NOTICE

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

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

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

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

AVIS

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

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

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

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.
NESTED-ERROR REGRESSION MODELS AND SMALL AREA ESTIMATION COMBINING CROSS-SECTIONAL AND TIME SERIES DATA

by

Mingyu Yu

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

Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University
Ottawa, Canada
November, 1993

©Copyright 1993
Mingyu Yu
The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

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

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

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

ISBN 0-315-89723-6
### The Humanities and Social Sciences

**Communications and the Arts**
- Architecture: 0739
- Art History: 0700
- Cinema: 0770
- Dance: 0738
- Fine Arts: 0737
- Information Science: 0733
- Journalism: 0791
- Library Science: 0798
- Mass Communications: 0730
- Music: 0743
- Speech Communication: 0745
- Theater: 0755

**Education**
- General: 0515
- Administration: 0514
- Adult and Continuing: 0513
- Agricultural: 0517
- Art: 0273
- Bilingual and Multicultural: 0282
- Business: 0268
- Community College: 0275
- Curriculum and Instruction: 0277
- Early Childhood: 0518
- Elementary: 0524
- Finance: 0277
- Guidance and Counseling: 0219
- Health: 0680
- Higher: 0745
- History: 0520
- Home Economics: 0278
- Industrial: 0521
- Language and Literature: 0279
- Mathematics: 0280
- Music: 0222
- Philosophy of: 0998
- Physical: 0523

**Psychology**
- General: 0679
- Language: 0679
- Literature: 0679
- Linguistics: 0679
- Modern: 0679
- Ancient: 0679
- Comparative: 0679
- Medieval: 0679
- African: 0679
- American: 0679
- Asian: 0679
- Canadian (English): 0679
- Canadian (French): 0679
- English: 0679
- German: 0679
- Latin American: 0679
- Middle Eastern: 0679
- Romance: 0679
- Slavic and East European: 0679

**Philosophy, Religion and Theology**
- General: 0680
- Biblical Studies: 0680
- Clergy: 0680
- History: 0680
- Philosophy of: 0680
- Religion: 0680

**Social Sciences**
- American Studies: 0323
- Anthropology: 0324
- Archaeology: 0324
- Cultural: 0324
- Economics: 0324
- Environmental: 0324
- Family: 0324
- General: 0324
- International Law: 0324
- International Relations: 0324
- Political: 0324
- Psychology: 0324
- Social: 0324
- Sociology: 0324

**The Sciences and Engineering**

**Biological Sciences**
- Agriculture: 0473
- Agronomy: 0473
- Animal Culture and Nutrition: 0473
- Animal Pathology: 0473
- Food Science and Technology: 0473
- Forestry and Wildlife: 0473
- Plant Culture: 0473
- Plant Pathology: 0473
- Plant Physiology: 0473
- Range Management: 0473
- Wood Technology: 0473

**Biology**
- General: 0306
- Anatomy: 0306
- Biochemistry: 0306
- Botany: 0306
- Cell: 0306
- Ecology: 0306
- Embryology: 0306
- Genetics: 0306
- Immunology: 0306
- Microbiology: 0306
- Molecular: 0306
- Neuroscience: 0306
- Oceanography: 0306
- Physiology: 0306
- Radiation: 0306
- Veterinary Science: 0306

**Geology**
- General: 0370
- Geology: 0370
- Geophysics: 0370
- Hydrology: 0370
- Mineralogy: 0370
- Paleobotany: 0370
- Paleocology: 0370
- Paleontology: 0370
- Paleophysiology: 0370
- Physical Geology: 0370
- Physical Oceanography: 0370

**Health and Environmental Sciences**
- Environmental Sciences: 0768
- Health Sciences: 0768
- General: 0768
- Audiology: 0768
- Chemotherapy: 0768
- Dentistry: 0768
- Education: 0768
- Hospital Management: 0768
- Human Development: 0768
- Immunology: 0768
- Medical and Surgery: 0768
- Mental Health: 0768
- Nursing: 0768
- Nutrition: 0768
- Obstetrics and Gynecology: 0768
- Occupational Health and Safety: 0768
- Ophthalmology: 0768
- Pathology: 0768
- Pharmacology: 0768
- Physical Therapy: 0768
- Public Health: 0768
- Radiology: 0768
- Speech Pathology: 0768
- Toxicology: 0768
- Home Economics: 0768

**Physical Sciences**
- Pure Sciences: General: 0485
- Analytical: 0485
- Biochemistry: 0485
- Inorganic: 0485
- Nuclear: 0485
- Organic: 0485
- Pharmacological: 0485
- Physical: 0485
- Physics: 0485
- Mathematics: 0485
- General: 0485
- Acoustics: 0485
- Astronomy and Astrophysics: 0485
- Atmospheric Science: 0485
- Oceanography: 0485
- Electronics and Electrical Engineering: 0485
- Electrical and Computer Engineering: 0485
- Material Science: 0485
- Mechanical: 0485
- Mining: 0485
- Nuclear: 0485
- Petroleum: 0485
- Sanitary and Municipal Engineering: 0485
- System Science: 0485

