

**Robust Estimation of Variance Components in
Small Area Estimation**

by

Golshid Chatrchi

**A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements
for the degree of**

Master of Science

in

Probability and Statistics

**Carleton University
Ottawa, Ontario**

© 2012

Golshid Chatrchi



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-93570-5

Our file Notre référence

ISBN: 978-0-494-93570-5

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

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

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Small area estimation has become more popular in recent years due to an increasing demand for reliable small area statistics. In this thesis, we mainly concentrate on the robust estimation of variance components in the context of small area estimation.

A robust method based on robustified maximum likelihood (RML) equations (Sinha and Rao, 2009) is considered under the one-fold nested error regression model. It is shown that the optimization method used, Newton-Raphson method, has some restrictions and may fail to converge. An alternative iterative algorithm, fixed-point method, is proposed and Henderson's method III estimator is also suggested for the initial values of the algorithm. Moreover, two other robust methods, Fellner's method and Huggins' method, are studied and a simulation study is conducted to compare the behaviour of these estimators under the contaminated distributions. Finally, the RML method is extended to the two-fold nested error regression model with equal error variances.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my thesis supervisors, Professor J. N. K. Rao and Professor Sanjoy K. Sinha, for their constant guidance and encouragement. They gave me tremendous support with their valuable advice and useful feedback. Indubitably, I would not have been able to complete this thesis without their help.

I would also like to thank Professors M. Alvo, J. Nielsen and J. Galdo for serving as members of my advisory committee and for their valuable comments on my thesis research.

I would like to thank Professor M. Mojirsheibani who introduced me to the field of small area estimation and taught me some statistics courses. I am also grateful to Professor I. Molina, Professor M. Torabi and Dr. L. Dumitrescu for their valuable suggestions and helpful discussions we had throughout the preparation of this thesis. My thanks also go to the School of Mathematics and Statistics, Carleton University, for providing the necessary research facilities for completing this thesis.

Finally, I wish to express my deep gratitude to my husband, Amir, for his continuous support and encouragement and without whom this thesis may never have seen the light.

To my parents, Mina and Mostafa

Table of Contents

Abstract	ii
Acknowledgments	iii
Table of Contents	v
List of Tables	viii
List of Figures	ix
List of Acronyms	x
1 Introduction	1
1.1 An Overview of Small Area Estimation	1
1.1.1 Area Level Model	2
1.1.2 Unit Level Model	4
1.1.3 General Linear Mixed Model with Block Diagonal Covariance Structure	5
1.1.4 Model-based Estimation	6
1.1.5 EBLUP	6
1.2 A Brief Overview of Robust Statistics	8
1.3 Thesis Outline	11

2	Robust Small Area Estimation	12
2.1	RML method	12
2.2	Practical Consideration of RML Method	16
2.2.1	Fixed-Point approach	16
2.2.2	Henderson's method III	19
2.2.3	Example	21
2.3	Other Robust Methods	26
2.3.1	Fellner's Method	26
2.3.2	Huggins' Method	29
2.4	Simulation Study	32
3	Robust Estimation Under a Two-Fold Nested Error Regression	
	Model	36
3.1	Two-Fold Nested Error Regression Model	36
3.1.1	Description of the model	37
3.1.2	EBLUP estimator	38
3.2	Robust Estimation	39
3.2.1	Robust estimator of β	40
3.2.2	Robust estimator of θ	40
3.3	Simulation Study	46
4	Summary and Future Research	50
4.1	Summary	50
4.2	Future Work	51
A		53
A.1	ML Equations	53
A.2	Derivative of $\Phi(\theta)$	55

A.3	Estimating random effects	56
A.4	Explicit form of \mathbf{V}_i^{-1} for the two-fold nested error regression model .	58
	References	63

List of Tables

2	ML and RML estimates of model parameters (Newton-Raphson vs. fixed-point)	21
3	Simulated biases and mean squared error of model parameters :No contamination	33
4	Simulated biases and mean squared error of model parameters: Contamination in random errors	34
5	Simulated biases and mean squared error of model parameters: Contamination in area-specific random effects	34
6	Simulated biases and mean squared error of model parameters: Contamination in both random errors and area-specific random effects . .	35
7	Simulated biases and mean squared errors of the estimated model parameters- 5% contamination	48
8	Simulated biases and mean squared errors of the estimated model parameters- 10% contamination	49

List of Figures

1	Variance components estimation: Newton-Raphson vs. fixed-point . .	22
2	Newton-Raphon method: Choosing an alternative starting value . . .	23
3	Newton-Raphon method: Setting constraints	24

List of Acronyms

Acronyms	Definition
ANOVA	analysis of variance
BLUE	best linear unbiased estimator
BLUP	best linear unbiased prediction
EBLUP	empirical best linear unbiased prediction
EM	expectation-maximization
GLS	generalized least square
ML	maximum likelihood
MME	mixed model equations
MSE	mean squared error
OLS	ordinary least square
REML	restricted maximum likelihood
RML	robustified maximum likelihood

Chapter 1

Introduction

The growing demand for reliable small area statistics from both the public and private sectors has made small area estimation an important topic in survey sampling. A comprehensive accounts of methods and models of small area estimation have been studied in the literature. In recent years, model based indirect estimators have been widely used in small area estimation. These estimators are efficient under model idealized assumptions, but they can be unduly influenced by deviations from assumed structures. Robust approaches have been developed to provide resistant estimators for small area estimation. Section 1.1 of this chapter presents an overview of small area estimation, while section 1.2 gives a brief review of robust statistics. Finally, section 1.3 provides the thesis outline.

1.1 An Overview of Small Area Estimation

The motivation of most survey sampling is to provide reliable estimates for the total population of interest. It is also of interest to produce similar estimates for sub-populations or domains. A domain may denote a geographical area, such as a county, a province or a socio-demographic area like a census division. Direct domain estimators use only domain-specific sample data. These estimators are not reliable for small

domains due to insufficient domain-specific sample sizes. This necessitates the use of indirect estimators that borrow strength from other related domains. Such estimators are based on implicit or explicit models that provide a link to related areas through supplementary data such as census information and administrative data. Generally, indirect domain estimators can be classified into two categories, traditional estimators based on implicit linking models and model-based estimators based on explicit linking models.

Traditional estimators such as synthetic estimators assume that small areas have the same characteristic as the larger ones. These estimators may have small mean squared error if the implicit linking model is close to the true model. However, violation from assumptions can cause high bias.

Model-based indirect estimators define a model so that related areas are incorporated in the estimation process. The most commonly used techniques for small area estimation are the empirical best linear unbiased prediction (EBLUP), the empirical Bayes (EB) procedures and the hierarchical Bayes (HB). These estimators have some advantages for small area estimation. For instance, optimal estimators can be obtained under assumed models with respect to the nature of the response variable. These models can be classified into two types:

- (i) Area level models that relate small area direct estimators to area-specific covariates.
- (ii) Unit level models that relate the unit values of study variable to unit-specific covariates.

We study these models and their properties in sections 1.1.1 and 1.1.2.

1.1.1 Area Level Model

A basic area level model, known as the Fay-Herriot model (Fay and Herriot, 1979), assumes that the area mean, \bar{Y}_i , is related to area-specific auxiliary data. Such

models are used when unit level data are not available. A representation of the basic area level model is given by:

$$\bar{y}_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, k,$$

where \bar{y}_i is a direct estimator, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ is a $p \times 1$ vector of regression coefficients, v_i are area-specific random effects and e_i are the sampling errors. It is assumed that v_i and e_i are independent and distributed as $N(0, \sigma_v^2)$ and $N(0, \sigma_{ei}^2)$, where σ_{ei}^2 is assumed to be known.

The above model comes from the combination of the sampling model and the linking model. More precisely, it is assumed that the population area mean, \bar{Y}_i has the following form and the area-specific auxiliary data $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ are available for area $i = 1, \dots, k$ (consider only the case where all areas are sampled):

$$\bar{Y}_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i, \quad i = 1, \dots, k. \quad (1.1.1)$$

For making inference about the \bar{Y}_i , we assume that a direct survey estimators \bar{y}_i is available that is,

$$\bar{y}_i = \bar{Y}_i + e_i, \quad i = 1, \dots, k, \quad (1.1.2)$$

where e_i is the sampling error. Combining (1.1.1) and (1.1.2) we obtain the following model

$$\bar{y}_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, k. \quad (1.1.3)$$

It should be mentioned that (1.1.3) contains both design-induced errors, e_i , and the model errors, v_i .

1.1.2 Unit Level Model

A one-fold unit level model relates the unit values of the study variable, y_{ij} , to unit-specific auxiliary data, x_{ij} . It assumes that unit-specific covariates are available for all the N_i population elements in each small area. The population model is given by

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, N_i, \quad (1.1.4)$$

where v_i are assumed to be iid random variables with mean zero and variance σ_v^2 . In addition, the sampling errors e_{ij} are assumed to be iid random variables with mean zero and variance σ_e^2 . It is assumed that v_i and e_{ij} are independent. Also, normality of the v_i s and e_{ij} s is assumed. If N_i is large and the sampling fraction $f_i = n_i/N_i$ is ignorable, the mean of the i^{th} area can be written as:

$$\mu_i = \bar{\mathbf{X}}_i^T \boldsymbol{\beta} + v_i, \quad i = 1, \dots, k,$$

where $\bar{\mathbf{X}}_i$ is the known mean of x_{ij} for the i^{th} area. We assume that the population model holds for the sample. That is, we assume sample selection bias is absent. Hence our sample model is

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}; \quad i = 1, \dots, k; \quad j = 1, \dots, n_i,$$

where $n_i (\geq 1)$ is the sample size in area i .

In section 1.1.3, we introduce the general linear mixed model with block diagonal covariance structure. A special case of the general linear model with a block diagonal covariance structure covers the area level model and the unit level model.

1.1.3 General Linear Mixed Model with Block Diagonal Covariance Structure

The two basic models, area level model and unit level model, and their extensions can be presented as a general linear mixed model with block diagonal covariance structure. The term linear mixed model denotes the mixture of fixed and random effects in a linear model. A general linear mixed model is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \quad (1.1.5)$$

where \mathbf{y} is the $n \times 1$ vector of sample observation, \mathbf{X} is the $n \times p$ known matrix, \mathbf{Z} is the $n \times h$ known matrix of full rank, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown fixed effects, \mathbf{v} and \mathbf{e} are independent random vectors with mean zero and covariance matrices \mathbf{G} and \mathbf{R} that depend on some variance parameters, $\boldsymbol{\theta} = (\sigma_1^2, \dots, \sigma_q^2)$. We further assume that $\boldsymbol{\theta}$ belongs to a specified subset of Euclidean q -space such that $Var(\mathbf{y}) = \mathbf{V} = \mathbf{V}(\boldsymbol{\theta}) = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^T$ be non-singular for all $\boldsymbol{\theta}$ belonging to the subset, where $Var(\mathbf{y})$ denotes the variance-covariance matrix of \mathbf{y} . Fixed effects are introduced to model the mean of \mathbf{y} and the random effects are considered to control the variance-covariance of \mathbf{y} ; see McCulloch and Searle (2001). The goal is to estimate a linear combination of fixed effects, $\boldsymbol{\beta}$, and the realized values of, \mathbf{v} : $\boldsymbol{\mu} = \mathbf{l}^T\boldsymbol{\beta} + \mathbf{m}^T\mathbf{v}$, where \mathbf{l} and \mathbf{m} are specified constant vectors.

A special case of the general linear mixed model (1.1.5) with block diagonal covariance structure covers the basic area level model and unit level models.

A block diagonal matrix is a square matrix with square matrices (\mathbf{A}_k) as diagonal elements and $\mathbf{0}$ s as off-diagonal elements. It is denoted as $\mathbf{A} = diag_{1 \leq i \leq k}(\mathbf{A}_i)$.

For (1.1.5), $\mathbf{y} = col_{1 \leq i \leq k}(\mathbf{y}_i)$, $\mathbf{X} = col_{1 \leq i \leq k}(\mathbf{X}_i)$, $\mathbf{Z} = diag_{1 \leq i \leq k}(\mathbf{Z}_i)$, $\mathbf{v} = col_{1 \leq i \leq k}(\mathbf{v}_i)$, $\mathbf{e} = col_{1 \leq i \leq k}(\mathbf{e}_i)$, where k is the number of small areas, \mathbf{y}_i is the $n_i \times 1$ vector of sample observation in the i^{th} area, \mathbf{X}_i is $n_i \times p$ and \mathbf{Z}_i is $n_i \times h_i$

($\sum_{i=1}^k n_i = n$ and $\sum_{i=1}^k h_i = h$).

Covariance matrix \mathbf{V} has the block diagonal structure:

$$\mathbf{V} = \text{diag}_{1 \leq i \leq k}(\mathbf{V}_i) = \text{diag}_{1 \leq i \leq k}(\mathbf{R}_i + \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i^T), \text{ where } \mathbf{R} = \text{diag}_{1 \leq i \leq k}(\mathbf{R}_i) \text{ and } \mathbf{G} = \text{diag}_{1 \leq i \leq k}(\mathbf{G}_i).$$

The model (1.1.5), with a block diagonal covariance matrix \mathbf{V} can be written as k sub-models:

$$\mathbf{y}_i = \mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{v}_i + \mathbf{e}_i, \quad i = 1, \dots, k. \quad (1.1.6)$$

Considering (1.1.6), the interest is to predict

$$\mu_i = \mathbf{l}_i^T \boldsymbol{\beta} + m_i^T \mathbf{v}_i, \quad i = 1, \dots, k.$$

In this study we focus on model (1.1.6) with a block diagonal covariance structure.

1.1.4 Model-based Estimation

1.1.5 EBLUP

For the general linear mixed model (1.1.6) with block diagonal covariance structure, the best linear unbiased prediction (BLUP) estimator of $\mu_i = \mathbf{l}_i^T \boldsymbol{\beta} + \mathbf{m}_i^T \mathbf{v}_i$ is

$$\tilde{\mu}_i(\boldsymbol{\theta}, \mathbf{y}_i) = \tilde{\mathbf{l}}_i^T \tilde{\boldsymbol{\beta}} + \mathbf{m}_i^T \mathbf{G}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\beta}}), \quad (1.1.7)$$

where

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left(\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i \right)$$

is the generalized least square (GLS) estimator of $\boldsymbol{\beta}$ (Henderson, 1975).

Since (1.1.7) depends on the unknown vector of variance components, $\boldsymbol{\theta}$, we replace $\boldsymbol{\theta}$ by an estimator $\hat{\boldsymbol{\theta}}$ to obtain the empirical BLUP (EBLUP) estimator $\hat{\mu}_i = \tilde{\mu}_i(\hat{\boldsymbol{\theta}}, \mathbf{y})$.

For example, assuming normality, we can use maximum likelihood (ML) or restricted

maximum likelihood (REML) estimator of $\boldsymbol{\theta}$. Under the area level model, using (1.1.7) the BLUP estimator of μ_i is given by:

$$\tilde{\mu}_i(\boldsymbol{\theta}, \mathbf{y}_i) = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}^* + \gamma_i(\mathbf{y}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}), \quad (1.1.8)$$

where

$$\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_{ei}^2),$$

and

$$\tilde{\boldsymbol{\beta}}^* = \left(\sum_{i=1}^k \mathbf{x}_i \mathbf{x}_i^T / (\sigma_v^2 + \sigma_{ei}^2) \right)^{-1} \left(\sum_{i=1}^k \mathbf{x}_i \mathbf{y}_i / (\sigma_v^2 + \sigma_{ei}^2) \right)$$

and the EBLUP of μ_i is given by $\hat{\mu}_i = \tilde{\mu}_i(\hat{\boldsymbol{\theta}}, \mathbf{y}_i)$.

Similarly, for the unit level model, the BLUP of $\mu_i = \bar{\mathbf{X}}_i \boldsymbol{\beta} + v_i$ is given by

$$\tilde{\mu}_i(\boldsymbol{\theta}, \mathbf{y}_i) = \bar{\mathbf{X}}_i^T \tilde{\boldsymbol{\beta}} + \gamma_i(\bar{\mathbf{y}}_i - \bar{\mathbf{x}}_i^T \tilde{\boldsymbol{\beta}}), \quad (1.1.9)$$

where

$$\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / n_i),$$

and

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i \right).$$

Again, ML or REML estimators of σ_e^2 and σ_v^2 may be used to get the EBLUP estimator of μ_i .

