \sum_k \bar{c}_{ijk}^2 &= \sum_j \sum_k (c_{ijk} - \bar{c}_{ij.})^2 \leq 2 \sum_j \sum_k c_{ijk}^2 + 2 \sum_j n_{ij} c_{ij.}^2 \\ &\leq 4 \sum_j \sum_k c_{ijk}^2 \end{aligned}$$

by applying the inequality $\sum_i (a_i - b_i)^2 \leq 2(\sum_i a_i^2 + \sum_i b_i^2)$ followed by the inequality $n_{ij} c_{ij.}^2 \leq \sum_k c_{ijk}^2$ (since $\sum_k (c_{ijk} - \bar{c}_{ij.})^2 \geq 0$). Thus

$$E(\sum_j \sum_k \bar{c}_{ijk}^2)^{2+\delta'} \leq 2^{4+\delta} E(\sum_j \sum_k c_{ijk}^2)^{2+\delta'},$$

and upon applying Minkowski's inequality, we have

$$E(\sum_j \sum_k \bar{c}_{ijk}^2)^{2+\delta'} \leq 2^{4+\delta} (\sum_j \sum_k (E(c_{ijk})^{4+\delta})^{1/(2+\delta')})^{2+\delta'} < \infty$$

using the initial assumption. $\blacksquare$

Lemma B.3 *Assuming the conditions of Theorem 4.1, we have that*

(a) $\bar{\sigma}_u^2 - \sigma_u^2 = o_p(1).$

(b) $t^{-1}[\sum_i \tau_{1i} T_i - \sigma_u^2 \sum_i \tau_{1i}^2] = o_p(1).$

$$(c) \quad t^{-1}[\sum_i T_i^2 - (\sum_i \tau_{1i})^2 \text{var}(\sigma_u^2) - \sigma_u^4 \sum_i \tau_{1i}^2] = o_p(1).$$

Proof (a) We apply Lemma B.1.1 to the sequence $T_i - E(T_i)$, but first show that the condition

$$t^{-(1+\delta)} \sum_i E|T_i - E(T_i)|^{1+\delta} = o(1) \quad (B.2)$$

is satisfied. First consider

$$\begin{aligned} E|T_i|^{1+\delta} &= E|\sum_j \sum_k \bar{u}_{ijk}^2 - \nu_{1i}^{-1} \nu_{2i} \sum_j \sum_k \bar{e}_{ijk}^2|^{1+\delta} \\ &\leq 2^\delta E|\sum_j \sum_k \bar{u}_{ijk}^2|^{1+\delta} + 2^\delta |\nu_{1i}^{-1} \nu_{2i}|^{1+\delta} E|\sum_j \sum_k \bar{e}_{ijk}^2|^{1+\delta} \end{aligned}$$

by the C_r inequality. Now by regularity condition (R9), we have that $E|u_{ij}|^{4+\delta} < \infty$. However, by Lemma B.4 ahead, this implies that $E[\sum_j \sum_k u_{ijk}^2]^{2+\delta'} < \infty$ for $0 < \delta < 1$ and $\delta = 2\delta'$. Thus, $E[\sum_j \sum_k u_{ijk}^2]^{1+\delta} < \infty$ since any lower order moment is also bounded. Now, it was established that $E[\sum_j \sum_k \bar{e}_{ijk}^2]^{2+\delta'} < \infty$ in Lemma B.1(a) and so $E[\sum_j \sum_k \bar{e}_{ijk}^2]^{1+\delta} < \infty$. In addition, we show that

$$\nu_{1i}^{-1} \nu_{2i} = O(1) \quad (B.3)$$

as follows:

$$\nu_{1i} = \sum_j n_{ij} - m_i - [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2]^{-1} \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 = O(1)$$

and is bounded away from zero, using regularity conditions (R1), (R2) and (R3).

Also, $\nu_{2i} = n_i - 1 - [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2]^{-1} \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 = O(1)$ using regularity conditions (R1), (R5) and (R6). Hence $E|T_i|^{1+\delta} < \infty$ which in turn implies that $E|T_i| < \infty$ since any lower order moment is also bounded. Thus, using the C_r inequality, we need only show that $t^{-(1+\delta)} \sum_i E|T_i|^{1+\delta} = o(1)$ to satisfy (B.2). However, since $E|T_i|^{1+\delta} < \infty$, then $t^{-(1+\delta)} \sum_i E|T_i|^{1+\delta} = O(t^{-\delta})$.