**Psychology**
- General: 0621
- Behavioral: 0621
- Cognitive: 0621
- Developmental: 0621
- Experimental: 0621
- Industrial: 0621
- Personality: 0621
- Physiological: 0621
- Psychodynamic: 0621
- Social: 0621
- Statistics: 0621

**Statistics**
- General: 0683
- Actuarial: 0683
- Agricultural: 0683
- Biological: 0683
- Business: 0683
- Educational: 0683
- Engineering: 0683
- Experimental: 0683
- Industrial: 0683
- Psychological: 0683
- Social: 0683

**Technology**
- General: 0794
- Aerospace: 0794
- Aeronautical: 0794
- Automotive: 0794
- Biomedical: 0794
- Chemical: 0794
- Civil: 0794
- Electronic and Electrical Engineering: 0794
- Electrical and Computer Engineering: 0794
- Material Science: 0794
- Mechanical: 0794
- Mining: 0794
- Nuclear: 0794
- Petroleum: 0794
- Sanitary and Municipal Engineering: 0794
- System Science: 0794

- Geology: 0428
- Operations Research: 0428
- Plastics Technology: 0428
- Textile Technology: 0428

- General: 0357
- Aerospace: 0357
- Aeronautical: 0357
- Automotive: 0357
- Biomedical: 0357
- Chemical: 0357
- Civil: 0357
- Electronic and Electrical Engineering: 0357
- Electrical and Computer Engineering: 0357
- Material Science: 0357
- Mechanical: 0357
- Mining: 0357
- Nuclear: 0357
- Petroleum: 0357
- Sanitary and Municipal Engineering: 0357
- System Science: 0357
The undersigned hereby recommend to
The Faculty of Graduate Studies and Research
acceptance of the thesis.

NESTED-ERROR REGRESSION MODELS AREA ESTIMATION
COMBINING CROSS-SECTIONAL AND TIME SERIES DATA
submitted by
Mingyu Yu
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

[Signature]
Chairman, Department of Mathematics and Statistics

[Signature]
Thesis Supervisor

[Signature]
External Examiner

Carleton University
January, 1994
Abstract

This dissertation consists of two parts. In Part I, the one and two-fold nested-error regression models with unequal error variances are investigated. Approximate minimum norm quadratic unbiased (AMINQU) estimators for the variance components and error variances are derived by ignoring lower order terms in the minimum norm quadratic unbiased (MINQU) estimation equations (C. R. Rao and Kleffe, 1988). The AMINQU estimators of variance components are shown to be asymptotically normal. Consistent variance estimators are obtained using the substitution method. Asymptotically correct confidence intervals on the variance components are constructed using these variance estimators.

Truncated AMINQU estimators provide positive estimators of weights for weighted least squares (WLS) estimation of regression parameters. Asymptotic normality of the resulting WLS estimators is established, and a consistent estimator of the asymptotic covariance matrix is also obtained. An alternative WLS estimator of regression parameters, obtained by constructing weights from AMINQU estimators of variance components and within group sample variances, is also studied.

For the one-fold nested-error model, the delete-group jackknife method is applied to the AMINQU estimators to reduce the bias. It is also used to estimate the variance of the AMINQU estimators as well as the covariance matrix of the WLS estimators of regression parameters. Consistency of the jackknife variance and covariance estimators is established and asymptotically correct jackknife confidence intervals on the parameters are obtained. Results of a limited simulation study and the proposed methods are presented.

The AMINQU estimation method is also applied to two-fold nested-error re-
gression models, but explicit expressions could be obtained only for a special set of prior values of variance components. These AMINQU estimators of variance components, along with the within group sample variances, are used to construct weights for WLS estimator of regression parameters. Asymptotic normality of the induced WLS regression estimator is established. A consistent estimator of the asymptotic covariance matrix of the WLS estimator is also obtained.

In Part II of this dissertation, a model involving random effects and autocorrelated errors is proposed for small area estimation, using both time series and cross-sectional data. The sampling errors are assumed to have a known block diagonal covariance matrix. This model is an extension of a well-known model, due to Fay and Herriot (1979), for cross-sectional data. A two-stage estimator of a small area mean for the current period is obtained under the proposed model with known autocorrelation, by first deriving the best linear unbiased prediction (BLUP) estimator assuming known variance components, and then replacing them with their consistent estimators. Extending the approach of Prasad and Rao (1986, 1990) for the Fay–Herriot model, an estimator of mean squared error (MSE) of the two-stage estimator, correct to a second order approximation for a small or moderate number of time points and a relatively large number of small areas, is obtained. The case of unknown autocorrelation is also discussed. A consistent method of estimating the autocorrelation, though not very adequate in small or moderate samples, is suggested. Two other alternatives are also proposed. Two-stage approximations to MSE and estimators of MSE are also obtained. A simulation study on the efficiency of the proposed two-stage estimators and the accuracy of the proposed estimators of MSE are presented.