Empirical best linear unbiased prediction (EBLUP) estimators of small area means under basic area level and basic unit level models can be highly affected by the occurrence of outliers. Robust approaches have been developed to remedy this problem. These methods down weight influential observations in the data and provide resistant estimators for model parameters.

1.2 A Brief Overview of Robust Statistics

The motivation of robust statistics is to produce estimators that are not affected by small deviations from idealized assumptions. Examples of departures from the model assumptions include the presence of outliers in the data, rounding and grouping errors, and departure from an assumed sample distribution. Outliers are often the most risky type of errors. Generally, an outlier is an observation which is numerically far from the rest of the data. Outliers happen very frequently in real data mainly due to measurement errors. Misplaced decimal points, recording errors, exceptional phenomena and including members of other populations are common source of outliers. These observations often remain hidden to users and do not show up in the usual residual plots and may spoil the results of statistical analysis. Many robust and resistant methods have been developed to provide estimators that are less sensitive to outliers.

A comprehensive history of robust estimation is provided by Hampel et al. (1986). They noted that the discussion of outliers and robust methods date back to Bernoulli (1777). First formal rules were proposed by Peirce (1852) and Chauvenet (1863), followed by Stone (1873), Thompson (1935), Pearson and Chandra (1936) and many others. In 1960, Tukey summarized earlier work and suggested some useful robust alternatives that made robust estimation a general research area. The foundation of robust estimation was made by Huber (1964) and Hampel (1968). Huber's (1964) paper "Robust estimation of a location parameter" is an important pioneering work. He introduced a resistant class of estimators, called "M-estimators", with useful properties like consistency and asymptotic normality.

Robust methods of direct area-specific estimators of a population total Y have been studied under a model-based approach, (Valliant et al., 2000). Chambers (1986) and Welsh and Ronchetti (1998) worked on robust estimation of Y when outliers occur

in data. An outlier-robust method was proposed as an alternative to the customary regression estimator $\hat{Y} = \sum_{i \in s} y_i + \sum_{i \in \bar{s}} (\hat{\beta} x_i)$ based on the ratio model $y_i = \beta x_i + \epsilon_i$, $i = 1, \dots, N$, where ϵ_i 's are independent and have mean zero and variance $\sigma_i^2 \propto x_i$, $\hat{\beta}$ is the weighted least-squares estimator of β , s is the sample and \bar{s} is the set of non-sample units. The robust estimator is significantly more efficient than \hat{Y} when ϵ_i has mixture distribution $\epsilon_i = (1 - \delta_i)\epsilon_{1i} + \delta_i\epsilon_{2i}$, where δ_i indicates a random variable with small $\pi_i = P(\delta_i = 1)$, and the variance of ϵ_{2i} is much larger than the variance of ϵ_{1i} . In this mixture model, only a few of the y -values may be outliers.

In general, robust estimators can be classified as M-estimators, L-estimators and R-estimators (Hampel et al., 1986). M-estimators are based on maximum likelihood estimators, L-estimators are a linear combination of order statistics and R-estimators rely on ranks of residuals. M-estimators are the most widely used robust estimators. These estimators replace the square of residuals with an alternative loss function of the data, $\{y_i, i = 1, \dots, n\}$ and the parameters, θ , to limit the effect of outliers. More generally, M-estimation attempts to minimize $\sum_{i=1}^n \rho(y_i, \theta)$, for a specified function ρ with derivative ψ . Two popular ψ -functions are the Tukey biweight (also known as bisquare) and the Huber's ψ .

• Tukey biweight

Tukey (1970) proposed an M-estimator that excludes any extreme outlier by giving it a zero weight. The ρ -function for a specified tuning constant, c , is given by

$$\rho_c(r) = \begin{cases} \frac{1}{6}(c^6 - (c^2 - r^2)^3) & \text{if } |r| < c \\ 0 & \text{if } |r| \geq c \end{cases}$$

The corresponding derivative $\psi(r)$ is given by

$$\psi_c(r) = \begin{cases} r(c^2 - r^2)^2 & \text{if } |r| < c \\ 0 & \text{if } |r| \geq c \end{cases}$$

• **Huber**

Huber (1964) suggested the following M-estimator that applies a least absolute deviation to outliers:

$$\rho_c(r) = \begin{cases} \frac{1}{2}r^2 & \text{if } |r| < c \\ c|r| - \frac{1}{2}c^2 & \text{if } |r| \geq c \end{cases}$$

and

$$\psi_c(r) = \begin{cases} r & \text{if } |r| < c \\ c \operatorname{sign}(r) & \text{if } |r| \geq c \end{cases} \quad (1.2.1)$$

For Tukey's biweight, under standard normal distribution 95% asymptotic efficiency will be obtained when $c=4.6851$; for Huber's function the corresponding value is $c=1.345$. For more information about the choice of the tuning constant, refer to Hampel et al. (1986).

Regarding outliers in small area estimation, Datta and Lahiri (1995) proposed a robust Bayesian method using a Cauchy distribution on the contaminated areas and a scale mixture of normal for the rest of the data. They assumed that outlying areas are known. In addition, Bell and Huang (2006) introduced a method for the area-level model (1.1.3) based on t-distribution with $k(> 2)$ degree of freedom to deal with outliers. They assumed that outliers can occur in model errors v_i or the sampling errors e_i . Fellner (1986) considered robust estimation of β and \mathbf{v} under the general linear mixed model (1.1.5), using Huber's (1973) ψ -function. He solved robust "mixed

model” equations to get $\tilde{\beta}$ and $\tilde{\nu}$ for given θ . Sinha and Rao (2009) proposed another method, based on Robustified Maximum Likelihood (RML) equations, which is more efficient than Fellner’s method. In this thesis, we mainly focus on Sinha and Rao’s RML method.

1.3 Thesis Outline

The main focus of this thesis is on the robust estimation of variance components for the unit level model in the presence of outliers. Chapter 2 deals with the RML method (Sinha and Rao, 2009). The model is described and a number of practical considerations are highlighted. Some difficulties with the Newton-Raphson method is outlined and an alternative iterative method is proposed for solving the RML equations. Also, Henderson’s method III estimates are suggested for the starting values of the iterative algorithm. Then an example is provided to present the failure of the Newton-Raphson method. Moreover, two other robust methods, Fellner’s method and Huggins’ method, are considered. The methods are described and a simulation study is conducted to assess the performance of these estimators and the RML estimator under the contamination model proposed by Sinha and Rao (2009).

Chapter 3 extends the RML method to the two-fold nested error regression model with equal error variances. This model is appropriate for the two-stage sampling design. Similar to Chapter 2, the fixed-point method is used for solving the RML equations and Henderson’s method III estimates are chosen as the initial values of the iterative algorithm. In the end, the result of a simulation study is reported.

Chapter 4 summarizes the results of the thesis and offers suggestions for future research.

Chapter 2

Robust Small Area Estimation

Various definitions of robust ML and REML have been investigated so far. Stahel and Welsh (1997) and Richardson and Welsh (1995) discussed a number of these methods. In this chapter, we primarily consider a recent robust method by (Sinha and Rao, 2009) based on ML estimators. Then, in section 2.3 two other methods, namely Fellner's (1986) and Huggins' (1993) methods, are studied.

2.1 RML method

Sinha and Rao (2009) considered the general linear mixed model with block diagonal covariance structure, namely

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{v}_i + \mathbf{e}_i \quad i = 1, \dots, k,$$

where the vectors \mathbf{y}_i are independent with means equal to $\mathbf{X}_i\boldsymbol{\beta}$ and variances $\mathbf{V}_i = \mathbf{R}_i + \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i^T = \mathbf{V}_i(\boldsymbol{\theta})$, where $\mathbf{V}_i(\boldsymbol{\theta})$ depends on the vector of unknown variance components, $\boldsymbol{\theta}$. see Rao (2003), page 107. Under normality, maximum likelihood

(ML) estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be obtained by solving the following equations:

$$\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) = 0, \quad (2.1.1)$$

$$\sum_{i=1}^k \left\{ (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_l} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) - \text{tr} \left(\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_l} \right) \right\} = 0, \quad l = 1, \dots, q, \quad (2.1.2)$$

where θ_l is the l^{th} component of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^T$. Equations (2.1.1) and (2.1.2) are given in Appendix A, section A.1.

One obvious indication of outliers in the data is when some of the fitted values, $\hat{\mathbf{y}}_i$, are unusually far from the observed values, \mathbf{y}_i . To handle such outliers Sinha and Rao (2009) robustified the ML equations (2.1.1) and (2.1.2) as follows:

$$\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) = 0, \quad (2.1.3)$$

$$\phi_l(\boldsymbol{\theta}) = \sum_{i=1}^k \left\{ \boldsymbol{\psi}^T(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_l} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) - \text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_l}) \right\} = 0 \quad l = 1, \dots, q, \quad (2.1.4)$$

where $\boldsymbol{\psi}(\cdot)$ is the Huber ψ -function given by (1.2.1), $\mathbf{r}_i = \mathbf{U}_i^{-1/2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})$, where \mathbf{U}_i is a diagonal matrix with diagonal elements equal to diagonal elements of \mathbf{V}_i , \mathbf{K}_i is a diagonal matrix chosen as $\mathbf{K}_i = c \mathbf{I}_i$ where c is the tuning constant, $c = E[\psi_b^2(u)]$ and $u \sim N(0, 1)$, and \mathbf{I}_i is the identity matrix of the same order as \mathbf{V}_i .

In order to obtain the robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, equations (2.1.3) and (2.1.4) need to be solved iteratively. For solving (2.1.3), the Newton-Raphson method based on a first-order Taylor series expansion around an initial value, $\boldsymbol{\beta}_0$, is used to create

a set of linear equations. We can show that

$$\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \psi(\mathbf{r}_i(\boldsymbol{\beta})) \approx \sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \psi(\mathbf{r}_i(\boldsymbol{\beta}_0)) - \sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i(\boldsymbol{\beta}_0) \mathbf{X}_i (\boldsymbol{\beta} - \boldsymbol{\beta}_0), \quad (2.1.5)$$

where $\mathbf{D}_i(\boldsymbol{\beta})$ is a diagonal matrix with its j^{th} diagonal element $D_{ij} = \psi'_b(r_{ij}) = (\partial/\partial r_{ij})\psi_b(r_{ij})$. The derivative of the Huber ψ -function with respect to r_{ij} is 1 if $|r_{ij}| \leq b$ and 0 otherwise.

From (2.1.5), iterative equations of the following form are obtained:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left\{ \sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i(\boldsymbol{\beta}^{(m)}) \mathbf{X}_i \right\}^{-1} \left\{ \sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \psi(\mathbf{r}_i(\boldsymbol{\beta}^{(m)})) \right\}, \quad (2.1.6)$$

for the m^{th} iteration ($m = 0, 1, 2, \dots$).

For (2.1.4), Sinha and Rao (2009) suggested using the Newton-Raphson method.

Newton-Raphson Algorithm

The Newton-Raphson method is commonly used for maximizing non-linear functions and for finding approximations to the roots of a real-valued function. The main idea of the algorithm is to construct a quadratic approximation to the function of interest around some initial values (first guess) and adjust the parameter value to that which maximizes the quadratic approximation. The process is repeated until convergence is reached.

Suppose a function f is defined over the real x , and its derivative f' exists. Iterations are given by

$$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}, \quad t = 1, 2, \dots$$

Accordingly, for the equation (2.1.4) the iterative equations are given by

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} - \left\{ \boldsymbol{\Phi}'(\boldsymbol{\theta}^{(m)}) \right\}^{-1} \boldsymbol{\Phi}(\boldsymbol{\theta}^{(m)}), \quad (2.1.7)$$

where $\Phi(\boldsymbol{\theta}) = (\phi_1, \dots, \phi_q)^T$ with its l^{th} element as defined in (2.1.4) and $\Phi'(\boldsymbol{\theta})$ is the derivative of $\Phi(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. The derivative of $\Phi(\boldsymbol{\theta})$ is obtained in Appendix A.2.

The complete algorithm involving both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be presented as follows

1. Choose initial values of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.
Set $m = 0$.
2. Calculate $\boldsymbol{\beta}^{m+1}$ and $\boldsymbol{\theta}^{m+1}$ from iterative equations (2.1.6) and (2.1.7).
Set $m = m + 1$.
3. Continue step 2 until convergence is achieved. Let $\hat{\boldsymbol{\beta}}_M$ and $\hat{\boldsymbol{\theta}}_M$ denote the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ at convergence.

After computing the robust estimators $\hat{\boldsymbol{\beta}}_M$ and $\hat{\boldsymbol{\theta}}_M$, area-specific random effects \mathbf{v}_i can be estimated by solving the following equation by (Fellner, 1986).

$$\mathbf{Z}_i^T \mathbf{R}_i^{-1/2} \psi \left(\mathbf{R}_i^{-1/2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{v}_i) \right) - \mathbf{G}_i^{-1/2} \psi \left(\mathbf{G}_i^{-1/2} \mathbf{v}_i \right) = 0, \quad i = 1, \dots, k. \quad (2.1.8)$$

Again, the Newton-Raphson method is used to solve the above equations. Then the robust estimators $\hat{\mathbf{v}}_{iM}$ are used to predict $\boldsymbol{\mu}_i = \mathbf{l}_i^T \boldsymbol{\beta} + \mathbf{m}_i^T \mathbf{v}_i$ as $\hat{\boldsymbol{\mu}}_{iM} = \mathbf{l}_i^T \hat{\boldsymbol{\beta}}_M + \mathbf{m}_i^T \hat{\mathbf{v}}_{iM}$ for specified vectors \mathbf{l}_i and \mathbf{m}_i . The predictor $\hat{\boldsymbol{\mu}}_{iM}$ is called the robust EBLUP estimator (REBLUP) of $\boldsymbol{\mu}_i$ and it is the same as EBLUP for the choice of $\psi_b(r) = r$.

Sinha and Rao's simulation results revealed that the RML method is significantly more efficient than Fellner's (1986) method for estimating $\boldsymbol{\theta}$, described in Chapter 3. However the Newton-Raphson method used for obtaining robust estimates of parameters may have some limitations. Section 2.2 explains some difficulties with the Newton-Raphson algorithm and suggests an alternative approach.

2.2 Practical Consideration of RML Method

Newton-Raphson's method usually converges very fast towards the solution. However, there are some difficulties with the method and in some situations the method fails to converge. Newton-Raphson method requires that the derivative be calculated directly, but the derivative may not be easily obtainable. For example, the convergence may fail if the function is not continuously differentiable in a neighbourhood of the root or if the derivative is not continuous at the root. Also, the Newton-Raphson method is very sensitive to the starting point. Kellison (1975) summarized the conditions under which the method may fail. The stability of the Newton-Raphson method may be affected if the initial value is taken at or near a maximum/minimum point or it is taken at or near a point of inflection.

Using Newton-Raphson optimization for estimating the variance components, we may encounter some of the above difficulties and in certain conditions the algorithm does not converge. Therefore, it is preferable to use a technique that does not need the derivative computation. In addition, choosing the starting value (first guess) of the algorithm is an important issue to be considered. Section 2.2.1 presents an alternative method, fixed-point iteration, that does not require the derivative computation. Section 2.2.2 suggests using Henderson's (1953) method III estimators of variance components θ as the initial value of the fixed-point algorithm.

2.2.1 Fixed-Point approach

The fixed-point iterative method relies on the fixed-point theorem:

"If $g(x)$ is a continuous function for all $x \in [a, b]$, then g has a fixed point in $[a, b]$." This can be proven by assuming that $g(a) \geq a$ and $g(b) \leq b$. Since g is continuous the intermediate value theorem guarantees that there exists a c such that $g(c) = c$.

Fixed-Point iterative method

Suppose x_0 is the root of a function $f(x)$. Then f can be manipulated algebraically to produce $x_0 = g(x_0)$, where x_0 is the fixed point of g . The algorithm of this method can be presented as:

1. Choose an initial value, x_0
2. Compute $x_{n+1} = g(x_n)$ for $n = 0, 1, 2, \dots$
3. Continue step 2 until convergence is achieved.