Now, applying Lemma B.1.1 to the sequence $T_i - E(T_i)$, we get

$$t^{-1} \sum_i (T_i - E(T_i)) = t^{-1} [\sigma_u^2(\sum_i \tau_{1i}) - \sigma_u^2(\sum_i \tau_{1i} + O(1)) - \sigma_e^2 O(1)] = o_p(1)$$

since $\sum_i \tau_{1i} = n - \sum_i n_i^{-1} \sum_j n_{ij}^2 + O(1)$ and since

$$E(T_i) = \sigma_u^2(n_i - n_i^{-1} \sum_j n_{ij}^2) + \sigma_e^2 O(t^{-1}).$$

Thus, $t^{-1}(\bar{\sigma}_u^2 - \sigma_u^2) \sum_i \tau_{1i} = o_p(1)$, or equivalently, $\bar{\sigma}_u^2 - \sigma_u^2 = o_p(1)$ since $\sum_i \tau_{1i}/t$ is bounded.

(b) We apply Lemma B.1.1 to the sequence $\tau_{1i} T_i - \tau_{1i} E(T_i)$. First, we need the following condition to be satisfied:

$$t^{-(1+\delta)} \sum_i E|\tau_{1i} T_i - E(\tau_{1i} T_i)|^{1+\delta} = o(1).$$

The proof of this is similar in nature to that of (B.2).

Now, applying Lemma B.1.1, we have that

$$\begin{aligned} t^{-1} \sum_i (\tau_{1i} T_i - E(\tau_{1i} T_i)) &= t^{-1} [\sum_i \tau_{1i} T_i - \sigma_u^2 \sum_i \tau_{1i} (\tau_{1i} + O(t^{-1})) - \sigma_e^2 O(t^{-1}) \sum_i \tau_{1i}] \\ &= o_p(1), \end{aligned}$$

or equivalently,

$$t^{-1} [\sum_i \tau_{1i} T_i - \sigma_u^2 \sum_i \tau_{1i}^2] = o_p(1)$$

since $\sum_i \tau_{1i}/t$ is bounded.

(c) We apply Lemma B.1.1 to the sequence $T_i^2 - E(T_i^2)$. First, we need the following condition to be satisfied:

$$t^{-(1+\delta)} \sum_i E|T_i^2 - E(T_i^2)|^{1+\delta} = o(1).$$

The proof of this is similar in nature to that of (B.2).

Now, applying Lemma B.1.1, we have that

$$\begin{aligned} t^{-1} \sum_i (T_i^2 - E(T_i^2)) &= t^{-1} \sum_i T_i^2 - t^{-1} \text{var}(\sum_i T_i) - t^{-1} \sum_i (E(T_i))^2 \\ &= t^{-1} \sum_i T_i^2 - t^{-1} (\sum_i \tau_{1i})^2 \text{var}(\bar{\sigma}_u^2) - t^{-1} \sigma_u^4 \sum_i (\tau_{1i} + O(t^{-1}))^2 = o_p(1), \end{aligned}$$

or equivalently that

$$t^{-1} [\sum_i T_i^2 - (\sum_i \tau_{1i})^2 \text{var}(\bar{\sigma}_u^2) - \sigma_u^4 \sum_i \tau_{1i}^2] = o_p(1)$$

since $\sum_i \tau_{1i}/t$ is bounded. ■

Lemma B.4 Assuming the conditions of Theorem 4.1, we have that if regularity condition (R9) holds, i.e., $\sup_{1 \leq i \leq t} \sup_{1 \leq j \leq m_i} E|u_{ij}|^{4+\delta} < \infty$, then $\sup_{1 \leq i \leq t} E[\sum_j \sum_k u_{ijk}^2]^{2+\delta'} < \infty$ for $0 < \delta < 1$ and $2\delta' = \delta$.