A hierarchical Bayes (HB) method is also applied to our model. Small area means are estimated by their posterior means, and their posterior variances are used as measures of uncertainty associated with the estimators. Evaluation of posterior means and posterior variances through numerical integration or Gibbs sampling is outlined.
Acknowledgements

First of all, I would like to express my sincere thanks to my thesis supervisor, Dr. J. N. K. Rao for the knowledge and skills he has taught me, as well as for his kindness and support during the past few years. Without his instruction and encouragement, this thesis may not have come into existence.

My thanks also go to Dr. J. Shao from whom I benefitted a lot, especially in the area of large sample theory.

I appreciate the opportunity given to me by the Department of Mathematics and Statistics to study at Carleton University. In particular, I would like to extend my thanks to Dr. M. Csörgő whose support helped to make it possible for me to come to Carleton University.

My sincere thanks are also extended to Drs. D.R. Krewski, B. MacGibbon, J. Oommen and R. Sitter for the time and hard work they put in for my thesis examination.

Finally, I want to thank my wife, Sophie Qixin Chen, and my parents for their love and support.
## Contents

Acceptance Sheet ................................................................. ii  
Abstract .................................................................................. iii  
Acknowledgements ...................................................................... v  
Contents .................................................................................... vi  
List of Tables ................................................................................ x  

1 Introduction ................................................................................ 1  
1.1 Introduction to One-Fold Nested Error Linear Regression Models .. 1  
  1.1.1 Classical linear regression model ........................................... 3  
  1.1.2 Linear regression models with unequal error variances .......... 4  
  1.1.3 Nested error regression models ............................................ 8  
  1.1.4 Common mean case ............................................................. 9  
  1.1.5 Random intercept model ..................................................... 10  
1.2 Introduction to Small Area Estimation ..................................... 11  
  1.2.1 Cross-sectional models ....................................................... 12  
    Regression synthetic estimators ................................................. 12  
    Two-stage estimation ............................................................. 13  
    Empirical Bayes Estimation ...................................................... 15  
    Hierarchical Bayes estimation ............................................... 16  
  1.2.2 Cross-sectional and time series models ................................ 17  
    Previously proposed models .................................................... 18  
    Proposed model .................................................................... 20  
1.3 Thesis Outline ........................................................................... 20
1.3.1 Outline for Part I ................................................. 20
1.3.2 Outline for Part II ................................................. 23

I Inference for One and Two-Fold Nested-Error Regression Models 25

2 Estimating Variance Components 26
   2.1 WLS Estimators .................................................. 26
   2.2 MINQUE Procedure with a Priori Values ...................... 29
   2.3 ASR Estimator .................................................. 31
   2.4 AMIQUE Procedure .............................................. 31
   2.5 Asymptotic Properties of $\sigma^2_v$ .......................... 35
      2.5.1 The model assumptions and preliminary results .......... 36
      2.5.2 The asymptotic normality of AMINQUE .................. 40
      2.5.3 Estimation of asymptotic variance ....................... 44

3 Estimating Regression Coefficients 46
   3.1 Notations and Some Preliminary Results ...................... 47
   3.2 Asymptotic Distribution of $\hat{\beta}_w$ ...................... 50
   3.3 Substitution Estimator of $V_k$ .............................. 54
   3.4 Another Estimator of $\beta$ .................................. 56

4 Jackknife Estimators 60
   4.1 Review of Jackknife Methodology ............................. 61
   4.2 Jackknife Estimators and Variance Estimators for $\sigma^2_v$ .... 63
   4.3 A Jackknife Variance Estimator for $\hat{\beta}_w$ .............. 73
   4.4 A Jackknife variance estimator for $\hat{\beta}_w$ .............. 79
   4.5 Simulation Results .......................................... 80
      4.5.1 Description of the design ................................. 80
6.5.1 Autocorrelation known ........................................ 126
6.5.2 Autocorrelation unknown ...................................... 127

7 Hierarchical Bayes Analysis ........................................ 134
  7.1 HB Estimator: $\rho$ Known ..................................... 134
    7.1.1 Numerical integration .................................... 135
    7.1.2 Gibbs sampling .......................................... 137
  7.2 HB Estimation: $\rho$ Unknown .................................. 139

8 Summary and Proposed Future Research ......................... 142
  8.1 Summary for Part I ........................................... 142
  8.2 Summary for Part II .......................................... 144
  8.3 Proposed Further Research .................................. 145

Bibliography .......................................................... 147
List of Tables