Anderson (1973) proposed the fixed point algorithm for obtaining ML estimates of linear mixed models (see Rao (2003), chapter 6). Fried et al. (2011) also used fixed-point algorithm for RML equations in the specified case of a mean nested error model: $y_{ij} = \mu + u_i + e_{ij}$. We apply this method to the RML equations (2.1.4) under the nested error regression models with respect to $\boldsymbol{\theta}$ where $\boldsymbol{\theta} = (\sigma_e^2, \sigma_v^2)^T$. Consider RML equation with respect to $\boldsymbol{\theta}$:

$$\sum_{i=1}^k \left\{ \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) - \text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l}) \right\} = 0.$$

This equation can be written as

$$\sum_{i=1}^k \left\{ \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) - \text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{-1} \mathbf{V}_i) \right\} = 0, \quad (2.2.1)$$

using $\mathbf{V}_i^{-1} \mathbf{V}_i = \mathbf{I}_{n_i}$, where \mathbf{I}_{n_i} is the identity matrix of the same order as \mathbf{V}_i . Then by replacing \mathbf{V}_i with $\mathbf{V}_i = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}'_{n_i}$ (2.2.1) changes to

$$\sum_{i=1}^k \left\{ \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) - \text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{-1} \begin{pmatrix} \mathbf{I}_{n_i} & \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \end{pmatrix} \begin{pmatrix} \sigma_e^2 \\ \sigma_v^2 \end{pmatrix}) \right\} = 0.$$

Evaluating $\frac{\partial \mathbf{V}_i}{\partial \theta_i}$, we get

$$\sum_{i=1}^k \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) = \sum_{i=1}^k \text{tr} \left(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} (\mathbf{I}_{n_i} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) \begin{pmatrix} \sigma_e^2 \\ \sigma_v^2 \end{pmatrix} \right),$$

$$\sum_{i=1}^k \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) = \sum_{i=1}^k \text{tr} \left(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \mathbf{V}_i^{-1} (\mathbf{I}_{n_i} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) \begin{pmatrix} \sigma_e^2 \\ \sigma_v^2 \end{pmatrix} \right).$$

Let

$$\begin{pmatrix} \sum_{i=1}^k \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{U}_i^{1/2} \boldsymbol{\psi}^t(\mathbf{r}_i) \\ \sum_{i=1}^k \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \mathbf{U}_i^{1/2} \boldsymbol{\psi}^t(\mathbf{r}_i) \end{pmatrix} = a(\boldsymbol{\theta}).$$

Then

$$\boldsymbol{\theta} = \begin{pmatrix} \sigma_e^2 \\ \sigma_v^2 \end{pmatrix} = [\mathbf{A}(\boldsymbol{\theta})]^{-1} a(\boldsymbol{\theta}),$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{i=1}^k \text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} \mathbf{I}_{n_i}) & \sum_{i=1}^k \text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) \\ \sum_{i=1}^k \text{tr}(\mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \mathbf{V}_i^{-1} \mathbf{I}_{n_i}) & \sum_{i=1}^k \text{tr}(\mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) \end{bmatrix}.$$

So, the fixed point algorithm of RML equations for the m^{th} iteration ($m=0,1,2,\dots$) can be presented as follows:

$$\boldsymbol{\theta}^{(m+1)} = \begin{pmatrix} \sigma_e^{2(m+1)} \\ \sigma_v^{2(m+1)} \end{pmatrix} = [\mathbf{A}(\boldsymbol{\theta}^{(m)})]^{-1} a(\boldsymbol{\theta}^{(m)}). \quad (2.2.2)$$

Iterative equation (2.2.2) is more stable than the Newton-Raphson method and it typically converges in 10 to 15 steps. Like any other iterative algorithm, the fixed-point algorithm requires initial values for the parameters. As a result, suggesting a well-defined starting value for $\boldsymbol{\theta}$ could facilitate the procedure. We have chosen Henderson's method III estimates as the initial values of the algorithm. Section 2.2.2 gives a brief introduction to Henderson's method III.

2.2.2 Henderson's method III

Henderson (1953) proposed three methods, known as Henderson's methods I, II and III, based on analysis of variance (ANOVA) estimation. The method III is more popular since it is valid for unbalanced cases.

The basic idea of Henderson's method III comes from decomposing the sum of squares of residuals by fitting various sub-models. First, consider the unit level model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}$ and rewrite it as

$$\mathbf{y} = \mathbf{W}\boldsymbol{\beta}^* + \mathbf{e}, \quad (2.2.3)$$

where $\mathbf{W} = (\mathbf{X}|\mathbf{Z})$ and $\boldsymbol{\beta}^* = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{v} \end{pmatrix}$. Treating $\boldsymbol{\beta}^*$ as fixed factor, the partition in total sum of squares, $SST(\boldsymbol{\beta}, \mathbf{v}) = \mathbf{y}^T \mathbf{y}$, under model (2.2.3) is given by

$$\begin{aligned} SSR(\boldsymbol{\beta}, \mathbf{v}) &= \mathbf{y}^T \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y}, \\ SSE(\boldsymbol{\beta}, \mathbf{v}) &= \mathbf{e}^T \mathbf{e} = \left[(\mathbf{I}_n - \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}) \mathbf{y} \right]^T \left[(\mathbf{I}_n - \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}) \mathbf{y} \right]. \end{aligned} \quad (2.2.4)$$

Then, reduce model (2.2.3) to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (2.2.5)$$

Similarly, the partition in total sum of squares under model (2.2.5) is given by

$$\begin{aligned}
 SST(\boldsymbol{\beta}) &= \mathbf{y}^T \mathbf{y}, \\
 SSR(\boldsymbol{\beta}) &= \mathbf{y}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \\
 SSE(\boldsymbol{\beta}) &= \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = \left[(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}) \mathbf{y} \right]^T \left[(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}) \mathbf{y} \right].
 \end{aligned} \tag{2.2.6}$$

In terms of sum of squares, dropping \mathbf{v} in (2.2.5) can be expressed as

$$SSR(\mathbf{v}|\boldsymbol{\beta}) = SSR(\boldsymbol{\beta}, \mathbf{v}) - SSR(\boldsymbol{\beta}). \tag{2.2.7}$$

Considering the expected value of (2.2.4), the unbiased estimator of σ_e^2 is given by

$$\hat{\sigma}_e^2 = \frac{SSE(\boldsymbol{\beta}, \mathbf{v})}{n - \text{rank}(\mathbf{W})}. \tag{2.2.8}$$

Similarly, using (2.2.7), the estimator of σ_v^2 is given by

$$\hat{\sigma}_v^2 = \frac{SSE(\mathbf{v}|\boldsymbol{\beta}) - \hat{\sigma}_e^2(\text{rank}(\mathbf{W}) - \text{rank}(\mathbf{X}))}{\text{tr}[\mathbf{Z}^T (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) \mathbf{Z}]}. \tag{2.2.9}$$

Estimators (2.2.8) and (2.2.9) are used as the initial values in the fixed-point algorithm (2.2.2) to get robust estimates of σ_e^2 and σ_v^2 .

Henderson's method III estimators are unbiased and they are valid for all mixed models. A detailed account of Henderson's method III is given by Searle et al. (1992). Section 2.2.3, provides an example that illustrates the failure of Newton-Raphson method in solving RML equations.

2.2.3 Example

In this section, we provide an example that demonstrates the failure of the Newton-Raphson method. A single data set is generated from the unit-level model and the parameters of the model are estimated by the RML method using the fixed-point iteration and the Newton-Raphson iteration. The following model with a single auxiliary variable is considered:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij} \quad i = 1, \dots, 40 \quad j = 1, \dots, 4, \quad (2.2.10)$$

where x_{ij} is generated from $N(1, 1)$ normal distribution with mean equal to 1 and standard deviation equal to 1 and $(\beta_0, \beta_1) = (1, 1)$. Further, both v_i and e_{ij} are generated from the contaminated distribution $(1 - 0.05)N(0, 1) + 0.05N(0, 25)$, which means 0.95 proportion of v_i and e_{ij} comes from the true distribution $N(0, 1)$ and the remaining 0.05 proportion is generated from the arbitrary contaminated distribution $N(0, 25)$. We focus on the estimation of parameters (β_0, β_1) and (σ_e^2, σ_v^2) using the RML method with fixed-point iteration and the Newton-Raphson method. The starting value for (σ_e^2, σ_v^2) is chosen as Henderson's method III estimators and for (β_0, β_1) ordinary least squares (OLS) estimators are used. For the generated data set, we have $(\beta_0^{(0)}, \beta_1^{(0)}) = (0.47, 1.32)$ and $(\sigma_e^{2(0)}, \sigma_v^{2(0)}) = (2.35, 4.49)$.

Table 2: ML and RML estimates of model parameters (Newton-Raphson vs. fixed-point)

Coefficients	ML	RML(Newton-Raphson)	RML(fixed-point)
β_0	0.59	0.61	0.89
β_1	1.21	1.19	1.17
σ_e^2	2.33	3.27	1.96
σ_v^2	4.36	NA	1.22

Table 2 shows the result of estimation with fixed-point iteration and Newton-Raphson. It shows that using the fixed-point iteration appears to be more stable than the Newton-Raphson method. Particularly for σ_v^2 , the Newton-Raphson method does not converge.

Figure 1 presents the process of Newton-Raphson and fixed-point iterations for finding the robust estimates of the variance components, σ_e^2 and σ_v^2 . As can be seen in Figure 1, choosing Henderson's method III estimates for the starting values, the Newton-Raphson fails to converge in estimating σ_v^2 . Also, the fixed-point method provides a better estimate for σ_e^2 than the Newton-Raphson method. A debating point of this figure is the spike in the σ_e^2 iterations, when using the Newton-Raphson method. A small change in the iterative values can drastically change the gradient matrix in equations (2.1.6) or (2.1.7) and that causes a sudden fluctuation.

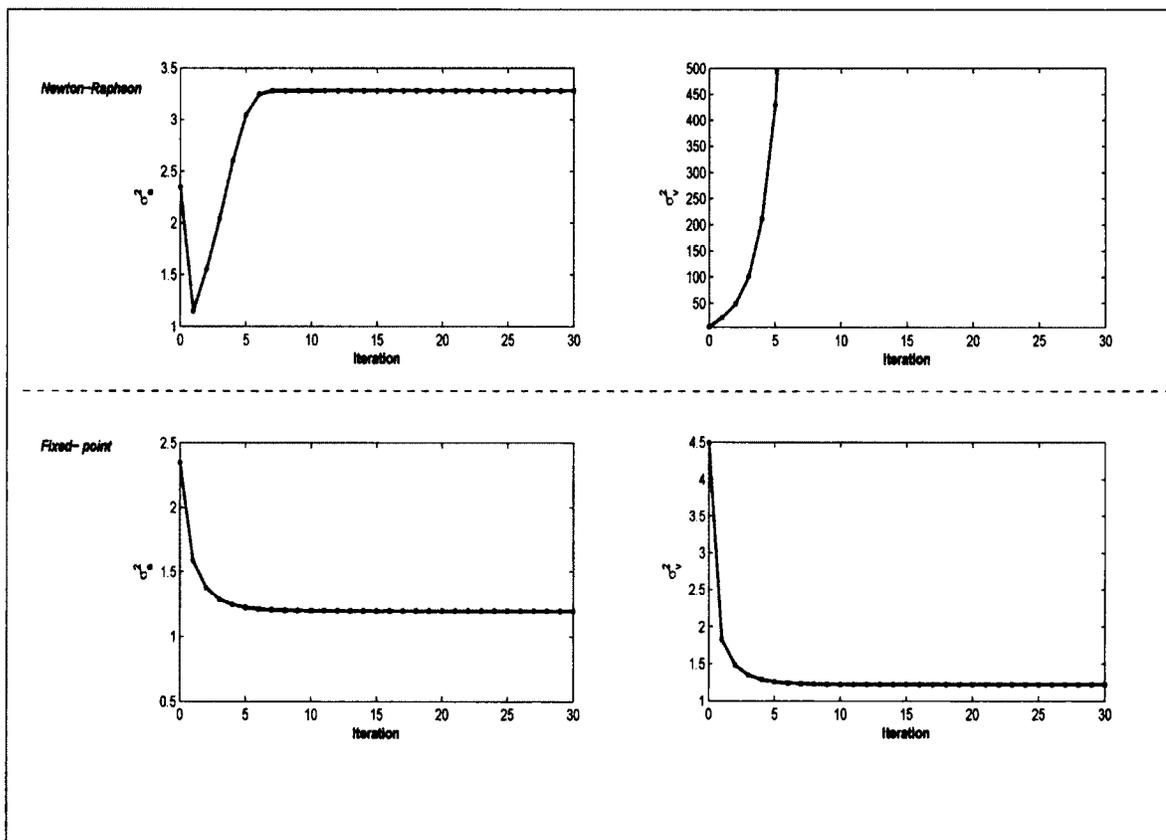


Figure 1: Variance components estimation: Newton-Raphson vs. fixed-point

Repeating the procedure on different data sets, we noticed that the Newton-Raphson algorithm may converge to values outside the parameter space (negative estimate of σ_e^2 and σ_v^2). In such cases, the gradient matrix in equation (2.1.6) becomes computationally singular.

Some alternative approaches are suggested by West et al. (2007) to remedy these problems. For example, choosing an alternative starting value may solve the problem. For the generated data set, choosing the initial values of the variance components very close to the true value, the Newton-Raphson algorithm converges and it provides the same results as the fixed-point iteration. Figure 2 presents the Newton-Raphson iteration, choosing the initial value very close to the true value. This method is only good for simulation studies where the true values are known.

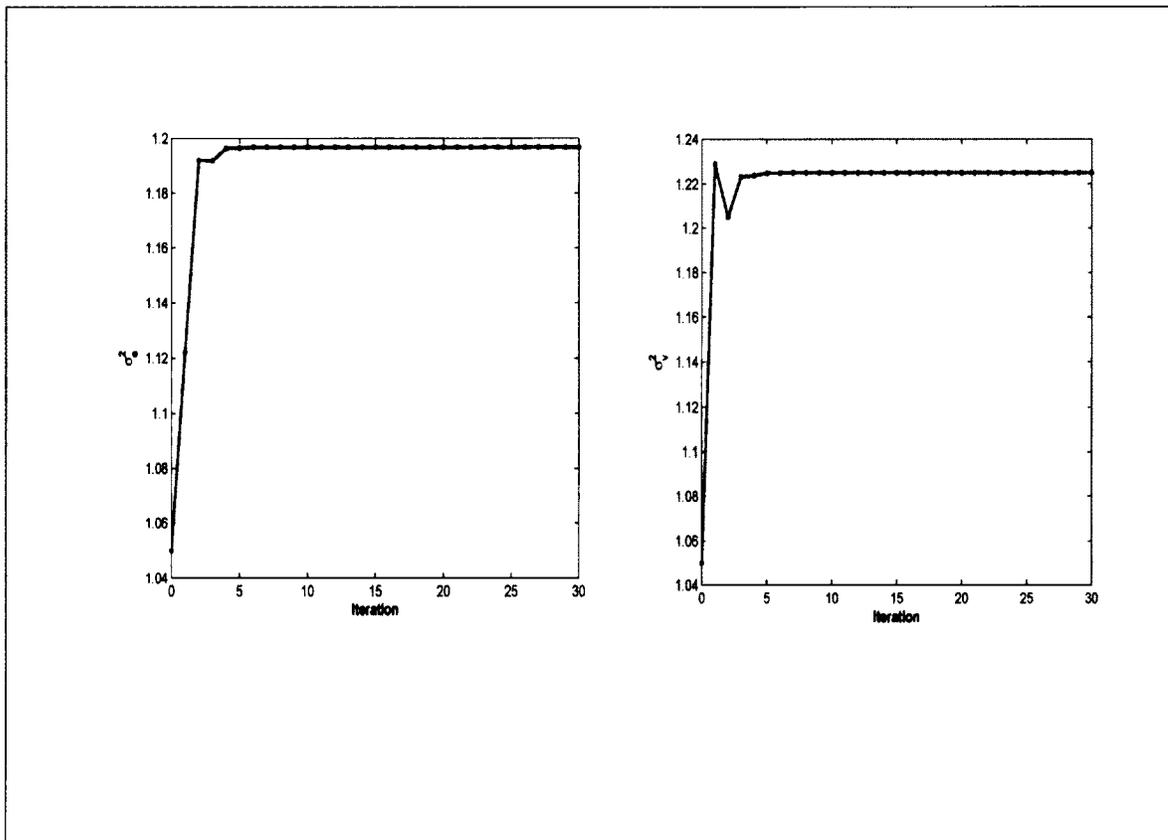


Figure 2: Newton-Raphon method: Choosing an alternative starting value

Moreover, setting some constraints for iterative values may solve the computational problem with the Newton-Raphson algorithm. Defining some boundaries for iterative estimates controls the Newton-Raphson procedure and guarantees convergence to some extent. For the data set generated from the model (2.2.10) with specified β_0 , β_1 , σ_e^2 and σ_v^2 , setting the following constraints leads to convergence.