Proof Now

$$\begin{aligned} \sum_j \sum_k u_{ijk}^2 &= \sum_j \sum_k [(e_{ijk} - \bar{e}_{i..}) + (u_{ij} - \bar{u}_{i.})]^2 \\ &\leq 2[\sum_j \sum_k (e_{ijk} - \bar{e}_{i..})^2 + \sum_j \sum_k (u_{ij} - \bar{u}_{i.})^2] \\ &\leq 4[\sum_j \sum_k e_{ijk}^2 + n_i \bar{e}_{i..}^2 + \sum_j \sum_k u_{ij}^2 + n_i \bar{u}_{i.}^2] \\ &\leq 8[\sum_j \sum_k e_{ijk}^2 + \sum_j \sum_k u_{ij}^2] \end{aligned}$$

by applying the inequalities $\sum_i (a_i + b_i)^2 \leq 2(\sum_i a_i^2 + \sum_i b_i^2)$ and $\sum_i (a_i - b_i)^2 \leq 2(\sum_i a_i^2 + \sum_i b_i^2)$, followed by the inequalities $n_i \bar{e}_{i..}^2 \leq \sum_j \sum_k e_{ijk}^2$ (since $\sum_j \sum_k (e_{ijk} - \bar{e}_{i..})^2 \geq 0$) and $n_i \bar{u}_{i.}^2 \leq \sum_j \sum_k u_{ij}^2$ (since $\sum_j \sum_k (u_{ij} - \bar{u}_{i.})^2 \geq 0$). Thus $E[\sum_j \sum_k \bar{u}_{ijk}^2]^{2+\delta'} \leq 8^{2+\delta'} E[\sum_j \sum_k e_{ijk}^2 + \sum_j \sum_k u_{ij}^2]^{2+\delta'}$, and upon applying the C_r inequality and Minkowski's inequality, we have

$$\begin{aligned} E[\sum_j \sum_k u_{ijk}^2]^{2+\delta'} &\leq 2^{7+2\delta} \{E[\sum_j \sum_k e_{ijk}^2]^{2+\delta'} + E[\sum_j \sum_k u_{ij}^2]^{2+\delta'}\} \\ &\leq 2^{7+2\delta} \{(\sum_j \sum_k (E(e_{ijk})^{4+\delta})^{1/(2+\delta')})^{2+\delta'} + (\sum_j \sum_k (E(u_{ij})^{4+\delta})^{1/(2+\delta')})^{2+\delta'}\} < \infty \end{aligned}$$

using the initial assumptions. ■

Lemma B.5 Assuming the conditions of Theorem 4.1, we have that

- (i) $\text{var}(\beta^{**} - \beta)^2 = O(t^{-2})$.
- (ii) $\text{var}(d_i(\beta^* - \beta)) = O(t^{-1})$.
- (iii) $\text{cov}(\sigma_u^2, (\beta^{**} - \beta)^2) = O(t^{-2})$.
- (iv) $\text{cov}(\sigma_u^2, (\beta^* - \beta)^2) = O(t^{-2})$.
- (v) $\text{cov}(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i(\beta^* - \beta), \sigma_u^2) = O(t^{-1})$.
- (vi) $\text{cov}((\beta^* - \beta)^2, (\beta^{**} - \beta)^2) = O(t^{-2})$.
- (vii) $\text{cov}(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i(\beta^* - \beta), (\beta^{**} - \beta)^2) = O(t^{-1})$.
- (viii) $\text{cov}(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i(\beta^* - \beta), (\beta^* - \beta)^2) = O(t^{-1})$.

Proof (i) This proof is similar to the proof that $\text{var}(\beta^* - \beta)^2 = O(t^{-2})$ so is omitted here.

$$\begin{aligned} \text{(ii)} \quad \text{var}(d_i(\beta^* - \beta)) &= E[d_i^2(\beta^* - \beta)^2] - [E(d_i(\beta^* - \beta))]^2 \\ &\leq E[d_i^2(\beta^* - \beta)^2] + [E(d_i(\beta^* - \beta))]^2 \leq (E(d_i^4)E(\beta^* - \beta)^4)^{1/2} + E(d_i^2)E(\beta^* - \beta)^2 \end{aligned}$$

using Holder's inequality. Now in Lemma 4.2(b), we showed that $E(d_i^4) < \infty$ and thus $E(d_i^2) < \infty$. Finally, $E(\beta^* - \beta)^2 = \text{var}(\beta^*) = O(t^{-1})$ from the proof of Lemma 4.2(a), and $E(\beta^* - \beta)^4 = O(t^{-2})$ since $\text{var}(\beta^* - \beta)^2 = O(t^{-2})$.