4.1 Relative Biases (%) for Estimators of $\sigma_v^2$ ........................................ 86
4.2 Coverage Probability (CP %) and Length of Confidence Interval (LCI) for Estimators of $\sigma_v^2$ at Level $1 - \alpha = 90\%$, when Errors Follow Normal Distribution ........................................ 87
4.3 Coverage Probability (CP %) and Length of Confidence Interval (LCI) for estimators of $\sigma_v^2$ at level $1 - \alpha = 90\%$, when Errors follow Double Exponential Distribution ........................................ 88
4.4 Percentage Efficiency of $\hat{\beta}_w$ Relative to $\hat{\beta}_{w*}$ and $\hat{\beta}_{w*}$ ........................................ 89
4.5 Coverage Probability (CP %) of Confidence Intervals on $\beta$ for Four Variance Estimators at level $1 - \alpha = 90\%$ ........................................ 90
4.6 Average Lengths of Confidence Intervals (LCI) for Four Variance Estimators at Level $1 - \alpha = 90\%$ ........................................ 91
4.7 Relative Biases (RB %) OF Four Variance Estimators ........................................ 92
4.8 Coefficient of Variation (CV %) of Four Variance Estimators ........................................ 93

6.1 Gain in efficiency (%) of the two-stage estimator $\hat{\theta}_{1T}(\rho)$ over the Fay-Herriot estimator: $\rho = 0.2$ ........................................ 129
6.2 Gain in efficiency (%) of the two-stage estimator $\hat{\theta}_{1T}(\rho)$ over the Fay-Herriot estimator: $\rho = 0.4$ ........................................ 129
6.3 Relative bias (%) of $\text{mse}[\hat{\theta}_{1T}(\rho)]$ and $\text{mse}_N[\hat{\theta}_{1T}(\rho)]$ (in brackets): $\rho = 0.2$ ........................................ 130
6.4 Relative bias (%) of $\text{mse}[\hat{\theta}_{1T}(\rho)]$ and $\text{mse}_N[\hat{\theta}_{1T}(\rho)]$ (in brackets): $\rho = 0.4$ ........................................ 130
6.5 Relative bias (%) of $\text{mse}[\hat{\theta}_{1T}(\rho_0)]$: $\rho_0 = 0.1, 0.3, 0.5, 0.6, \ T = 5$ ........ 131

6.6 Relative bias (%) of $\text{mse}[\hat{\theta}_{1T}(\rho_0)]$: $\rho_0 = 0.2, 0.3, 0.5, 0.6, \rho = 0.4, \ T = 5$ .................................................. 132

6.7 Relative bias (%) of $\text{mse}[\hat{\theta}_{1T}(\hat{\rho}_N)]$: $\rho = 0.2$ .................. 133

6.8 Relative bias (%) of $\text{mse}[\hat{\theta}_{1T}(\hat{\rho}_N)]$: $\rho = 0.4$ .................. 133
Chapter 1

Introduction

This dissertation consists of two parts. In Part I, we study one and two-fold nested-error linear regression models with unequal error variances. Results on this topic are presented in Chapters 2–5. In Part II, we study small area estimation by combining time series and cross-sectional data. Our results on this topic are presented in Chapters 6 and 7. These two parts are somewhat unrelated, but both parts are concerned with models involving random effects.

In this chapter, we present some background and known methodologies concerning these two parts. We review some nonparametric methods for one-fold nested-error linear regression models in Section 1.1. In Section 1.2, we present some background for small area estimation. In Section 1.3, we give an outline of the work reported in the following Chapters.

1.1 Introduction to One-Fold Nested Error Linear Regression Models

The following one-fold nested error linear model with unequal error variances is considered in Chapters 2–4:

$$y_{ij} = x_i' \beta + \varepsilon_{ij} \quad \text{with} \quad \varepsilon_{ij} = v_i + e_{ij}, \quad (1.1)$$
CHAPTER 1. INTRODUCTION

\[ i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \quad \sum_{i=1}^{k} n_i = n, \]

where \( y_{ij} \) is the \( j \)-th observation within \( i \)-th group, \( x_i = (x_{i1}, \ldots, x_{ip})' \) is a vector of known constants (design points or auxiliary variable associated with the \( i \)-th group), \( \beta = (\beta_1, \ldots, \beta_p)' \) is a vector of unknown parameters called regression parameters, \( \nu_i \)'s are independent identically distributed variables with mean 0 and \( \sigma_\nu^2 \) (denoted as \( \nu_i \sim N(0, \sigma_\nu^2) \)), also for given \( i \), \( e_{ij} \sim N(0, \sigma_e^2) \), and \( \nu_i \)'s and \( e_{ij} \)'s are mutually independent. Note that \( \sigma_\nu^2 \)'s and \( n_i \)'s are not necessarily equal. If \( n_i \)'s are equal, the model is called a balanced design model. If \( \sigma_\nu^2 \)'s are equal, the model is called a nested-error linear model with equal error variances. The main interest of this dissertation is the estimation of \( \sigma_\nu^2, \sigma_e^2 \) and \( \beta \) and the related statistical inferences.