- If $\sigma_e^{2(m)} < 0$ then set $\sigma_e^{2(m)} = 0.01$.
- If $\sigma_e^{2(m)} > 10$ then set $\sigma_e^{2(m)} = 1.5$.
- If $\sigma_v^{2(m)} < 0$ then set $\sigma_v^{2(m)} = 0.01$.
- If $\sigma_v^{2(m)} > 10$ then set $\sigma_v^{2(m)} = 1.5$.

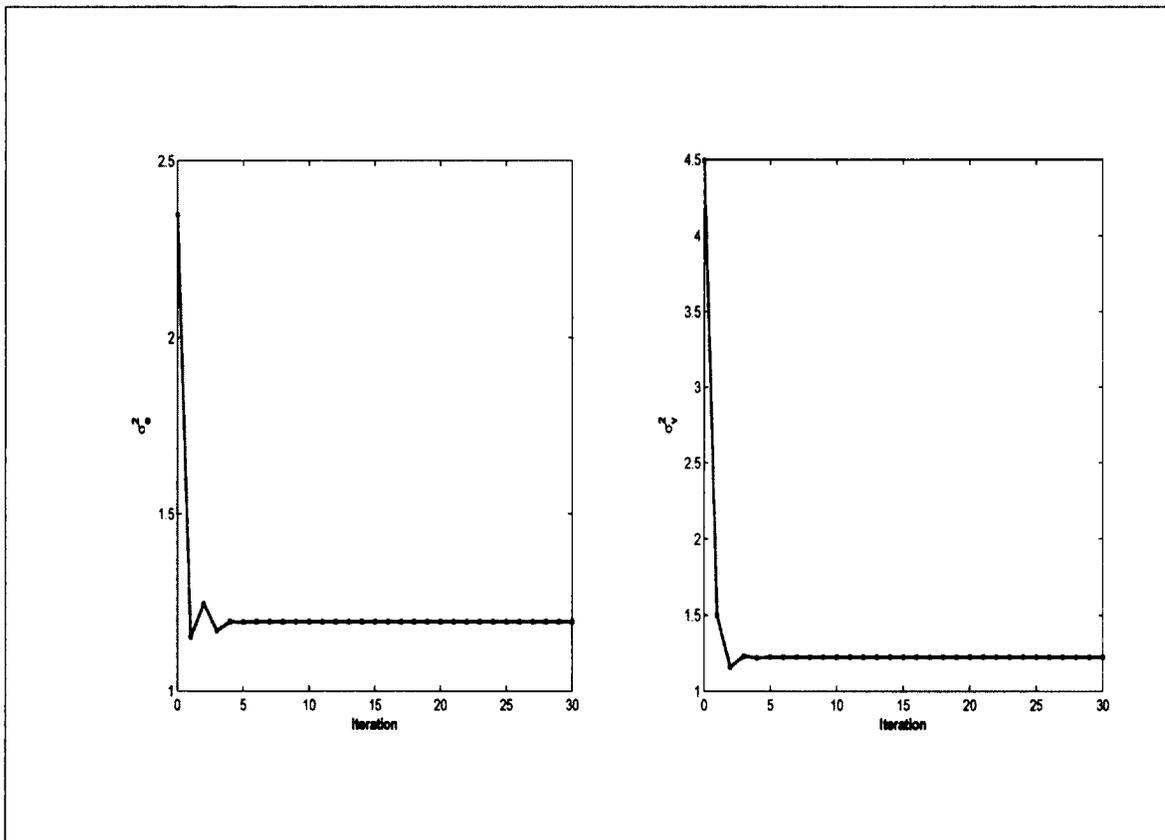


Figure 3: Newton-Raphson method: Setting constraints

Although setting such restrictions solves the problem, defining a criterion for the side conditions is not straightforward specially for inference from actual data. It

should be mentioned that convergence issues with the Newton-Raphson method are not limited to robust estimates. Difficulty in convergence of Newton-Raphson algorithm is a frequent problem in the analysis of linear mixed models (LMM). In the context of LMMs, the Newton-Raphson algorithm is the most commonly used algorithm in ML and REML estimation. Although this algorithm require calculation of the partial derivative and its iterations are relatively time consuming, the convergence is faster than the other available methods such as the expectation-maximization (EM) algorithm and the Fisher scoring algorithm (Little and Rubin, 2002). For more explanations on EM algorithm and Fisher scoring algorithm may be referred to Searle et al. (1992).

The main computational struggle is the estimation of the covariance matrix, $\mathbf{V}_i = \mathbf{R}_i + \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i^T$, i.e, we need to ensure that \mathbf{G}_i and \mathbf{R}_i be positive definite. More precisely, the diagonal elements of \mathbf{G}_i and \mathbf{R}_i should be greater than zero.

Considering the above points and the computational difficulties of the Newton-Raphson method, we recommend using the fixed-point approach which is more stable and does not require the derivative computation.

The available software for fitting LMMs have different error messages or notes when problems arise. Some software stops the procedures, some others elude the positive definiteness criteria and gives estimate out of the parameter space (negative or zero estimations).

West et al. (2007) summarized the procedures of some software such as SAS, R and Stata.

- **SAS Proc Mixed** in SAS fits a variety of linear mixed models to the data. If Proc Mixed converges to an estimated solution for variance components that leads to a non-positive definite covariance matrix, SAS stops the procedure and sets the negative variance components to 0 in the output. The following note is displayed in the log window.

"Estimated G matrix is not positive definite"

The variance-covariance matrix of random effects is called the G matrix by SAS. This means that the variance components estimates might be either negative or very close to zero. In this case, SAS provides an extra option that relaxes the positive definite requirement and enables the user to investigate the problem. The **nobound** option gives an invalid estimates for the variance components.

- **R** `lme` function in package `nlme` fits a linear mixed-effects model in the formulation described in Laird and Ware (1982). When computational problems occur in estimating the covariance matrix, the estimates cannot be achieved and R displays the following message

"Error in lme.formula: iteration limit reached without convergence"

2.3 Other Robust Methods

Sections 2.1 and 2.2 focused on the RML method of Sinha and Rao (2009), explained the procedure and suggested an alternative iterative algorithm for robust estimation of the variance components. As mentioned earlier, different methods have been proposed to deal with outliers in individual observations and in the area-level random effects. In this section, we study two other robust methods of estimating variance components and compare them in terms of efficiency by conducting a simulation study.

2.3.1 Fellner's Method

Fellner (1986) proposed a method of robust estimation of β and \mathbf{v} under the linear mixed model. His method consists of two stages. First, robust estimators $\tilde{\beta}$ and $\tilde{\mathbf{v}}$ are obtained for a specified θ and then $\tilde{\beta}$ and $\tilde{\mathbf{v}}$ are substituted in the robustified Henderson's equations for the variance components. We consider the linear mixed

model with block diagonal covariance structure and mixed model equations (MME):

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{v}_i + \mathbf{e}_i; \quad i = 1, \dots, k.$$

Mixed model equations (Henderson (1959)) are derived by maximizing the joint density of $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_k^T)$ and $\mathbf{v} = (\mathbf{v}_1^T, \dots, \mathbf{v}_k^T)$ with respect to $\boldsymbol{\beta}$ and \mathbf{v} , assuming normality MME equations are given by

$$\sum_{i=1}^k \mathbf{X}_i^T \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{v}_i) = \mathbf{0}; \quad i = 1, \dots, k, \quad (2.3.1)$$

$$\mathbf{Z}_i^T \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{v}_i) - \mathbf{G}_i^{-1} \mathbf{v}_i = \mathbf{0}; \quad i = 1, \dots, k. \quad (2.3.2)$$

Equations (2.3.1) and (2.3.2) can be written compactly as

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix},$$

where $\mathbf{y} = \text{col}_{1 \leq i \leq k}(\mathbf{y}_i)$, $\mathbf{X} = \text{col}_{1 \leq i \leq k}(\mathbf{X}_i)$, $\mathbf{Z} = \text{diag}_{1 \leq i \leq k}(\mathbf{Z}_i)$, $\mathbf{G} = \text{diag}_{1 \leq i \leq k}(\mathbf{G}_i)$, $\mathbf{R} = \text{diag}_{1 \leq i \leq k}(\mathbf{R}_i)$.

Fellner (1986) robustified the MME equations (2.3.1) and (2.3.2) as follows

$$\sum_{i=1}^k \mathbf{X}_i^T \mathbf{R}_i^{-1/2} \boldsymbol{\psi}(\mathbf{R}_i^{-1/2} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{v}_i)) = \mathbf{0}; \quad i = 1, \dots, k, \quad (2.3.3)$$

and

$$\mathbf{Z}_i^T \mathbf{R}_i^{-1/2} \boldsymbol{\psi}(\mathbf{R}_i^{-1/2} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{v}_i)) - \mathbf{G}_i^{-1/2} \boldsymbol{\psi}(\mathbf{G}_i^{-1/2} \mathbf{v}_i) = \mathbf{0}; \quad i = 1, \dots, k, \quad (2.3.4)$$

where ψ_b is the Huber's ψ -function. $\psi(x) = x \min(1, b/|x|)$, where $b > 0$ is the tuning constant. A common choice of b is 1.345 that provides 95% asymptotic efficiency under the normal distribution. The choice of $b = \infty$ leads to standard BLUE $\tilde{\beta} = \tilde{\beta}(\theta)$ of β and BLUP estimator $\tilde{v} = \tilde{v}(\theta)$ of v .

Solving equations (2.3.3) and (2.3.4) gives the robust estimators $\tilde{\beta}_M(\theta)$ and $\tilde{v}_M(\theta)$ for fixed θ . Substituting these estimators in the robustified version of the Henderson's (1973) iterative algorithm provides the estimators of variance components.

For the unit level model with $\theta = (\sigma_v^2, \sigma_e^2)^T$, robustified Fellner's iterative algorithm is given by

$$\tilde{\sigma}_v^{2(m+1)} = \tilde{\sigma}_v^{2(m)} \sum_{i=1}^k \frac{\psi_b(\tilde{v}_i^{(m)}/\tilde{\sigma}_v^{(m)})^2}{\{c(k - \lambda^{(m)})\}},$$

$$\tilde{\sigma}_e^{2(m+1)} = \tilde{\sigma}_e^{2(m)} \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{\psi_b(\tilde{e}_{ij}^{(m)}/\tilde{\sigma}_e^{(m)})^2}{[c\{(n_0 - p) - (k - \lambda^{(m)})\}]},$$

where $n_0 = \sum_i n_i$, $\tilde{\sigma}_e^{2(m)}$ and $\tilde{\sigma}_v^{2(m)}$ are the estimated values in the m^{th} iteration. $c = E[\psi_b^2(u)]$ for $u \sim N(0, 1)$ and λ^m is the value of $tr(\mathbf{A}_{22})/\tilde{\sigma}_v^2$ evaluated at $\theta^{(m)}$, where

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1}.$$

Fellner calculated the tuning constant c constant as $c = P_3(r^2) + r^2[1 - P_1(r^2)]$, where P_3 is the cumulative chi-squared distribution with parameter equal to 3. He noted that for $r = 2$ the asymptotic efficiency of the estimators is 92.13% and $c = 0.9205$ which means under normality the maximum bias of variance components fluctuates between 4% to 8%. However, in the absence of normality, the biases of variance components could be significant.

It should be mentioned that a nice property of Fellner's method is the use of Henderson's equations for variance components estimation that is not affected by the constraints on the parameter space: $\sigma_e^2 > 0$, $\sigma_v^2 > 0$ (Harville, 1977). If the starting

values of the algorithm are strictly positive, then at every iteration the estimators remain positive. See Rao (2003)- page 102.

2.3.2 Huggins' Method

Huggins (1993) proposed a robust version of the likelihood for multivariate normal data. His method require the assumption that \mathbf{y} has a symmetric distribution. The log likelihood under normality for the linear model with block diagonal structure is given by

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) \propto \frac{-1}{2} \sum_{i=1}^k \log |\mathbf{V}_i| - \frac{1}{2} \sum_{i=1}^k \{(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})\}.$$

Huggins defined the corresponding robustified log likelihood as

$$l^* \propto \frac{-1}{2} \sum_{i=1}^k (\mathbf{K} \log |\mathbf{V}_i|) - \sum_{i=1}^k \rho(\tilde{\mathbf{r}}_i),$$

where $\tilde{\mathbf{r}}_i = \mathbf{V}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})$, $\mathbf{K} = E(\psi(\tilde{\mathbf{r}}_i)\tilde{\mathbf{r}}_i)$, ρ is a twice continuously differentiable even function and $\psi \equiv \rho'$.

The ML estimates can be found by minimizing (2.3.5)

$$\mathbf{M}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^k (\mathbf{K} \log |\mathbf{V}_i|) + \sum_{i=1}^k \rho(\mathbf{V}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})). \quad (2.3.5)$$

Direct differentiation of \mathbf{M} with respect to $\boldsymbol{\beta}$ and θ_l gives

$$\frac{\partial \mathbf{M}}{\partial \boldsymbol{\beta}} = \frac{\partial \sum_{i=1}^k \rho(\tilde{\mathbf{r}}_i)}{\partial \boldsymbol{\beta}} = - \sum_{i=1}^k \psi(\tilde{\mathbf{r}}_i)^T \mathbf{V}_i^{-1/2} \mathbf{X}_i, \quad (2.3.6)$$

$$\frac{\partial \mathbf{M}}{\partial \theta_l} = \frac{1}{2} \frac{\partial \left(\sum_{i=1}^k \mathbf{K} \log |\mathbf{V}_i| \right)}{\partial \theta_l} + \frac{\partial \left(\sum_{i=1}^k \rho(\tilde{\mathbf{r}}_i) \right)}{\partial \theta_l}$$

$$= \frac{\mathbf{K}}{2} \sum_{i=1}^k \text{tr}(\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l}) + \sum_{i=1}^k \boldsymbol{\psi}(\tilde{\mathbf{r}}_i)^T \frac{\partial \mathbf{V}_i^{-1/2}}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{1/2} \tilde{\mathbf{r}}_i. \quad (2.3.7)$$

Equations (2.3.6) and (2.3.7) have zero expectations. For (2.3.6),

$$E \left(- \sum_{i=1}^k \{ \boldsymbol{\psi}(\tilde{\mathbf{r}}_i)^T \mathbf{V}_i^{-1/2} \mathbf{X}_i \} \right) = - \sum_{i=1}^k \{ E \left(\boldsymbol{\psi}(\tilde{\mathbf{r}}_i)^T \mathbf{V}_i^{-1/2} \mathbf{X}_i \right) \} = 0,$$

since $\boldsymbol{\psi}$ is an odd function, $\tilde{\mathbf{r}}_i$ is symmetric and $E(\boldsymbol{\psi}(\tilde{\mathbf{r}}_i)) = 0$.