(iii) Define

$$\begin{aligned} d_i^* &= \sum_j \sum_k (x_{ijk} - \bar{x}_{i..}) \bar{u}_{ijk} \\ &= \sum_j \sum_k (y_{ijk} - \bar{y}_{i..})(x_{ijk} - \bar{x}_{i..}) - \beta \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2 \end{aligned}$$

so that $(\beta^{**} - \beta) = [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2]^{-1} \sum_i d_i^*$. Define $q_i^* = T_i - E(T_i)$ so that $\sum_i q_i^* = \bar{\sigma}_u^2 (\sum_i \tau_{1i}) - \sigma_u^2 (n - \sum_i n_i^{-1} \sum_j n_{ij}^2) - \sigma_e^2 O(1)$, or equivalently, $\bar{\sigma}_u^2 = (\sum_i \tau_{1i})^{-1} (\sum_i q_i^* + \sigma_u^2 (n - \sum_i n_i^{-1} \sum_j n_{ij}^2) + \sigma_e^2 O(1))$. Thus, we can write

$$\text{cov}(\bar{\sigma}_u^2, (\beta^{**} - \beta)^2) = (\sum_i \tau_{1i})^{-1} [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{i..})^2]^{-2} \text{cov}((\sum_i d_i^*)^2, \sum_i q_i^*).$$

It suffices to show that $\text{cov}((\sum_{\dot{i}} d_i^*)^2, \sum_{\dot{i}} q_i^*) = O(t)$. To do so, we proceed exactly as we did in the proof of Lemma 4.2(b), where we showed that $\text{cov}((\sum_{\dot{i}} d_i)^2, \sum_{\dot{i}} q_i) = O(t)$.

(iv) The proof of this is similar in nature to that of (iii), so is omitted here.

(v) Now

$$\begin{aligned} & \text{cov}(\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i (\beta^* - \beta), \sigma_u^2) \\ &= [\sum_{\dot{i}} \sum_j \sum_k (x_{ijk} - x_{ij})^2]^{-1} (\sum_{\dot{i}} \tau_{1i})^{-1} \text{cov}[(\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i)(\sum_{\dot{i}} d_i), \sum_{\dot{i}} q_i^*]. \end{aligned}$$

Thus, it suffices to show that $|\text{cov}[(\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i)(\sum_{\dot{i}} d_i), \sum_{\dot{i}} q_i^*]| = O(t)$. Now

$$\begin{aligned} |\text{cov}[(\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i)(\sum_{\dot{i}} d_i), \sum_{\dot{i}} q_i^*]| &= |E[(\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i)(\sum_{\dot{i}} d_i)(\sum_{\dot{i}} q_i^*)]| \\ &= |E[(\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i^2)(\sum_{\dot{i}} q_i^*)] + E[(\sum_{\dot{i} \neq j} \nu_{1i}^{-1} \nu_{2i} d_i d_j)(\sum_{\dot{i}} q_i^*)]| \end{aligned}$$

since $(\sum_{\dot{i}} a_i)(\sum_{\dot{i}} b_i) = \sum_{\dot{i}} a_i b_i + \sum_{\dot{i} \neq j} a_i b_j$. But the second term on the right hand side of the equality is zero since d_i is independent of d_j for $i \neq j$, and since $(d_i, q_i^*; i = 1, \dots, t)$ are independent with $E(d_i) = F(q_i^*) = 0$. Since d_i^2 is independent of q_j^* , we have

$$\begin{aligned} |E[(\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i^2)(\sum_{\dot{i}} q_i^*)]| &= |E[\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i^2 q_i^*] + E[\sum_{\dot{i} \neq j} \nu_{1i}^{-1} \nu_{2i} d_i^2 q_j^*]| \\ &= |E[\sum_{\dot{i}} \nu_{1i}^{-1} \nu_{2i} d_i^2 q_i^*]| \leq \sum_{\dot{i}} E|\nu_{1i}^{-1} \nu_{2i} d_i^2 q_i^*| \\ &\leq \sum_{\dot{i}} |\nu_{1i}^{-1} \nu_{2i}| (E|d_i^4| E|q_i^{*2}|)^{1/2} = O(t) \end{aligned}$$

upon applying Holder's inequality and noting that $E|d_i^4| < \infty$ and $E|q_i^{*2}| < \infty$. The proof that $E|q_i^{*2}| < \infty$ is similar to that of $E|q_i^2| < \infty$ from Lemma 4.2(b).