Model (1.1) has gained extensive application in agriculture, biology, medicine, and industry. For illustrations of the model, we present the following examples which are special cases of model (1.1). With \( x_i \beta = \mu \), the model can be applied to estimate the average difference between two treatment means over a series of randomized experiments conducted at a set of different locations or at different times. With \( x_i \beta = \alpha + \beta x_i \), \( y_{ij} \) could represent the difference in the responses of two specific treatments in the \( j \)-th block of the \( i \)-th experiment, with \( x_i \) representing the duration of \( i \)-th experiment. More examples can be found in Henderson (1975), P. S. R. S. Rao, Kaplan and Cochran (1981), and P. S. R. S. Rao and Kuranchie (1988), among others.

Model (1.1) includes many special cases, which have received considerable attention in the literature. Some of these papers date back to the 1930s, e.g. Cochran (1937) and Yates and Cochran (1938), while many others are of recent origin. When \( \sigma_\nu^2 \) and \( \sigma_e^2 \) are restricted or when \( x_i \beta \) takes special forms, model (1.1) reduces to the following simpler models.

Case 1. \( \sigma_\nu^2 = 0, \sigma_e^2 = \sigma_e^2 \). In this case model (1.1) is the classic linear regression model.

Case 2. \( \sigma_\nu^2 = 0, \sigma_e^2 \) not all equal. This model is usually called a linear regression model with unequal error variances. This model has received much attention in the last twenty years (see e.g. Rao, 1973; Fuller and Rao, 1978; Carroll and Ruppert
CHAPTER 1. INTRODUCTION


Case 3. $\sigma_i^2 \neq 0, \sigma_i^2 = \sigma^2$. This model is called a nested error regression model (see Fuller and Battese, 1973). The special case, $x_i'\beta = \mu$, yields the classical one-fold (or one-way) random effects model (see Searle, 1971).

Case 4. $x_i'\beta = \mu$ or $x_i'\beta = \alpha + \beta x_i$. P. S. R. S. Rao, Kaplan and Cochran (1981) considered the first case, or the common mean case, while P. S. R. S. Rao and Kuranchie (1988) discussed the second case of a single auxiliary variable $x_i$.

In the following subsections, some well-known results on these special cases are reviewed.

1.1.1 Classical linear regression model

The classic linear regression model is of form

$$y_i = x_i'\beta + e_i, \quad i = 1, 2, \ldots, k$$

(1.2)

where $\beta$ is a $p \times 1$ vector of unknown parameters, $e_i \sim N(0, \sigma^2)$. This is a special case of model (1.1) with $\sigma_i^2 = 0, \sigma_i^2 = \sigma^2$ and $n_i \equiv 1$. Let $Y = (y_1, \ldots, y_k)'$, $X = (x_1, \ldots, x_k)'$, and $e = (e_1, \ldots, e_k)'$, then model (1.2) can be written as

$$Y = X\beta + e.$$ 

Given $X$ is of full column rank, the regression parameter $\beta$ is estimated through the ordinary least squares (OLS) procedure which gives

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y.$$ 

The error variance $\sigma^2$ is estimated by the sum of squares of residuals divided by $k - p$, i.e.,

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta}_{OLS})(Y - X\hat{\beta}_{OLS})}{k - p} = \frac{Y'(I - X(X'X)^{-1}X')Y}{k - p}.$$ 

The detailed inference for this model can be found in standard text books (see e.g. Seber, 1977; Searle, 1981; Cook and Weisberg, 1982). Further discussion is therefore omitted.
1.1.2 Linear regression models with unequal error variances

When $\sigma_i^2 = 0$ and $\sigma_i^2$ are not all equal, model (1.1) becomes

$$y_{ij} = x_i'\beta + e_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \quad \sum_{i=1}^k n_i = n. \quad (1.3)$$

In the literature this model (1.3) is also called a heteroscedastic linear regression model.

There are, in general, three types of approaches for making inferences on $\beta$ from this model:

(A) Parametric approach. In this approach, $\sigma_i^2$ is assumed to be a specified function $h$ of the design points $x_i$, the regression coefficient $\beta$ and some auxiliary parameter $\theta$ (a $q \times 1$ vector of unknown parameters), or $\sigma_i^2$ can be expressed as

$$\sigma_i^2 = h(x_i, \beta, \theta), \quad (1.4)$$

see e.g., Carroll (1982), Davidian and Carroll (1987), Carroll and Ruppert (1988). The inference on $\beta$ involves starting with the initial values of $\beta$ and $\theta$, and then estimating $\sigma_i^2$ using the above function. Then $\beta$ is updated by the weighted least squares (WLS) estimator of $\beta$ from model (1.3) using the estimated $\sigma_i^2$. The method also specifies an updating method for $\theta$. With the updated $\beta$ and $\theta$, $\sigma_i^2$ is updated from (1.4), and the iteration is continued until estimators of $\sigma_i^2$ and $\beta$ converge.