For (2.3.7), using $\frac{\partial \mathbf{V}_i^{-1/2}}{\partial \boldsymbol{\theta}_l} = -\mathbf{V}_i^{-1/2} \frac{\partial \mathbf{V}_i^{1/2}}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{-1/2}$, we can write

$$E \left(\sum_{i=1}^k \{ \boldsymbol{\psi}(\tilde{\mathbf{r}}_i)^T \frac{\partial \mathbf{V}_i^{-1/2}}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{1/2} \tilde{\mathbf{r}}_i \} \right) = - \sum_{i=1}^k \{ \mathbf{K} \text{tr}(\mathbf{V}_i^{1/2} \frac{\partial \mathbf{V}_i^{-1/2}}{\partial \boldsymbol{\theta}_l}) \}. \quad (2.3.8)$$

In addition, considering the point that $\mathbf{V}_i^{-1/2} \mathbf{V}_i^{-1/2 T} = \mathbf{V}_i^{-1}$, the first part of the (2.3.7) can be written as

$$\sum_{i=1}^k \frac{1}{2} \text{tr}(\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l}) = \sum_{i=1}^k \text{tr}(\mathbf{V}_i^{-1/2} \frac{\partial \mathbf{V}_i^{1/2}}{\partial \boldsymbol{\theta}_l}). \quad (2.3.9)$$

Equations (2.3.8) and (2.3.9) show that (2.3.7) also has zero expectation. To ensure the robustness it is required that both $\boldsymbol{\psi}(\tilde{\mathbf{r}}_i)$ and $\boldsymbol{\psi}(\tilde{\mathbf{r}}_i)\tilde{\mathbf{r}}_i$ be bounded (Richardson and Welsh, 1995).

Huggins (1993) performed a simulation study of size 100 on pedigree data with a complex covariance structure. He defined contamination via the elements of $\tilde{\mathbf{r}}_i$. More precisely, a model is denoted as $p\%$ contaminated when with probability of $1 - p$ each element of $\tilde{\mathbf{r}}_i$ is an observation from $N(0, 1)$ and with probability $1 - p$ it is an observation from $N(0, 25)$. This type of contamination does not indicate if outliers are generated in the distribution of area-specific random effect, \mathbf{v}_i , or if it is affecting the error term \mathbf{e}_i . However, in the small-area estimation context it is preferable to identify the outliers in the random effects and the random errors.

His simulation results show that under the suggested contamination model the robust method has a good efficiency for the estimation of model parameters. We study the performance of Huggins robust method for the nested error model under the contamination model proposed by Sinha and Rao (2009), where the outliers are produced in the distributions of v_i and e_{ij} .

In order to find robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, equations (2.3.6) and (2.3.7) should be solved iteratively. For equation (2.3.6), we use a first-order Taylor series expansion around $\boldsymbol{\beta}_0$:

$$\sum_{i=1}^k \{\boldsymbol{\psi}\{\tilde{\mathbf{r}}_i(\boldsymbol{\beta})\}^T \mathbf{V}_i^{-1/2} \mathbf{X}_i\} \approx \sum_{i=1}^k \{\boldsymbol{\psi}\{\tilde{\mathbf{r}}_i(\boldsymbol{\beta}_0)\}^T \mathbf{V}_i^{-1/2} \mathbf{X}_i\} - \sum_{i=1}^k \mathbf{X}_i^T (\mathbf{V}_i^{-1/2})^T \mathbf{D}_i(\boldsymbol{\beta}_0) \mathbf{V}_i^{-1/2} \mathbf{X}_i (\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

where $\mathbf{D}_i(\boldsymbol{\beta})$ is a diagonal matrix that its diagonal elements is equal to $\psi'(\tilde{r}_{ij})$. So, the iterative equation is given by

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left\{ \sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1/2 T} \mathbf{D}_i(\boldsymbol{\beta}^{(m)}) \mathbf{V}_i^{-1/2} \mathbf{X}_i \right\}^{-1} \sum_{i=1}^k \boldsymbol{\psi}\{\tilde{\mathbf{r}}_i(\boldsymbol{\beta}^{(m)})\}^T \mathbf{V}_i^{-1/2} \mathbf{X}_i. \quad (2.3.10)$$

Since the covariance matrix has a block diagonal structure, we apply the fixed-point iteration for equation (2.3.7). considering the block diagonal structure for the covariance matrix. Using the Cholesky decomposition, we can obtain the explicit form of $\mathbf{V}_i^{1/2}$ and $\mathbf{V}_i^{-1/2}$. This facilitates the computation of derivatives with respect to variance components.

Cholesky decomposition

Assume A is a symmetric positive definite square matrix, then $A = U^T U = L L^T$, where U is a unique upper triangular matrix and L is a unique lower triangular matrix (Gentle, 1998).

Since $\mathbf{V} = \text{diag}_{1 \leq i \leq k} \mathbf{V}_i = \text{diag}_{1 \leq i \leq k} \{\sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}'_{n_i}\}$, we can write

$$\mathbf{V}_i^{1/2} = \sigma_e \left(\mathbf{I}_{n_i} - \frac{1}{n_i} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \right) + (\sigma_e^2 + n_i \sigma_v^2)^{1/2} \frac{\mathbf{1}_{n_i} \mathbf{1}'_{n_i}}{n_i}, \quad (2.3.11)$$

$$\mathbf{V}_i^{-1/2} = \sigma_e^{-1} \left(\mathbf{I}_{n_i} - \frac{1}{n_i} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \right) + (\sigma_e^2 + n_i \sigma_v^2)^{-1/2} \frac{\mathbf{1}_{n_i} \mathbf{1}'_{n_i}}{n_i}. \quad (2.3.12)$$

Taking derivative of (2.3.11) and (2.3.12) with respect to $\boldsymbol{\theta}$, the iterative equation for estimating variance components can be expressed as follows

$$\boldsymbol{\theta}^{(m+1)} = \begin{pmatrix} \sigma_e^{2(m+1)} \\ \sigma_v^{2(m+1)} \end{pmatrix} = \frac{-2}{\mathbf{K}} [\mathbf{A}(\boldsymbol{\theta}^{(m)})]^{-1} a(\boldsymbol{\theta}^{(m)}),$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{i=1}^k \{tr(\mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} \mathbf{I}_{n_i})\} & \sum_{i=1}^k \{tr(\mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i})\} \\ \sum_{i=1}^k \{tr(\mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \mathbf{V}_i^{-1} \mathbf{I}_{n_i})\} & \sum_{i=1}^k \{tr(\mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}'_{n_i})\} \end{bmatrix},$$

$$a(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{i=1}^k \boldsymbol{\psi}^T(\tilde{\mathbf{r}}_i) \left[\frac{-1}{2} (\sigma_e^2)^{-3/2} (\mathbf{I}_{n_i} - \frac{1}{n_i} \mathbf{1}_{n_i} \mathbf{1}'_{n_i}) - \frac{1}{2n_i} (\sigma_e^2 + n_i \sigma_v^2)^{-3/2} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \right] \mathbf{V}_i^{1/2} \tilde{\mathbf{r}}_i \\ \sum_{i=1}^k \boldsymbol{\psi}^T(\tilde{\mathbf{r}}_i) \left[-\frac{1}{2} (\sigma_e^2 + n_i \sigma_v^2)^{-3/2} \mathbf{1}_{n_i} \mathbf{1}'_{n_i} \right] \mathbf{V}_i^{1/2} \tilde{\mathbf{r}}_i \end{bmatrix}.$$

Section 2.4 compares the performances of Huggins' method, Fellner method and the RML method under the contamination model proposed by Sinha and Rao (2009).

2.4 Simulation Study

We ran four sets of simulations with 500 simulation runs, using the nested error model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}, \quad i = 1, \dots, 40, \quad j = 1, \dots, 4,$$

where the auxiliary variable, x_{ij} , was generated from $N(1,1)$ and kept fixed in all the simulation . Further, $(\beta_0, \beta_1) = (1,1)$ and it remained unchanged for all the simulations. The area-specific random effects, v_i , and the random errors, e_{ij} , were generated from the contaminated distribution $(1-\lambda)N(0,1)+\lambda N(0,25)$, which means $(1-\lambda)$ proportion of v_i and e_{ij} comes from the true distribution $N(0,1)$ and the remaining λ proportion were generated from the arbitrary contaminated distribution $N(0,25)$. We have chosen $\lambda = 0.1$. Four possible combinations were considered:

1. No contamination: e_{ij} and v_i were generated from standard normal.
2. Contamination in random errors only: 10% of e_{ij} were generated from $N(0,25)$ and the remaining 90% were generated from $N(0,1)$.
3. Contamination in area-specific random effects only: 10% of v_i were generated from $N(0,25)$ and the remaining 90% were generated from $N(0,1)$.
4. Contamination in both random errors and area-specific random effects: 10% of e_{ij} and 10% of v_i were generated from $N(0,25)$ and the remaining 90% of e_{ij} and v_i were generated from $N(0,1)$.

Table 3 to Table 6 present the biases and mean squared errors of the estimated model parameters using the three robust methods: Fellner, Huggins and RML, and the ML method, under the four possible contaminations. When there are no outliers in the data, Huggins, RML and ML performed analogously and Huggins and RML method are as efficient as the ML method. On the contrary, Fellner's method provided poorer results and displayed larger biases and MSE.

Table 3: Simulated biases and mean squared error of model parameters :No contamination

Parameter	ML		Huggins		Fellner		RML	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0	0.000	0.05	-0.011	0.046	-0.003	0.035	-0.010	0.039
β_1	-0.002	0.006	0.007	0.007	0.007	0.008	-0.003	0.009
σ_e^2	-0.019	0.019	-0.005	0.021	0.106	0.038	0.027	0.024
σ_v^2	-0.013	0.043	-0.026	0.083	0.151	0.119	-0.045	0.091

Table 4: Simulated biases and mean squared error of model parameters: Contamination in random errors

Parameter	ML		Huggins		Fellner		RML	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0	-0.009	0.08	0.006	0.053	0.007	0.076	0.001	0.044
β_1	-0.003	0.031	-0.001	0.126	-0.009	0.025	0.005	0.013
σ_e^2	2.361	7.524	1.13	1.45	0.598	0.408	0.523	0.337
σ_v^2	-0.040	0.231	-0.199	0.122	0.282	0.192	-0.103	0.109

In the presence of outliers in the random errors (Table 4) ML behaviour changed dramatically and exhibited huge biases for variance component $\hat{\sigma}_e^2$. The RML method performed better than Huggins and Fellner's methods in terms of the bias and the efficiency. Among the robust methods, Huggins method appears to provide the worst results.

Table 5: Simulated biases and mean squared error of model parameters: Contamination in area-specific random effects

Parameter	ML		Huggins		Fellner		RML	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0	-0.001	0.09	0.018	0.082	0.006	0.091	-0.008	0.052
β_1	0.000	0.007	-0.001	0.008	-0.002	0.008	0.006	0.008
σ_e^2	0.013	0.021	-0.008	0.019	0.128	0.043	0.028	0.026
σ_v^2	2.34	7.31	1.746	5.011	0.554	0.48	0.364	0.299

Table 6: Simulated biases and mean squared error of model parameters: Contamination in both random errors and area-specific random effects

Parameter	ML		Huggins		Fellner		RML	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0	0.015	0.132	-0.002	0.091	0.033	0.134	-0.002	0.064
β_1	0.000	0.031	-0.005	0.013	-0.002	0.024	0.005	0.015
σ_e^2	2.342	6.321	1.240	1.73	0.637	0.473	0.586	0.404
σ_v^2	2.271	8.716	1.337	3.11	0.748	0.785	0.411	0.347

When we consider outliers in the random effect v_i , Table 5 shows that the RML method provides better results than other methods. Huggins' method is heavily influenced by the outliers and produced larger biases and mean squared errors comparing to other two robust methods.

Considering outliers in both random errors and random effects, still RML appears to provide better result than Huggins and the Fellner's methods. Similarly, Fellner's method performed better than Huggins, but less efficient than the RML method.

Chapter 3

Robust Estimation Under a Two-Fold Nested Error Regression Model

Chapter 2 dealt with the one-fold nested error regression model. An extension of this model involves two levels of nested errors. In this chapter, we study robust estimation of model parameters under the two-fold nested error regression model. As stated earlier, the focus of this thesis is on Sinha and Rao's (2009) RML method. We apply the RML method to the two-fold nested error regression model and investigate its behaviour in the presence of outliers by performing a simulation study.

3.1 Two-Fold Nested Error Regression Model

The two-fold nested error regression model is a natural extension of the one-fold nested error regression model. Fuller and Battese (1973) studied this model for balanced case with constant variances. Stukel and Rao (1999) considered the two-fold nested error regression model with unequal error variances and obtained the BLUP estimator of a small area mean explicitly.

3.1.1 Description of the model

Under the two-fold nested error regression model, the i^{th} small area has M_i clusters or primary units, where the j^{th} cluster in the i^{th} area contains N_{ij} elements or subunits. Suppose (y_{ijl}, x_{ijl}) is the response value and the auxiliary information for the l^{th} unit in the j^{th} cluster from the i^{th} area, where $l = 1, \dots, N_{ij}; j = 1, \dots, M_i; i = 1, \dots, k$. A sample, s_i , of m_i clusters is selected from the i^{th} small area and in the next step a sub-sample, s_{ij} , of n_{ij} units is selected from the j^{th} sample cluster.

A representation of the two-fold nested error regression model for the sample is given by Stukel and Rao (1999):

$$y_{ijl} = \mathbf{x}_{ijl}^T \boldsymbol{\beta} + v_i + u_{ij} + e_{ijl} \quad l = 1, \dots, n_{ij}; j = 1, \dots, m_i; i = 1, \dots, k, \quad (3.1.1)$$

where $v_i \sim iid(0, \sigma_v^2)$, $u_{ij} \sim iid(0, \sigma_u^2)$ and $e_{ijl} = k_{ijl} \tilde{e}_{ijl}$ with $\tilde{e}_{ijl} \sim iid(0, \sigma_e^2)$ and known constant k_{ijl} . In addition, $\{v_i\}$, $\{u_{ij}\}$ and $\{\tilde{e}_{ijl}\}$ are mutually independent.

We focus on Stukel and Rao's model (3.1.1) with equal error variances by assuming $k_{ijl} = 1$. This model is the special case of the general linear mixed model with block diagonal covariance structure.

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{v}_i + \mathbf{e}_i; \quad i = 1, \dots, k, \quad (3.1.2)$$

where

$$\begin{aligned} \mathbf{y}_i &= col_{1 \leq j \leq m_i}(\mathbf{y}_{ij}), \quad \mathbf{y}_{ij} = col_{1 \leq l \leq n_{ij}}(y_{ijl}), \quad \mathbf{X}_i = col_{1 \leq j \leq m_i}(col_{1 \leq l \leq n_{ij}}(\mathbf{x}_{ijl}^T)), \\ \mathbf{e}_i &= col_{1 \leq j \leq m_i}(col_{1 \leq l \leq n_{ij}}(e_{ijl})), \quad \mathbf{Z}_i = (\mathbf{z}_i \mid \mathbf{Z}_{2i}), \\ \mathbf{z}_i &= col_{1 \leq j \leq m_i}(col_{1 \leq l \leq n_{ij}}(1)) = col_{1 \leq j \leq m_i}(\mathbf{z}_{ij}), \quad \mathbf{Z}_{2i} = diag_{1 \leq j \leq m_i}(\mathbf{z}_{ij}), \\ \mathbf{v}_i &= \begin{pmatrix} v_i \\ \mathbf{v}_{2i} \end{pmatrix}, \quad \mathbf{v}_{2i} = col_{1 \leq j \leq m_i}(\mathbf{u}_{ij}), \text{ and } \boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T. \end{aligned}$$

In addition, \mathbf{e}_i and \mathbf{v}_i are independently distributed with mean $\mathbf{0}$ and covariance matrices \mathbf{R}_i and \mathbf{G}_i depending on the vector of variance components, $\boldsymbol{\theta}$.

Note that the covariance matrix of $\mathbf{y} = \text{col}_{1 \leq i \leq k}(\mathbf{y}_i)$ has block diagonal structure with blocks $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{R}_i + \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i^t$, where $\mathbf{G}_i = \text{var}\left(\begin{smallmatrix} v_i \\ \mathbf{v}_{2i} \end{smallmatrix}\right) = \begin{pmatrix} \sigma_v^2 & \mathbf{0}^T \\ \mathbf{0} & \sigma_u^2 \mathbf{I}_{m_i} \end{pmatrix}$, $\mathbf{R}_i = \text{diag}_{1 \leq j \leq m_i}(\text{diag}_{1 \leq l \leq n_{ij}}(\sigma_e^2 k_{ijl}^2))$.