(vi) Now

$$\begin{aligned} & \text{cov}[(\beta^* - \beta)^2, (\beta^{**} - \beta)^2] \\ &= E[(\beta^* - \beta)^2 (\beta^{**} - \beta)^2] - E(\beta^* - \beta)^2 E(\beta^{**} - \beta)^2 \\ &\leq [E(\beta^* - \beta)^4 E(\beta^{**} - \beta)^4]^{1/2} - \text{var}(\beta^*) \text{var}(\beta^{**}) = O(t^{-2}) \end{aligned}$$

by applying Holder's inequality and noting that $\text{var}(\beta^* - \beta)^2 = O(t^{-2})$, $\text{var}(\beta^{**} - \beta)^2 = O(t^{-2})$, $\text{var}(\beta^*) = O(t^{-1})$ and $\text{var}(\beta^{**}) = O(t^{-1})$, as shown previously.

(vii) Now

$$\begin{aligned} & \text{cov}[\sum_i \nu_{1i}^{-1} \nu_{2i} d_i (\beta^* - \beta), (\beta^{**} - \beta)^2] \\ &= [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2]^{-1} [\sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2]^{-2} \text{cov}[(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i)(\sum_i d_i), (\sum_i d_i^*)^2]. \end{aligned}$$

Thus it suffices to show that $\text{cov}[(\sum_i \nu_{1i}^{-1} \nu_{2i} d_i)(\sum_i d_i), (\sum_i d_i^*)^2] = O(t^2)$. One can argue this using reasoning along the lines of the proof of (v).

(viii) This proof is similar in nature to (vii), so is omitted here. ■

References

- [1] Arvesen, J.N. (1969). "Jackknifing U-Statistics", *The Annals of Mathematical Statistics*, **40**, 2076-2100.
- [2] Arvesen, J.N. and Schmitz, T.H. (1970). "Robust Procedures for Variance Component Problems Using the Jackknife", *Biometrics*, **26**, 677-686.
- [3] Battese, G.E. and Fuller, W.A. (1981). "Prediction of County Crop Areas Using Survey and Satellite Data", *Proceedings of the Section on Survey Research Methods*, pp. 500-505.
- [4] Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). "An Error Components Model for Prediction of County Crop Areas Using Survey and Satellite Data", *Journal of the American Statistical Association*, **83**, 28-36.
- [5] Bhat, B.R. (1981). *Modern Probability Theory*, Wiley Eastern Ltd.
- [6] Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Mass.
- [7] Broemeling, L.D. (1969). "Confidence Regions for Variance Ratios of Random Models", *Journal of the American Statistical Association*, **64**, 660-664.
- [8] Bogue, D.J. (1950). "A Technique for Making Extensive Population Estimates", *Journal of the American Statistical Association*, **45**, 149-163.
- [9] Bogue, D.J. and Duncan, B.D. (1959). "A Composite Method for Estimating Post-censal Population of Small Areas by Age, Sex and Colour", *Vital Statistics-Special Report 47* (6): 160-185, National Office of Vital Statistics, Washington, D.C.
- [10] Calvin, J.A. and Sedransk, J. (1991). "Bayesian and Frequentist Predictive Inference for the Patterns of Care Studies", *Journal of the American Statistical Association*, **86**, 36-48.
- [11] Choudhry, G.H. and Rao, J.N.K. (1988). "Evaluation of Small Area Estimators: An Empirical Study", paper presented at the Symposium on Small Area Statistics, New Orleans.
- [12] Datta, G.S. and Ghosh, M. (1990). "Bayesian Prediction in Linear Models: Applications to Small Area Estimation", unpublished manuscript.
- [13] Dempster, A.P. and Ryan, L.M. (1985). "Weighted Normal Probability Plots", *Journal of the American Statistical Association*, **80**, 845-850.
- [14] 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.