The advantages of this approach are: (1) The estimator of $\sigma_i^2$ is consistent, and, (2) The estimator of $\beta$ is almost as efficient as the WLS estimator with known $\sigma_i^2$. The disadvantages of the scheme are: (1) The specification of the function $h$ is usually artificial, and, (2) The iteration procedure may not converge.

(B) Nonparametric approach. This approaches generally lacks the advantages of the former parametric approaches. However, in some cases, it is unreasonable to assume a relationship between the error variance $\sigma_i^2$ and the design points $x_i$ since there are occasions when we don’t have a priori information about the structure of $\sigma_i^2$. In this case a nonparametric approach is needed.
CHAPTER 1. INTRODUCTION

(C) Bayesian approach. A newer approach to the heteroscedastic linear model is the Bayesian approach. This approach assumes a priori distribution on $\sigma^2_i$ and the regression coefficients $\beta$. Then the inference is based on the posterior distributions of these parameters $\sigma^2_i$ and $\beta$ given the observations $\{y_i\}$ (See Shao, 1991).

Since the research presented in this dissertation is related to the nonparametric approach, we omit further discussions of approaches (A) and (C) and concentrate on a discussion of approach (B).

The matrix representation of model (1.3) can be written as

$$\mathcal{Y} = \mathcal{X}\beta + e,$$  

where

$$\mathcal{Y}_{n \times 1} = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{k1}, \ldots, y_{kn_k})',$$  

$$\mathcal{X}_{n \times p} = (x_{11}1_{n_1}', x_{21}1_{n_2}', \ldots, x_{k1}1_{n_k}'),$$  

$$e_{n \times 1} = (e_{11}, \ldots, e_{1n_1}, \ldots, e_{kn_1}, \ldots, e_{kn_k})',$$

with $1_t$ as a $t$-vector with all entries equal to one. The dispersion matrix of the model is given by

$$V = \text{Var}(e) = \text{block diag}(\sigma^2_i I_{n_1}, \ldots, \sigma^2_i I_{n_k}).$$  

We assume that $\mathcal{X}$ is of full column rank. The main interests of investigation are estimation of $\sigma^2_i$ and the estimation of $\beta$. The general steps of estimation are as follows.

Step 1: An estimate $\hat{\sigma}^2_i$ of $\sigma^2_i$ is either guessed (some prior values of $\sigma^2_i$), or is estimated using some procedure.

Step 2: With the estimator $\hat{\sigma}^2_i$ from step 1, $\beta$ is estimated using weighted least squares (WLS) procedure:

$$\hat{\beta}_w = (\mathcal{X}'\hat{V}^{-1}\mathcal{X})^{-1}\mathcal{X}'\hat{V}^{-1}\mathcal{Y}.$$  

where $\hat{V}$ is obtained by substituting $\hat{\sigma}^2_i$ into $\sigma^2_i$ in (1.6).
When \( n_i \geq 2 \) for all \( i \), an intuitive estimator of \( \sigma_i^2 \) is the sample variance in the \( i \)-th group

\[
s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \quad \text{for } n_i \geq 2,
\]

where \( \bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i \), the average within \( i \)-th group. This estimator, also referred to as the analysis of variance (ANOVA) estimator, is unbiased for \( \sigma_i^2 \). But for small or moderate \( n_i \), its variance is quite large. The estimator \( s_i^2 \) is, of course, not applicable if \( n_i \) equals 1. Meier (1953), Cochran and Carroll (1953) and Levy (1970) considered using \( s_i^2 \) in model (1.3) with \( x_i' \beta = \mu \). Jacquez et al (1968), William (1967), Bement and William (1969) used \( s_i^2 \) for model (1.3) with \( x_i' \beta = \alpha + \beta x_i \). Their extensive empirical studies showed that the induced WLS estimator of \( \beta \) from (1.7) by using \( \hat{\sigma}_i^2 = s_i^2 \) is not efficient for small \( n_i \).

C. R. Rao (1970) introduced minimum norm quadratic unbiased (MINQU) estimation for estimating \( \sigma_i^2 \), motivated by the work of Hartley, Rao, and Kiefer (1969) in the context of survey sampling. We now describe this method in detail.