3.1.2 EBLUP estimator

We wish to predict the population mean of the i^{th} small area, μ_i , under model (3.1.1). As mentioned in Chapter 1, for the general model (3.1.2) the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and the best linear unbiased prediction (BLUP) estimator of \mathbf{v}_i are given by

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left(\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i \right), \quad (3.1.3)$$

and

$$\tilde{\mathbf{v}}_i(\boldsymbol{\theta}) = \mathbf{G}_i \mathbf{Z}_i^t \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\beta}}),$$

where

$\boldsymbol{\theta} = (\sigma_e^2, \sigma_v^2, \sigma_u^2)^t$ and $\mathbf{V}_i(\boldsymbol{\theta}) = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 (\mathbf{1}_{n_i})(\mathbf{1}'_{n_i}) + \sigma_u^2 \text{diag}_{1 \leq j \leq m_i}(\mathbf{1}_{n_{ij}})(\mathbf{1}'_{n_{ij}})$. We can obtain the EBLUP estimator of μ_i by replacing $\boldsymbol{\theta}$ by its estimator $\hat{\boldsymbol{\theta}}$. Under normality maximum likelihood (ML) or restricted ML (REML) estimator of $\boldsymbol{\theta}$ may be used. The EBLUP estimator of μ_i is given by

$$\hat{\mu}_i = \bar{\mathbf{X}}_i \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}) + \tilde{\mathbf{v}}_i(\hat{\boldsymbol{\theta}}). \quad (3.1.4)$$

Stukel (1991) facilitated the computations of (3.1.4) by obtaining an explicit form of \mathbf{V}_i^{-1} ; see Appendix A.4. It should be mentioned that obtaining the explicit forms of the estimators may not be the top priority any more, due to the the development and universal availability of statistical software that facilitates the computations. However, for complex models with high dimensional data using the explicit forms

decreases the running time of the program.

3.2 Robust Estimation

As mentioned earlier, the EBLUP estimators of the small area means can be affected significantly by the presence of outliers in the data. This calls for using methods of estimation that limit the influence of outliers and produce more reliable estimators. In this section, we apply Sinha and Rao's (2009) method to the two-fold model and similar to Chapter 2 we focus on robust estimation of $\boldsymbol{\beta}$ and the variance components, σ_e^2 , σ_v^2 , σ_u^2 .

We consider Stukel and Rao's (1993) model (3.1.1) with equal variances ($k_{ijl} = 1$):

$$y_{ijl} = \mathbf{x}_{ijl}^T \boldsymbol{\beta} + v_i + u_{ij} + e_{ijl} \quad l = 1, \dots, n_{ij}; \quad j = 1, \dots, m_i; \quad i = 1, \dots, k. \quad (3.2.1)$$

In section 3.1.1, it is shown that (3.2.1) is a special case of the general linear mixed model with block diagonal covariance structure. The RML method could be easily extended to the two-fold model, since the RML estimators are defined based on the ML equations of the general linear mixed model with block diagonal covariance.

First, rewrite the model (3.2.1) as $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{v}_i + \mathbf{e}_i$ and then apply the following RML equations.

$$\sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) = 0, \quad (3.2.2)$$

$$\phi_l(\boldsymbol{\theta}) = \sum_{i=1}^k \left\{ \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_l} \mathbf{V}_i^{-1} \boldsymbol{\psi}(\mathbf{r}_i) - \text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_l}) \right\} = 0, \quad l = 1, \dots, q \quad (3.2.3)$$

where $\boldsymbol{\psi}(\cdot)$ is the Huber $\boldsymbol{\psi}$ -function (1.2.1), $\mathbf{r}_i = \mathbf{U}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})$ where $\mathbf{r}_i =$

$col_{1 \leq j \leq m_i}(\mathbf{r}_{ij}) = col_{1 \leq j \leq m_i}(col_{1 \leq l \leq n_{ij}}(r_{ijl}))$, \mathbf{U}_i is a diagonal matrix with diagonal elements equal to diagonal elements of \mathbf{V}_i , \mathbf{K}_i is a diagonal matrix chosen as $\mathbf{K}_i = c\mathbf{I}_i$ where c is the tuning constant, $c = E[\psi_b^2(u)]$ and $u \sim N(0, 1)$ and \mathbf{I}_i is the identity matrix of the same order as \mathbf{V}_i .

To find robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta} = (\sigma_e^2, \sigma_v^2, \sigma_u^2)^T$, equations (3.2.2) and (3.2.3) should be solved iteratively. Sections 3.2.1 and 3.2.2 deal with robust estimation of $\boldsymbol{\beta}$ and variance components $\boldsymbol{\theta}$ using RML method.

3.2.1 Robust estimator of $\boldsymbol{\beta}$

Using the Newton-Raphson method for (3.2.2), the iterative equations are given by

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left\{ \sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i(\boldsymbol{\beta}^{(m)}) \mathbf{X}_i \right\}^{-1} \left\{ \sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i(\boldsymbol{\beta}^{(m)})) \right\}, \quad (3.2.4)$$

where $\mathbf{D}_i(\boldsymbol{\beta})$ is a diagonal matrix with its j^{th} diagonal element $\mathbf{D}_{ij} = \psi'_b(\mathbf{r}_{ij}) = (\partial/\partial \mathbf{r}_{ij})\psi_b(\mathbf{r}_{ij})$ with $\mathbf{r}_{ij} = col_{1 \leq l \leq n_{ij}}(r_{ijl})$.

3.2.2 Robust estimator of $\boldsymbol{\theta}$

In order to obtain the robust estimator of variance components, equation (3.2.3) needs to be solved iteratively. To avoid the difficulties with the Newton-Raphson method, we used the fixed-point method. In addition, Henderson's method III estimates are used as the initial values of the algorithm.

Fixed-point iteration

This section is similar to section 2.2.1. Equation (3.2.3) can be written as follows

$$\sum_{i=1}^k \left\{ \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) - tr(\mathbf{K}_i \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_l} \mathbf{V}_i^{-1} \mathbf{V}_i) \right\} = 0. \quad (3.2.5)$$

Letting $\mathbf{T}_i = \text{diag}_{1 \leq m \leq j}(\mathbf{1}_{n_{ij}} \mathbf{1}_{n_{ij}}^t)$ and replacing V_i with $V_i = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t + \sigma_u^2 \text{diag}_{1 \leq m \leq j}(\mathbf{1}_{n_{ij}} \mathbf{1}_{n_{ij}}^t)$, equation (3.2.5) leads to

$$\boldsymbol{\theta} = [\mathbf{A}(\boldsymbol{\theta})]^{-1} a(\boldsymbol{\theta}),$$

where

$$\boldsymbol{\theta} = (\sigma_e^2, \sigma_v^2, \sigma_u^2)^t,$$

$$a(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{i=0}^k \{\boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \mathbf{I}_{n-i} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i)\} \\ \sum_{i=0}^k \{\boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \mathbf{1}_{n-i} \mathbf{1}_{n-i}^T \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i)\} \\ \sum_{i=0}^k \{\boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \mathbf{T}_i \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i)\} \end{bmatrix}$$

and,

$$\mathbf{A}(\boldsymbol{\theta}) =$$

$$\begin{bmatrix} \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} \mathbf{I}_{n_i})\} & \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t)\} & \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{I}_{n_i} \mathbf{V}_i^{-1} \mathbf{T}_i)\} \\ \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i (\mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t \mathbf{V}_i^{-1} \mathbf{I}_{n_i}))\} & \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t)\} & \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t \mathbf{V}_i^{-1} \mathbf{T}_i)\} \\ \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{T}_i \mathbf{V}_i^{-1} \mathbf{I}_{n_i})\} & \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{T}_i \mathbf{V}_i^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t)\} & \sum_{i=0}^k \{\text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{T}_i \mathbf{V}_i^{-1} \mathbf{T}_i)\} \end{bmatrix}.$$

Accordingly, the iterative equation has the following form

$$\boldsymbol{\theta}^{(m+1)} = [\mathbf{A}(\boldsymbol{\theta}^{(m)})]^{-1} a(\boldsymbol{\theta}^{(m)}). \quad (3.2.6)$$

Equation (3.2.6) needs a starting guess. Similar to Chapter 2, we choose Henderson's method III estimates for the starting values of $\boldsymbol{\theta}$. Hence, the complete algorithm for the two fold model can be described as follows:

1. Choose initial values of $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$ (Henderson's type III estimates).

Set $m = 0$.

2. Calculate $\boldsymbol{\beta}^{(m+1)}$ and $\boldsymbol{\theta}^{(m+1)}$ from iterative equations (3.2.4) and (3.2.6).

Set $m = m + 1$.

3. Continue step 2 until convergence is achieved.

Henderson's method III

Henderson's method III provides the estimations of variance components by fitting various sub-models. For the two-fold nested error regression model, (3.2.1), three steps are needed for estimating σ_e^2 , σ_u^2 and σ_v^2 .

First of all, the two-fold model, (3.1.1), can be written as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + (\mathbf{z}_i | \mathbf{Z}_{2i}) \begin{pmatrix} v_i \\ \mathbf{v}_{2i} \end{pmatrix} + \mathbf{e}_i; \quad i = 1, \dots, k, \quad (3.2.7)$$

where

$$\begin{aligned} \mathbf{y}_i &= \text{col}_{1 \leq j \leq m_i}(\text{col}_{1 \leq l \leq n_{ij}}(y_{ijl})), & \mathbf{X}_i &= \text{col}_{1 \leq j \leq m_i}(\text{col}_{1 \leq l \leq n_{ij}}(\mathbf{x}_{ijl}^T)), \\ \mathbf{e}_i &= \text{col}_{1 \leq j \leq m_i}(\text{col}_{1 \leq l \leq n_{ij}}(e_{ijl})), & \mathbf{z}_i &= \text{col}_{1 \leq j \leq m_i}(\text{col}_{1 \leq l \leq n_{ij}}(1)) = \text{col}_{1 \leq j \leq m_i}(\mathbf{z}_{ij}), \\ \mathbf{Z}_{2i} &= \text{diag}_{1 \leq j \leq m_i}(\mathbf{z}_{ij}), & \mathbf{v}_i &= \begin{pmatrix} v_i \\ \mathbf{v}_{2i} \end{pmatrix}, \mathbf{v}_{2i} = \text{col}_{1 \leq j \leq m_i}(\mathbf{u}_{ij}). \end{aligned}$$

The compact form of (3.2.7) is given by

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + (\mathbf{Z} | \mathbf{Z}_2) \begin{pmatrix} \mathbf{v} \\ \mathbf{v}_2 \end{pmatrix} + \mathbf{e}, \quad (3.2.8)$$

where

$$\begin{aligned} \mathbf{y} &= \text{col}_{1 \leq i \leq k}(\mathbf{y}_i) & \mathbf{X} &= \text{col}_{1 \leq i \leq k}(\mathbf{X}_i) \\ \mathbf{Z} &= \text{diag}_{1 \leq i \leq k}(\mathbf{z}_i) & \mathbf{Z}_2 &= \text{diag}_{1 \leq i \leq k}(\mathbf{Z}_{2i}) \\ \mathbf{v} &= \text{col}_{1 \leq i \leq k}(v_i) & \mathbf{v}_2 &= \text{col}_{1 \leq i \leq k}(\mathbf{v}_{2i}) \end{aligned}$$

and $\mathbf{e} = \text{col}_{1 \leq i \leq k}(\mathbf{e}_i)$.

Clearly, model (3.2.8) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{Z}_2\mathbf{v}_2 + \mathbf{e}. \quad (3.2.9)$$

Considering the fact that the equation (3.2.9) is a special case of the general linear mixed model, applying Henderson's method III is quite easy. First, we can rewrite (3.2.9) as

$$\mathbf{y} = \mathbf{W}_1\boldsymbol{\beta}^* + \mathbf{e}, \quad (3.2.10)$$

where

$\mathbf{W}_1 = (\mathbf{X}|\mathbf{Z}|\mathbf{Z}_2)$ and $\boldsymbol{\beta}^* = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{v} \\ \mathbf{v}_2 \end{pmatrix}$. Assuming $\boldsymbol{\beta}^*$ is a fixed effect, the partition of the total sum of squares, $SST(\boldsymbol{\beta}, \mathbf{v}, \mathbf{v}_2) = \mathbf{y}^t\mathbf{y}$ under model (3.2.10) is given by

$$\begin{aligned} SSR(\boldsymbol{\beta}, \mathbf{v}, \mathbf{v}_2) &= \mathbf{y}^t\mathbf{W}_1(\mathbf{W}_1^T\mathbf{W}_1)^{-1}\mathbf{W}_1^T\mathbf{y}, \\ SSE(\boldsymbol{\beta}, \mathbf{v}, \mathbf{v}_2) &= \mathbf{e}^t\mathbf{e} = \left[(\mathbf{I}_n - \mathbf{W}_1(\mathbf{W}_1^T\mathbf{W}_1)^{-1}\mathbf{W}_1^T)\mathbf{y} \right]^T \left[(\mathbf{I}_n - \mathbf{W}_1(\mathbf{W}_1^T\mathbf{W}_1)^{-1}\mathbf{W}_1^T)\mathbf{y} \right]. \end{aligned} \quad (3.2.11)$$

Analogously, reducing model (3.2.10) to

$$\mathbf{y} = \mathbf{W}_2\boldsymbol{\beta}^{**} + \boldsymbol{\epsilon}, \quad (3.2.12)$$

where

$\mathbf{W}_2 = (\mathbf{X}|\mathbf{Z})$ and $\boldsymbol{\beta}^{**} = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{v} \end{pmatrix}$. We can write

$$\begin{aligned} SST(\boldsymbol{\beta}, \mathbf{v}) &= \mathbf{y}^t\mathbf{y}, \\ SSR(\boldsymbol{\beta}, \mathbf{v}) &= \mathbf{y}^t\mathbf{W}_2(\mathbf{W}_2^T\mathbf{W}_2)^{-1}\mathbf{W}_2^T\mathbf{y}, \\ SSE(\boldsymbol{\beta}, \mathbf{v}) &= \boldsymbol{\epsilon}^t\boldsymbol{\epsilon} = \left[(\mathbf{I}_n - \mathbf{W}_2(\mathbf{W}_2^T\mathbf{W}_2)^{-1}\mathbf{W}_2^T)\mathbf{y} \right]^T \left[(\mathbf{I}_n - \mathbf{W}_2(\mathbf{W}_2^T\mathbf{W}_2)^{-1}\mathbf{W}_2^T)\mathbf{y} \right]. \end{aligned} \quad (3.2.13)$$

Reducing model (3.2.10) once again, we have

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma}. \quad (3.2.14)$$

The partition of total sum of squares, $SST(\boldsymbol{\beta}) = \mathbf{y}^t\mathbf{y}$, under model (3.2.14) is given by

$$\begin{aligned} SSR(\boldsymbol{\beta}) &= \mathbf{y}^t\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}, \\ SSE(\boldsymbol{\beta}) &= \boldsymbol{\gamma}^t\boldsymbol{\gamma} = \left[(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X})\mathbf{y} \right]^T \left[(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X})\mathbf{y} \right]. \end{aligned} \quad (3.2.15)$$

Taking the expectation of SSE in (3.2.11), we can obtain an unbiased estimator of σ_e^2 as

$$\begin{aligned} \hat{\sigma}_e^2 &= \frac{SSE(\boldsymbol{\beta}, \mathbf{v}, \mathbf{v}_2)}{n - \text{rank}(\mathbf{W}_1)} \\ &= \frac{\sum_{i=1}^k \sum_{j=1}^{m_i} \sum_{l=1}^{n_{ijl}} \hat{e}_{ijl}^{*2}}{n - \text{rank}(\mathbf{W}_1)}, \end{aligned} \quad (3.2.16)$$

where \hat{e}_{ijl}^{*2} is the residual corresponding to $(y_{ijl}, \mathbf{x}_{ijl})$ by fitting model (3.2.10) as a fixed effects model. Considering the reduced models, (3.2.12) and (3.2.14), we can write

$$SSR(\mathbf{v}_2|\boldsymbol{\beta}, \mathbf{v}) = SSR(\boldsymbol{\beta}, \mathbf{v}, \mathbf{v}_2) - SSR(\boldsymbol{\beta}, \mathbf{v}), \quad (3.2.17)$$

$$SSR(\mathbf{v}_2, \mathbf{v}|\boldsymbol{\beta}) = SSR(\boldsymbol{\beta}, \mathbf{v}, \mathbf{v}_2) - SSR(\boldsymbol{\beta}).$$

The corresponding expected values of (3.2.17) are given by

$$\begin{aligned} E[SSR(\mathbf{v}_2|\boldsymbol{\beta}, \mathbf{v})] &= \text{tr}(\mathbf{Z}_2^t\mathbf{M}_1\mathbf{Z}_2)\hat{\sigma}_u^2 + [\text{rank}(\mathbf{W}_1) - \text{rank}(\mathbf{W}_2)]\hat{\sigma}_e^2 \\ E[SSR(\mathbf{v}_2, \mathbf{v}|\boldsymbol{\beta})] &= \text{tr}(\mathbf{Z}_2^t\mathbf{M}_2\mathbf{Z}_2)\hat{\sigma}_u^2 + \text{tr}(\mathbf{Z}\mathbf{M}_2\mathbf{Z})\hat{\sigma}_v^2 + [\text{rank}(\mathbf{W}_1) - \text{rank}(\mathbf{X})]\hat{\sigma}_e^2, \end{aligned} \quad (3.2.18)$$

where $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{W}_2[\mathbf{W}_2^t\mathbf{W}_2]^{-1}\mathbf{W}_2^t$ and $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{X}[\mathbf{X}^t\mathbf{X}]^{-1}\mathbf{X}^t$.