- [15] Fay, R.E. and Herriot, R.A. (1979). "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data", *Journal of the American Statistical Association*, **74**, 269-277.
- [16] Fuller, W.A. (1987). *Measurement Error Models*, John Wiley & Sons.
- [17] Fuller, W.A. and Battese, G.E. (1973). "Transformations for Estimation of Linear Models with Nested-Error Structure", *Journal of the American Statistical Association*, **68**, 626-632.
- [18] Gelfand, A. and Smith, A.F.M. (1990). "Sampling Based Approaches to Calculating Marginal Densities", to appear in *Journal of the American Statistical Association*.
- [19] Ghosh, M. and Lahiri, P. (1987). "Robust Empirical Bayes Estimation of Means from Stratified Samples", *Journal of the American Statistical Association*, **82**, 1153-1162.
- [20] Ghosh, M. and Lahiri, P. (1988). "Bayes and Empirical Bayes Analysis in Multistage Sampling", *Statistical Decision Theory and Related Topics*, **1**, (eds. S.S. Gupta and J.O. Berger), Springer-Verlag, 195-212.
- [21] Ghosh, M. and Lahiri, P. (1989). "A Hierarchical Bayes Approach to Small Area Estimation with Auxiliary Information", to appear in the Proceedings of the Joint Indo-U.S. Workshop on Bayesian Inference in Statistics and Econometrics.
- [22] Ghosh, M. and Meeden, G. (1986). "Empirical Bayes Estimation in Finite Population Sampling", *Journal of the American Statistical Association*, **81**, 1058-1062.
- [23] Graybill, F.A. (1969). *Introduction to Matrices with Applications in Statistics*, Wadsworth Publishing Company.
- [24] Harter, R.M. (1983). "Small Area Estimation Using Nested-Error Models and Auxiliary Data", unpublished Ph.D. thesis, Iowa State University, Ames, Iowa.
- [25] Henderson, C.R. (1975). "Best Linear Unbiased Estimation and Prediction Under a Selection Model", *Biometrics*, **31**, 423-447.
- [26] Hinkley, D.V. (1977). "Jackknifing in Unbalanced Situations", *Technometrics*, **19**, 285-292.
- [27] Kackar, R.N. (1979). "Variance Approximations for Estimators of Fixed and Random Effects in Mixed Linear Models", unpublished Ph.D. thesis, Iowa State University, Ames, Iowa.
- [28] Kackar, R.N. and Harville, D.A. (1984). "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models", *Journal of the American Statistical Association*, **79**, 853-862.

- [29] Kleffe, J. and Rao, J.N.K. (1986). "The Existence of Asymptotically Unbiased Non-negative Quadratic Estimates of Variance Components in ANOVA Models", *Journal of the American Statistical Association*, **81**, 692-698.
- [30] Laird, N. M. and Louis, T.A. (1987). "Empirical Bayes Confidence Intervals Based on Bootstrap Samples", *Journal of the American Statistical Association*, **82**, 739-750.
- [31] LaMotte, L.R. (1973). "On Non-Negative Quadratic Unbiased Estimation of Variance Components", *Journal of the American Statistical Association*, **68**, 728-730.
- [32] Lange, N. and Ryan, L.M. (1989). "Assessing Normality in Random Effects Models", *Annals of Statistics*, **17**, 624-642.
- [33] Loève, M. (1963). *Probability Theory*, D. van Nostrand Company, Princeton.
- [34] Maler, D. and Sedransk, J. (1985). "Bayesian Inference for Finite Population Parameters in Multistage Cluster Sampling", *Journal of the American Statistical Association*, **80**, 897-902.
- [35] Miller, R.G. (1974). "An Unbalanced Jackknife", *Annals of Statistics*, **2**, 880-891.
- [36] NIDA Research Monograph 24 (1979). *Synthetic Estimates for Small Areas: Statistical Workshop Papers and Discussion*, (ed. J. Steinberg), U.S. Government Printing Office, Washington, D.C.
- [37] Platek, R., Rao, J.N.K., Särndal, C.E. and Singh, M.P. (Editors) (1987). *Small Area Statistics: An International Symposium*, John Wiley and Sons.
- [38] Prasad, N.G.N. (1985). "Small Area Estimation and Measurement of Response Error Variance in Surveys", unpublished Ph.D. thesis, Carleton University, Ottawa, Canada.
- [39] Prasad, N.G.N. and Rao, J.N.K. (1988). "Robust Tests and Confidence Intervals for Error Variance in a Regression Model and For Functions of Variance Components in an Unbalanced Random One-Way Model", *Communications in Statistics-Theory and Methods*, **A17**, 1111-1133.
- [40] Prasad, N.G.N. and Rao, J.N.K. (1990). "The Estimation of the Mean Squared Error of Small-Area Estimators", *Journal of the American Statistical Association*, **85**, 163-171.
- [41] Purcell, N.J. and Kish, L. (1980). "Postcensal Estimates for Local Areas (or Domains)", *International Statistical Review*, **48**, 3-18.
- [42] Quenouille, M.H. (1949). "Notes on Bias in Estimation", *Biometrika*, **43**, 353-360.
- [43] Quenouille, M.H. (1956). "Approximate Tests of Correlation in Time-Series", *Journal of the Royal Statistical Society, Ser. B*, **11**, 68-74.