For estimating \( p' \sigma^2 = \sum_k p_k \sigma_k^2 \), with \( \sigma^2 = (\sigma_1^2, \ldots, \sigma_k^2)' \) and \( p = (p_1, \ldots, p_k)' \), a quadratic form \( \mathcal{Y}' A \mathcal{Y} \) with a symmetric matrix \( A \) is used. Also \( \mathcal{Y}' A \mathcal{Y} \) is required to have the following two properties:

(i) It is invariant if \( \mathcal{Y} \) is replaced by \( \mathcal{Y} - \mathcal{X} \beta_0 \) for any fixed \( \beta_0 \), i.e. \( \mathcal{Y}' A \mathcal{Y} = (\mathcal{Y} - \mathcal{X} \beta_0)' A (\mathcal{Y} - \mathcal{X} \beta_0) \) which is equivalent to

\[
A \mathcal{X} = 0
\]

(1.8)

and so

\[
\mathcal{Y}' A \mathcal{Y} = e' A e.
\]

(ii) It is unbiased for \( p' \sigma^2 \) or \( E(\mathcal{Y}' A \mathcal{Y}) = tr(A V) = \sum_i p_i \sigma_i^2 \) which is equivalent to

\[
tr(A F_i) = p_i, \quad i = 1, \ldots, k,
\]

(1.9)

where \( F_i = \text{block diag}(0, \ldots, 0, I_{n_i}, 0, \ldots, 0) \) and \( tr(A) \) denote the trace of a matrix \( A \).
CHAPTER 1. INTRODUCTION

When $\beta$ is known, $e$ is also known. Then a reasonable estimator of $p^T \sigma^2$ is

$$\sum_{i=1}^{k} p_i n_i^{-1} \sum_{j=1}^{n_i} e_{ij}^2 = e' \Delta e,$$

where $\Delta = \text{block diag}(p_1 n_1^{-1}1_{n_1}, \ldots, p_k n_k^{-1}1_{n_k})$. C. R. Rao (1970) suggests that $A$ should be chosen such that the difference $e' A e - e' \Delta e$ is as small as possible. Since $e$ is practically unobservable, he proposes to minimize $\| A - \Delta \|$ under the Euclidean norm $\| \cdot \|$ subject to (1.8) and (1.9), which is equivalent to minimizing $tr(A^2)$ subject to (1.8) and (1.9). The solution is

$$\mathcal{Y}' A \mathcal{Y} = \sum_{i}^{k} \lambda_i e' F_i e, \quad (1.10)$$

where $\lambda_i$'s are such that (1.9) holds, and $e$ is the OLS residual vector, i.e.,

$$\hat{e} = \mathcal{Y} - \mathcal{X} \hat{\beta}_{OLS}, \quad (1.11)$$

and $\hat{\beta}_{OLS} = (\mathcal{X}' \mathcal{X})^{-1} \mathcal{X}' \mathcal{Y}$ is the ordinary least squares estimator.

Estimation of each individual $\sigma_i^2$ is a special case of estimating $p^T \sigma^2$. Therefore, the MINQU estimator of $\sigma^2$ is unbiased. However, it can possibly take negative value, which creates difficulty in using the WLS estimator of $\beta$ given by (1.7). Some modification is usually necessary.

Rao (1973) showed that the MINQUE of $\sigma_i^2$, say $\hat{\sigma}_{Mi}^2$, is close to the average of squared residuals in the sense that

$$\hat{\sigma}_{Mi}^2 = n_i^{-1} \sum_{j=1}^{n_i} \hat{e}_{ij}^2 + O_p(k^{-1}),$$

where $\hat{e}_{ij}$ is the element of $\hat{e}$ in (1.11) corresponding to $j$-th observation in $i$-th group. Motivated by this result, Fuller and Rao (1978) propose a two-step estimator of $\beta$ for model (1.3). First they use $\hat{\sigma}_i^2 = n_i^{-1} \sum_{j=1}^{n_i} \hat{e}_{ij}^2$ as an estimator of $\sigma_i^2$ (also called average of squared residual (ASR) estimator). Then using $\hat{\sigma}_i^2$ to estimate $V$ in (1.6), they obtained the two-step WLS estimator $\hat{\beta}_w$ of $\beta$ from (1.7). Under the assumption of boundedness of $\sigma_i^2$ and $n_i$, normally distributed errors $e_{ij}$, and some conditions on the design matrix $\mathcal{X}$, they derived the asymptotic distribution of $\hat{\beta}_w$ as $k \rightarrow \infty$.
and \( n_i \geq 3 \). The case of small \( n_i \) is important since it is often impractical to obtain more than 4–5 replicates at a point. Their result was improved by Shao (1989a). Shao derived a similar result under weaker conditions on the design matrix \( X \), and replaced the normality assumption by the assumption of a symmetric distribution with \((2+\delta)\)-moments (sufficient condition). Moreover, Shao provided a consistent estimator of the asymptotic variance of \( \hat{\beta}_w \). Shao (1989b) also obtained consistent jackknife estimators of all the terms in the asymptotic variance of \( \hat{\beta}_w \), after proving the inconsistency of the usual delete-one-pair jackknife variance estimator (see Wu, 1986). Rao (1980) proposed a delete–group jackknife variance estimator of \( \hat{\beta}_w \). It performed well in a simulation study for the special case \( x'_i \beta = \mu \). Shao and Rao (1993) proved the consistency of the delete–group jackknife variance estimator. The variance estimator is also robust against possible within group dependence among random errors \( e_{ij} \).