This gives unbiased estimators of $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ as

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{m_i} \sum_{l=1}^{n_{ijl}} \hat{e}_{ijl}^{**2} - (n - \text{rank}(\mathbf{W}_2))\hat{\sigma}_e^2}{\text{tr}(\mathbf{Z}_2^t \mathbf{M}_1 \mathbf{Z}_2)}, \quad (3.2.19)$$

where \hat{e}_{ijl}^{**2} denotes the residuals of model (3.2.13), treating $\boldsymbol{\beta}^{**}$ as fixed factor.

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{m_i} \sum_{l=1}^{n_{ijl}} \hat{e}_{ijl}^{***2} - (n - \text{rank}(\mathbf{X}))\hat{\sigma}_e^2 - \text{tr}(\mathbf{Z}_2^t \mathbf{M}_2 \mathbf{Z}_2)\hat{\sigma}_u^2}{\text{tr}(\mathbf{Z} \mathbf{M}_2 \mathbf{Z})}, \quad (3.2.20)$$

where \hat{e}_{ijl}^{***2} is the residual corresponding to $(y_{ijl}, \mathbf{x}_{ijl})$, obtained by fitting model (3.2.15).

Estimators (3.2.16), (3.2.19) and (3.2.20) are used as the starting values of $\boldsymbol{\theta}$ in the iterative equations (3.2.6).

Section 3.3 investigates the the behaviour of two-fold model in the presence of outliers.

3.3 Simulation Study

We performed 200 simulation runs, using the two-fold nested error regression model

$$y_{ijl} = \mathbf{x}_{ijl}^T + v_i + u_{ij} + e_{ijl} \quad l = 1, \dots, 5; \quad j = 1, \dots, 4; \quad i = 1, \dots, 40.$$

We assumed that the number of units in each cluster and the number clusters in each area are equal : $n_{ij} = n = 5$ and $m_i = m = 4$.

The auxiliary variable, \mathbf{x}_{ij} , was generated from $N(1, 1)$ and kept fixed in all the simulation runs. In addition, $(\beta_0, \beta_1) = (1, 1)$ and it remained unchanged for all the simulation runs. The area-specific random effects, v_i , the cluster-specific random effects, u_{ij} and the random errors, e_{ijl} , were generated from the contaminated distribution $(1 - \lambda)N(0, 1) + \lambda N(0, 25)$ which means $(1 - \lambda)$ proportion of v_i , u_{ij} and e_{ijl} comes from the true distribution $N(0, 1)$ and the remaining λ proportion were generated from the arbitrary contaminated distribution $N(0, 25)$. We have chosen $\lambda = 0.05$.

Eight combinations of contamination are possible:

1. No contamination (0,0,0): e_{ijl} , u_{ij} and v_i were generated from standard normal.
2. Contamination in random errors (e,0,0): only 5% of e_{ijl} were generated from $N(0,25)$ and the remaining 95% were generated from $N(0,1)$.
3. Contamination in area-specific random effects (0,v,0): only 5% of u_{ij} were generated from $N(0,25)$ and the remaining 95% were generated from $N(0,1)$.
4. Contamination in cluster-specific random effect (0,0,u): only 5% of v_i were generated from $N(0,25)$ and the remaining 95% were generated from $N(0,1)$.
5. Contamination in both random errors and cluster-specific random effects (e,0,u): 5% of e_{ijl} and 5% of u_{ij} were generated from $N(0,25)$ and the remaining 95% of

e_{ijl} and u_{ij} were generated from $N(0,1)$.

6. Contamination in both random errors and area-specific random effects $(e,v,0)$: 5% of e_{ijl} and 5% of v_i were generated from $N(0,25)$ and the remaining 95% of e_{ijl} and v_i were generated from $N(0,1)$.
7. Contamination in both cluster-specific random effects and area-specific random effects $(0,v,u)$: 5% of u_{ij} and 5% of v_i were generated from $N(0,25)$ and the remaining 95% of u_{ij} and v_i were generated from $N(0,1)$.
8. Contamination in random errors, cluster-specific random effects and area-specific random effects (e,v,u) : 5% of e_{ijl} , 5% of u_{ij} and 5% of v_i were generated from $N(0,25)$ and the remaining 95% of e_{ijl} , u_{ij} and v_i were generated from $N(0,1)$.

We only consider the four models that involve e : $(0, 0, 0)$, $(e, 0, 0)$, $(e, v, 0)$, $(e, 0, u)$ and (e, v, u) . Tables 7 shows the biases and mean squared errors of the estimated model parameters using the RML method and the ML method.

In the uncontaminated case, ML appears to be slightly better than RML method in terms of biases. The RML method is almost as efficient as the ML method. When we consider outliers in the random errors, e_{ijl} , the ML estimator of variance component $\hat{\sigma}_e^2$ is affected significantly and its bias increases to around 1.20. However, the RML methods gives better results for the estimation of variance components. In the case of outliers in both e_{ijl} and v_i or both e_{ijl} and u_{ij} the RML method provides better results than the ML method. Also, when we have outliers in random errors, cluster effects and area-specific random effects, the RML estimators again produce more stable results than the ML ones. Generally, variance components associate with contaminated distribution are the most affected by outliers.

Table 7: Simulated biases and mean squared errors of the estimated model parameters- 5% contamination

	Parameter	ML		RML	
		Bias	MSE	Bias	MSE
No Contamination	β_0	0.012	0.037	0.015	0.038
	β_1	-0.004	0.001	-0.004	0.001
	σ_e^2	-0.002	0.003	0.057	0.007
	σ_v^2	-0.023	0.083	-0.058	0.095
	σ_u^2	0.029	0.021	0.014	0.026
Contamination in e	β_0	-0.006	0.038	-0.005	0.036
	β_1	0.003	0.003	0.001	0.002
	σ_e^2	1.183	1.483	0.362	0.141
	σ_v^2	-0.045	0.083	-0.085	0.086
	σ_u^2	-0.003	0.037	-0.024	0.303
Contamination in (e, v)	β_0	-0.002	0.069	-0.007	0.044
	β_1	0.003	0.003	0.001	0.002
	σ_e^2	1.184	1.483	0.362	0.158
	σ_v^2	1.31	2.756	0.214	0.157
	σ_u^2	0.003	0.037	-0.016	0.031
Contamination in (e, u)	β_0	-0.007	0.045	-0.004	0.039
	β_1	0.002	0.003	0.001	0.002
	σ_e^2	1.184	1.483	0.384	0.158
	σ_v^2	-0.065	0.124	-0.102	0.096
	σ_u^2	1.248	2.213	0.281	0.138
Contamination in (e, v, u)	β_0	-0.001	0.073	-0.006	0.046
	β_1	0.003	0.003	0.002	0.002
	σ_e^2	1.192	1.51	0.407	0.177
	σ_v^2	1.42	2.841	0.217	0.167
	σ_u^2	1.251	2.155	0.302	0.155

In the next step, we repeated the simulation runs with $\lambda = 0.1$. Table 8 presents the simulated biases and the mean squared errors of the estimated model parameters for the RML and the ML methods.

Table 8: Simulated biases and mean squared errors of the estimated model parameters- 10% contamination

	Parameter	ML		RML	
		Bias	MSE	Bias	MSE
No	β_0	-0.007	0.04	0.005	0.041
	β_1	0.004	0.001	0.004	0.001
	σ_e^2	-0.012	0.003	0.049	0.007
Contamination	σ_v^2	-0.013	0.078	-0.049	0.087
	σ_u^2	-0.012	0.251	-0.027	0.027
Contamination in e	β_0	-0.009	0.033	-0.016	0.028
	β_1	0.004	0.005	-0.001	0.003
	σ_e^2	2.386	5.87	0.722	0.542
	σ_v^2	-0.046	0.105	-0.096	0.097
	σ_u^2	-0.024	0.051	-0.060	0.033
Contamination in (e, v)	β_0	-0.016	0.116	-0.005	0.064
	β_1	0.006	0.005	0.005	0.003
	σ_e^2	2.426	6.075	0.803	0.669
	σ_v^2	2.410	9.976	0.608	0.581
	σ_u^2	-0.014	9.976	-0.047	0.041
Contamination in (e, u)	β_0	-0.015	0.065	-0.009	0.047
	β_1	0.005	0.006	0.0041	0.003
	σ_e^2	2.426	6.074	0.806	0.671
	σ_v^2	-0.046	0.198	-0.112	0.113
	σ_u^2	2.42	6.813	0.634	0.502
Contamination in (e, v, u)	β_0	-0.032	0.093	-0.032	0.051
	β_1	-0.007	0.005	-0.002	0.003
	σ_e^2	2.386	5.871	0.896	0.836
	σ_v^2	2.424	9.467	0.616	0.658
	σ_u^2	2.450	7.081	0.720	0.649

Considering Table 7 and Table 8, it appears that the bias of estimators increases as the proportion of outliers increases.

Chapter 4

Summary and Future Research

This chapter has two parts. In section 4.1, the result of the thesis are discussed, while in section 4.2, some suggestions are made for future research.

4.1 Summary

This thesis deals with the robust estimation of variance components in nested error regression models used in small area estimation. In Chapter 2, we studied some outlier-robust methods for the one-fold nested error model and primarily focused on the RML method proposed by Sinha and Rao (2009). Some practical issues of the RML method are highlighted. Particularly, it is shown that the Newton-Raphson method used for obtaining the estimates of variance components, is not stable and in some cases it does not even converge. Moreover, it is shown that the Newton-Raphson algorithm is very sensitive to the choice of the starting values. To avoid convergence problems with the Newton-Raphson method, we suggested using the fixed-point method that is more stable and also does not need the gradient matrix computation. In addition, Henderson's method III estimators of variance components are used for the initial values of the fixed-point algorithm. Some alternative approaches, such as choosing

another starting values or setting some constraints for the iterative values, are presented to remedy the Newton-Raphson's convergence problems. Then, an example is provided to compare the performance of the Newton-Raphson method and the fixed-point method in estimating θ . Moreover, we studied Fellner's (1986) and Huggins' (1993) methods and compared the behaviour of these estimators under contaminated distributions. The simulation results showed that the RML method (Sinha and Rao, 2009) provides better results than Fellner's and Huggins' methods, in terms of efficiency and bias.

In Chapter 3 of this thesis, the RML method is extended to the two-fold nested error regression model, which is appropriate for a two-stage cluster sampling design. The two-fold model is expressed as a general linear mixed model. The fixed-point method is used for solving the RML equations and Henderson's method III estimators are used as the initial values of the variance components. A simulation study is performed to investigate the behaviour of this model in the presence of outliers.

4.2 Future Work

As mentioned earlier, this thesis concentrates on the robust estimation of variance components and does not cover the problems associated with the estimation of small area means and consequently many related topics remain to be considered for future research. Some of these topics are discussed below.

First important area to be considered for future research is the robust estimation of robust small area means under the two-fold nested error regression model. Also, estimating the mean squared prediction error (MSPE) needs investigation.

A second area is motivated by Sinha and Rao (2009). It would be helpful to consider other models for outliers, such as long-tailed distribution and outliers in the auxiliary variables and investigate the behaviour of the robust methods in the

presence of such outliers.

Finally, considering the optimization method used for solving the RML equations, one may consider other iterative techniques such as the EM algorithm. Extensive literature on the use of EM algorithm for the ML and REML methods of estimation exists (see, for example (Searle et al., 1992)), which could be extended to the RML method.

Appendix A

A.1 ML Equations

For the general linear mixed model with block diagonal covariance structure under normality, we have

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{v}_i + \mathbf{e}_i; \quad i = 1, \dots, k,$$

with $\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i)$ with $E[\mathbf{y}_i] = \mathbf{X}_i\boldsymbol{\beta}$, $\text{var}(\mathbf{y}_i) = \mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{V}_i(\boldsymbol{\theta})$
 $\mathbf{e}_i \sim N(0, \mathbf{R}_i)$ and $\mathbf{v}_i \sim N(0, \mathbf{G}_i)$. The joint density function of \mathbf{y}_i is

$$f(\mathbf{y}_1, \dots, \mathbf{y}_k | \boldsymbol{\theta}) = \prod_{i=1}^k \left\{ \frac{1}{(2\pi)^{n_i/2} |\mathbf{V}_i|^{1/2}} e^{-1/2[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})]} \right\}.$$

Thus the log-likelihood is

$$l \propto \frac{-k}{2} \log |\mathbf{V}_i| - \frac{1}{2} \sum_{i=1}^k \{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\}.$$

Direct differentiation of l with respect to $\boldsymbol{\beta}$ gives

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = -1/2 \sum_{i=1}^k \frac{\partial [\sum_{i=1}^k \{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\}]}{\partial \boldsymbol{\beta}}.$$

Using the fact that $\partial(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T/\partial\boldsymbol{\beta} = -\mathbf{X}_i$ we can write

$$\frac{\partial l}{\partial\boldsymbol{\beta}} = \sum_{i=1}^k \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$$

Similarly, differentiation with respect to θ_l gives

$$\frac{\partial l}{\partial\theta_l} = \frac{\partial(-k/2 \log |\mathbf{V}_i|)}{\partial\theta_l} - 1/2 \sum_{i=1}^k \frac{\partial \{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\}}{\partial\theta_l} \quad (\text{A.1.1})$$

Break equation (A.1.1) into two parts, I and II

$$I = \frac{\partial(-k/2 \log |\mathbf{V}_i|)}{\partial\theta_l} = -\frac{k}{2} \text{tr} \left(\mathbf{V}_i \frac{\partial \mathbf{V}_i}{\partial\theta_l} \right).$$

Using $\partial(\det(X)) = \det(X) \text{tr}(V^{-1} \partial X)$, we get

$$\begin{aligned} II &= -1/2 \sum_{i=1}^k \frac{\partial \{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\}}{\partial\theta_l} \\ &= -1/2 \sum_{i=1}^k \left\{ \frac{\partial}{\partial\theta_l} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right\} \\ &= -1/2 \sum_{i=1}^k \left\{ (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial\theta_l} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right\}. \end{aligned}$$

Thus

$$\frac{\partial l}{\partial\theta_l} = \sum_{i=1}^k \left\{ (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial\theta_l} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - \text{tr} \left(\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial\theta_l} \right) \right\}.$$

A.2 Derivative of $\Phi(\boldsymbol{\theta})$

Using the Newton-Raphson algorithm for the iterative equation (2.1.7), we need to obtain the derivative of $\Phi(\boldsymbol{\theta}) = (\phi_1, \dots, \phi_q)^T$, where

$$\phi_l(\boldsymbol{\theta}) = \sum_{i=1}^k \left\{ \boldsymbol{\psi}^t(\mathbf{r}_i) \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_l} \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} \boldsymbol{\psi}(\mathbf{r}_i) - \text{tr}(\mathbf{K}_i \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_l}) \right\}; \quad l = 1, \dots, q. \quad (\text{A.2.1})$$

Equation (A.2.1) can be described in matrix form

$$\Phi(\boldsymbol{\theta}) = \boldsymbol{\Psi}^t(\mathbf{r}) \mathbf{U}^{1/2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{1/2} \boldsymbol{\Psi}(\mathbf{r}) - \text{tr}(\mathbf{K} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l}). \quad (\text{A.2.2})$$

Equation (A.2.2) can be separated in two parts, I and II , where

$$I = \boldsymbol{\Psi}^t(\mathbf{r}) \mathbf{U}^{1/2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{1/2} \boldsymbol{\Psi}(\mathbf{r}),$$

and

$$II = \text{tr}(\mathbf{K} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l}).$$

where

$$\begin{aligned} \frac{\partial I}{\partial \theta_l} &= \frac{\partial(\boldsymbol{\Psi}^t(\mathbf{r}) \mathbf{U}^{1/2} \mathbf{V}^{-1} \mathbf{U}^{1/2} \frac{\partial \mathbf{V}}{\partial \theta_l})}{\partial \theta_l} \\ &= 2 \frac{\partial}{\partial \theta_l} \left[\boldsymbol{\Psi}^t(\mathbf{r}) \mathbf{U}^{1/2} \mathbf{V}^{-1} \right] \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{1/2} \boldsymbol{\Psi}(\mathbf{r}) \\ &= 2 \left\{ \left(\frac{\partial}{\partial \theta_l} \boldsymbol{\Psi}^t(\mathbf{r}) \right) \mathbf{U}^{1/2} \mathbf{V}^{-1} + \boldsymbol{\Psi}^t(\mathbf{r}) \left(\frac{\partial}{\partial \theta_l} \mathbf{U}^{1/2} \right) \mathbf{V}^{-1} - \boldsymbol{\Psi}^t(\mathbf{r}) \mathbf{U}^{1/2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \right\} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{1/2} \boldsymbol{\Psi}(\mathbf{r}), \end{aligned}$$

where $\frac{\partial}{\partial \theta_i} \Psi^t(\mathbf{r}) = (-\frac{1}{2} \Psi'(\mathbf{r}) \mathbf{U}^{-1} \mathbf{r})^t$ and $\frac{\partial \mathbf{U}^{1/2}}{\partial \theta_i} = \frac{1}{2} \mathbf{U}^{-1/2} \mathbf{I}$.