- [44] Rao, C.R. (1965). *Linear Statistical Inference and Its Applications*, Second Edition, Wiley Eastern Ltd.
- [45] Rao, C.R. and Kleffe, J. (1988). *Estimation of Variance Components and Applications*, North-Holland Series in Statistics and Probability, 3, (eds. P.R. Krishnaiah, C.R. Rao, M. Rosenblatt and Y.A. Rozanov), Elsevier Science Publishing Company.
- [46] Rives, N.W. (1976). "A Modified Housing-Unit Method for Small-Area Population Estimation", *Proceedings of the Social Statistics Section*, pp. 717-720, American Statistical Association, Washington, D.C.
- [47] Royall, R.M. (1976). "The Linear Least-Squares Prediction Approach to Two-Stage Sampling", *Journal of the American Statistical Association*, 71, 657-664.
- [48] Särndal, C.E. and Hidiroglou, M.A. (1989). "Small Domain Estimation: A Conditional Analysis", *Journal of the American Statistical Association*, 84, 266-275.
- [49] Schnable, W.L., Brock, D.B. and Schmack, G.A. (1977). "An Empirical Comparison of the Simple Inflation, Synthetic and Composite Estimators for Small Area Statistics", *Proceedings of the Social Statistics Section*, pp. 1017-1021, American Statistical Association, Washington, D.C.
- [50] Searle, S.R. (1971). *Linear Models*, John Wiley and Sons.
- [51] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons.
- [52] Spjøtvoll, E. (1967). "Optimum Invariant Tests in Unbalanced Variance Component Models", *The Annals of Mathematical Statistics*, 38, 422-428.
- [53] Tukey, J.W. (1958). "Bias and Confidence in Not-Quite Large Samples", *The Annals of Mathematical Statistics*, 29, 614.
- [54] U.S. Bureau of the Census (1949). "Illustrative Examples of Two Methods of Estimating the Current Population of Small Areas", *Current Population Report*, Ser. P-25, 20, U.S. Government Printing Office, Washington, D.C.
- [55] U.S. Bureau of Census (1966). "Methods of Population Estimation: Part 1", *Current Population Report*, Ser. P-25, 339, U.S. Government Printing Office, Washington, D.C.
- [56] U.S. Bureau of the Census (1969). "Estimates of the Population of Counties and Metropolitan Areas", July 1, 1966: A Summary Report", *Current Population Report*, Ser. P-25, 427, U.S. Government Printing Office, Washington, D.C.
- [57] Verdooren, L.R. (1988). "Exact Tests and Confidence Intervals for Ratio of Variance Components in Unbalanced Two- and Three-Stage Nested Designs", *Communications in Statistics- Theory and Methods*, A17, 1197-1230.
- [58] Wu, C.F.J. (1984). "Jackknife and Bootstrap Inference in Regression and a Class of Representations for the LSE", MRC Technical Summary Report #2675.

சென்னை, 15 சூன் -

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

END

0|2·0|3·9|2

FIN