Similarly,

$$\begin{aligned} \frac{\partial II}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i} \text{tr}(\mathbf{K} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i}) \\ &= \text{tr}(\frac{\partial}{\partial \theta_i} (\mathbf{K} \mathbf{V}^{-1} \frac{\partial \mathbf{V}^{-1}}{\partial \theta_i})) \\ &= -\text{tr}(\mathbf{K} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i}). \end{aligned}$$

Thus

$$\begin{aligned} \frac{\partial \Phi(\boldsymbol{\theta})}{\partial \theta_i} &= 2 \left\{ (-\frac{1}{2} \Psi'(\mathbf{r}) \mathbf{U}^{-1} \mathbf{r})^t \mathbf{U}^{1/2} \mathbf{V}^{-1} + \frac{1}{2} \Psi^t(\mathbf{r}) \mathbf{U}^{-1/2} \mathbf{V}^{-1} - \Psi^t(\mathbf{r}) \mathbf{U}^{1/2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{V}^{-1} \right\} \\ &\quad \times (\frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{V}^{-1} \mathbf{U}^{1/2} \Psi(\mathbf{r})) + \text{tr}(\mathbf{K} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i}). \end{aligned}$$

A.3 Estimating random effects

For estimating the area-specific random effects \mathbf{v}_i the following equation should be solved iteratively.

$$\mathbf{Z}_i^t \mathbf{R}_i^{-1/2} \boldsymbol{\psi} \left(\mathbf{R}_i^{-1/2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{v}_i) \right) - \mathbf{G}_i^{-1/2} \boldsymbol{\psi}(\mathbf{G}_i^{-1/2} \mathbf{v}_i) = 0; \quad i = 1, \dots, k \quad (\text{A.3.1})$$

(A.3.1) can be written in matrix form

$$\Upsilon(\mathbf{v}) = \mathbf{Z}^t \mathbf{R}^{-1/2} \boldsymbol{\psi} \left(\mathbf{R}^{-1/2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{v}) \right) - \mathbf{G}^{-1/2} \boldsymbol{\psi}(\mathbf{G}^{-1/2} \mathbf{v}) = 0.$$

Suppose,

$$\mathbf{P}_1 = \mathbf{Z}^t \mathbf{R}^{-1/2} \boldsymbol{\psi} \left(\mathbf{R}^{-1/2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{v}) \right),$$

and

$$\mathbf{P}_2 = \mathbf{G}^{-1/2} \boldsymbol{\psi}(\mathbf{G}^{-1/2} \mathbf{v}).$$

Using Newton-Raphson algorithm, the iterative equation can be described as follows:

$$\mathbf{v}^{(m+1)} = \mathbf{v}^{(m)} - \Upsilon'(\mathbf{v}^{(m)})^{-1} \Upsilon(\mathbf{v}^{(m)}),$$

Let $\mathbf{r} = \mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{v})$, then $\Upsilon'(\mathbf{v})$ can be written as:

$$\Upsilon'(\mathbf{v}) = \frac{\partial \mathbf{P}_1}{\partial \mathbf{v}} - \frac{\partial \mathbf{P}_2}{\partial \mathbf{v}}.$$

Clearly,

$$\begin{aligned} \frac{\partial \mathbf{P}_1}{\partial \mathbf{v}} &= \mathbf{Z}^t \mathbf{R}^{-1/2} \frac{\partial \psi(\cdot)}{\partial \mathbf{v}} \\ &= \mathbf{Z}^t \mathbf{R}^{-1/2} \frac{\partial \psi(\mathbf{r})}{\partial \mathbf{r}} \frac{\partial \mathbf{r}}{\partial \mathbf{v}}, \\ &= \mathbf{Z}^t \mathbf{R}^{-1/2} \mathbf{A} \{-\mathbf{R}^{-1/2} \mathbf{Z}\}. \end{aligned}$$

where \mathbf{A} is a diagonal matrix that its diagonal elements are $\psi'(r)$ and

$$\psi'(r) = \begin{cases} 1 & |r| < b \\ 0 & \text{otherwise} \end{cases}$$

Similarly,

$$\frac{\partial \mathbf{P}_2}{\partial \mathbf{v}} = \mathbf{G}^{-1/2} \mathbf{A}^* \mathbf{G}^{-1/2},$$

where \mathbf{A}^* is a diagonal matrix that its diagonal elements are $\psi'(\mathbf{G}^{-1/2} \mathbf{v})$.

A.4 Explicit form of \mathbf{V}_i^{-1} for the two-fold nested error regression model

Stukel (1991) provided an explicit form of \mathbf{V}_i^{-1} for the two-fold model. Considering the general model (3.1.2), the block of variance-covariance matrix is given by

$$\mathbf{V}_i = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 (\mathbf{1}_{n_i})(\mathbf{1}'_{n_i}) + \sigma_u^2 \text{diag}_{1 \leq j \leq m_i} (\mathbf{1}_{n_{ij}})(\mathbf{1}'_{n_{ij}}).$$

The above term is achieved by substituting the \mathbf{R}_i , \mathbf{G}_i and \mathbf{Z}_i in $\mathbf{V}_i = \mathbf{R}_i + \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i^T$. Thus, the block of the variance-covariance can be written as

$$\mathbf{V}_i = \mathbf{A}_i + \sigma_v^2 (\mathbf{1}_{n_i})(\mathbf{1}'_{n_i}),$$

where

$$\mathbf{A}_i = \text{diag}_{1 \leq j \leq m_i} (\sigma_e^2 \mathbf{I}_{n_{ij}} + \sigma_u^2 \mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}}) = \text{diag}_{1 \leq j \leq m_i} \mathbf{A}_{ij},$$

and $n_i = \sum_{j=1}^{m_i} n_{ij}$. Using the above expression and the following lemma (C. R. Rao (1965)), the inverse of \mathbf{V}_i is given by

$$\begin{aligned} \mathbf{V}_i^{-1} &= \sigma_e^{-2} \left\{ (\mathbf{I}_{n_i} - \text{diag}_{1 \leq j \leq m_i} \left(\frac{1}{n_{ij}} \gamma_{ij} (\mathbf{1}_{n_{ij}} \mathbf{1}'_{n_{ij}}) \right) \right. \\ &\quad \left. - \sigma_e^2 \sigma_u^{-2} \gamma_i (\sum_{j=1}^{m_i} \gamma_{ij})^{-1} [\text{col}_{1 \leq j \leq m_i} (1/n_{ij} \gamma_{ij} (\mathbf{1}_{n_{ij}}))] [\text{col}_{1 \leq j \leq m_i} (\frac{1}{n_{ij}} \gamma_{ij} (\mathbf{1}_{n_{ij}}))] \right\}, \end{aligned} \tag{A.4.1}$$

where

$$\begin{aligned} \gamma_{ij} &= \sigma_u^2 (\sigma_u^2 + \sigma_{j=1}^{m_i} / n_{ij})^{-1}, \\ \gamma_i &= \sigma_v^2 \left(\sigma_v^2 + \sigma_u^2 \left(\sum_{j=1}^{m_i} \gamma_{ij} \right)^{-1} \right)^{-1}. \end{aligned}$$

Lemma: Let \mathbf{C} be $k \times k$ non-singular matrix and \mathbf{a} and \mathbf{b} two $k \times 1$ vectors. Then the inverse of $\mathbf{C} + \mathbf{ab}'$ is given by $\mathbf{C}^{-1} - (\mathbf{1} + \mathbf{b}'\mathbf{C}^{-1}\mathbf{a})^{-1}\mathbf{C}^{-1}\mathbf{ab}'\mathbf{C}^{-1}$.

For \mathbf{V}_i , $\mathbf{C} = \mathbf{A}_i$, $\mathbf{a} = \sigma_v^2(\mathbf{1}_{n_i})$ and $\mathbf{b} = \mathbf{1}_{n_i}$. In the same way, for \mathbf{A}_{ij} , $\mathbf{C} = \sigma_e^2\mathbf{I}_{n_{ij}}$, $\mathbf{a} = \sigma_u^2\mathbf{1}_{n_{ij}}$ and $\mathbf{b} = \mathbf{1}_{n_{ij}}$.

Substituting equation (A.4.1) in (3.1.3), one can derive an explicit solution for $\tilde{\boldsymbol{\beta}}$.

See Stukel and Rao (1999).

Similarly, the BLUP estimator of $\tilde{\mathbf{v}}_i$ reduces to $\tilde{\mathbf{v}}_i = (\tilde{v}_i, \mathbf{v}_{2i})^T$ where

$$\tilde{v}_i = \gamma_i(\bar{y}_{i\gamma} - \bar{\mathbf{x}}_{i\gamma}^T \tilde{\boldsymbol{\beta}}),$$

$$\tilde{u}_{ij} = \gamma_{ij} \left(\sum_{l=1}^{n_{ij}} \frac{y_{ijl}}{n_{ij}} - \sum_{l=1}^{n_{ij}} \frac{\mathbf{x}_{ijl}}{n_{ij}} \tilde{\boldsymbol{\beta}} \right) - \gamma_i \gamma_{ij} (\bar{y}_{i\gamma} - \bar{\mathbf{x}}_{i\gamma}^T \tilde{\boldsymbol{\beta}}),$$

$$\bar{y}_{i\gamma} = \sum_{j=1}^{m_i} \gamma_{ij} \sum_{l=1}^{n_{ij}} \frac{y_{ijl}}{n_{ij}} / \sum_{j=1}^{m_i} \gamma_{ij},$$

$$\bar{\mathbf{x}}_{i\gamma} = \sum_{j=1}^{m_i} \gamma_{ij} \sum_{l=1}^{n_{ij}} \frac{\mathbf{x}_{ijl}}{n_{ij}} / \sum_{j=1}^{m_i} \gamma_{ij}.$$

Bibliography

- T. W. Anderson. Asymptotically efficient estimation of covariance matrices with linear covariance structure. *Annals of Statistics*, 1:135–141, 1973.
- W. R. Bell and E. T. Huang. Using the t-distribution to deal with outliers in small area estimation. *Proceedings of Statistics Canada Symposium*, 2006.
- D. Bernoulli. Dijudicatio maxime probabilis plurium observationum discrepantium ateq versimillima inductio inde formanda (english translated by c. g. allen 1961 in biometrika). *petropolit*, 1:3–33, 1777.
- R. L. Chambers. Outlier robust finite population estimation. *Journal of American Statistical Association*, 81:1063–1069, 1986.
- W. Chauvenet. *Method of Least Squares*. Dover, 1863.
- G. S. Datta and P. Lahiri. Robust hierachical bayes estimation of small area characteristics in the presence of covariates and outliers. *Journal of Multivariate Analysis*, 54:310–328, 1995.
- R. E. Fay and R. A. Herriot. Estimation of income for small places: An application of james-stein procedures to census data. *Journal of American Statistical Association*, 74:269–277, 1979.
- W. H. Fellner. Robust estimation of variance components. *Technometrics*, 28:51–60, 1986.

- R. Fried, I. Molina, B. Perez, and A. Thieler. Robustness analysis of unbalanced linear mixed modeling. In *ERCIM*, 2011.
- W. A. Fuller and G. E. Battese. Transformations for estimation of linear models with nested-error structure. *Journal of American Statistical Association*, 68:626–632, 1973.
- J. E. Gentle. *Numerical Linear Algebra for Applications in Statistics*. Springer, 1998.
- F. R. Hampel. *Contributions to the theory of robust estimation*. PhD thesis, University of California, Berkeley, 1968.
- F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics*. John Wiley & Sons, 1986.
- D. A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72:320–338, 1977.
- C. R. Henderson. Estimation of variance and covariance components. *Biometrics*, 9: 226–252, 1953.
- C. R. Henderson. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31:423–447, 1975.
- P. J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:73–101, 1964.
- R. M. Huggins. On the robust analysis of variance components models for pedigree data. *The Australian Journal of Statistics*, 35:43–57, 1993.
- S. G. Kellison. *Fundamentals of Numerical Analysis*. R. D. Irwin, 1975.

- N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.
- R. C. Littell, W. Stroup, and R. J. Freund. *SAS for Linear Models*. SAS Institute, 2002.
- R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, 2002.
- C. E. McCulloch and S. R. Searle. *Generalized, Linear, and Mixed Models*. John Wiley & Sons, 2001.
- E. S. Pearson and C. Chandra. The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, 28:308–320, 1936.
- B. Peirce. Criterion for the rejection of doubtful observations. *Astronomical Journal*, 2:161–163, 1852.
- J. N. K. Rao. *Small Area Estimation*. JOHN WILEY & SONS, INC, 2003.
- A. M. Richardson and A. H. Welsh. Robust restricted estimation in the mixed linear model. *Biometrics*, 51:1429–1439, 1995.
- S.R. Searle, G. Casella, and Charles E. McCulloch. *Variance Components*. Wiley-Interscience, 1992.
- S. K. Sinha and J. N. K. Rao. Robust small area estimation. *The Canadian Journal of Statistics*, 37:381–399, 2009.
- W. A. Stahel and A. Welsh. Approaches to robust estimation in the simplest variance components model. *Journal of Statistical Planning and Inference*, 57:295–319, 1997.
- E. J. Stone. On the rejection of discordant observations. *Roy. Astr. Soc*, 34:9–15, 1873.

- D. M. Stukel. *Small Area Estimation Under One and Two-fold Nested Error Regression Models*. PhD thesis, Carleton University, Mathematics and Statistics, 1991.
- D. M. Stukel and J. N. K. Rao. On small-area estimation under two-fold nested error regression models. *Journal of Statistical Planning and Inference*, 78:131–147, 1999.
- W. R. Thompson. On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Annal of Mathematical Statistics*, 6:214–219, 1935.
- J. W. Tukey. *A survey of sampling from contaminated distributions*. Stanford University Press, 1960.
- R. Valliant, A. H. Dorfman, and R. M. Royall. *Finite Population Sampling and Inference: A prediction Approach*. John Wiley & Sons, 2000.
- A. H. Welsh and E. Ronchetti. Bias-calibrated estimation from sample surveys containing outliers. *Journal of the Royal Statistical Society*, 60:413–428, 1998.
- B. T. West, K. B. Welch, and A. T. Galecki. *Linear Mixed Models A Practical Guide Using Statistical Software*. Chapman & Hall/CRC, 2007.