### 1.1.3 Nested error regression models

Fuller and Battese (1973) considered a nested error regression model with constant error variances, i.e. case 3 of our model (1.1), but they did not require replicated observations. Their approach is applied below to Case 3 of our model (1.1) with \( \sigma_i^2 \neq 0 \) and \( \sigma_i^2 = \sigma^2 \) for all \( i \).

A transformation of the model is used such that the variance of the transformed errors \( \varepsilon_{ij}^* \) are uncorrelated and its variance independent of \( \sigma_i^2 \), or \( \varepsilon_{ij}^* \overset{iid}{\sim} (0, \sigma_e^2) \):

\[
y_{ij} - w_i \bar{y}_i = \sum_{i=1}^{p} (1 - w_i) x_{ii} \beta_i + \varepsilon_{ij}^*
\]

where

\[
w_i = 1 - \left( \frac{\sigma_e^2}{(\sigma_e^2 + n_i \sigma_i^2)} \right)^{1/2},
\]

and

\[
\varepsilon_{ij}^* = \varepsilon_{ij} - w_i \varepsilon_i, \quad \text{with} \quad \varepsilon_i = \sum_{i=1}^{n_i} \varepsilon_{ij} / n_i.
\]

Since \( \sigma_i^2 \) and \( \sigma_e^2 \) are unknown, their estimators \( \hat{\sigma}_i^2 \) and \( \hat{\sigma}_e^2 \) are obtained through the...
well-known “fitting-of-constants” method proposed by Henderson (1953) as follows:

\[ \hat{\sigma}_v^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n - k}, \]

\[ \hat{\sigma}_v^2 = \max(0, \hat{\sigma}_v^2), \]

where

\[ \hat{\sigma}_v^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon} - (n - p)\hat{\sigma}_v^2}{n - tr[(\mathcal{X}'\mathcal{X})^{-1}\sum_{i=1}^{k} n_i^2 x_i'x_i]}, \]

and

\[ \hat{\varepsilon} = (I_n - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}')\mathcal{Y}, \]

the vector of OLS residual from the regression of \( \mathcal{Y} \) on \( \mathcal{X} \). Note that \( \hat{\sigma}_v^2 \) and \( \hat{\sigma}_v^2 \) are unbiased estimators of \( \sigma_v^2 \) and \( \sigma_v^2 \) respectively, while the truncated estimator \( \hat{\sigma}_v^2 \) is biased but consistent.

With these estimators of \( \sigma_v^2 \) and \( \sigma_v^2 \), an OLS estimator, \( \hat{\beta} \), of \( \beta \) is obtained from the following model:

\[ y_{ij} - \hat{\omega}_i \bar{y}_i = \sum_{k=1}^{p} (1 - \hat{\omega}_i) x_{ik} \beta_k + \varepsilon_i^* \]

with

\[ \hat{\omega}_i = 1 - [\hat{\sigma}_v^2/(\hat{\sigma}_v^2 + n_i \hat{\sigma}_v^2)]^{1/2}. \]

This estimator, \( \hat{\beta} \), is identical to the two-step WLS estimator obtained from (1.7) using the above estimators of \( \sigma_v^2 \) and \( \sigma_v^2 \). Fuller and Battese (1973) showed that \( \hat{\beta} \) is unbiased for \( \beta \) under the assumptions of symmetrically distributed errors \( \varepsilon_i^* \) with fourth moments and the existence of the expectation of \( (\hat{\sigma}_v^2)^{-1} \). Further they showed that \( \hat{\beta} \) has the same asymptotic distribution as the WLS estimator using the true values of \( \sigma_v^2 \) and \( \sigma_v^2 \) as weights.

### 1.1.4 Common mean case

P. S. R. S. Rao, Kaplan and Cochran (1981) discussed the special case of model (1.1) with \( x_i' \beta = \mu \). They summarized eight methods of estimating \( \sigma_v^2 \) and \( \sigma_v^2 \). These methods include the analysis of variance (ANOVA), the unweighted sums of squares.
(USS), the minimum norm quadratic unbiased (MINQU) estimation with a priori values (four different sets of prior values), the two-stage MINQU estimation, the average of squared residuals (ASR), the ASR-SSE type estimators, the estimators assuming that $\sigma_i^2$ are equal, and the maximum likelihood estimators (MLE). With these estimators of variance components, they obtained the weighted least squares (WLS) estimators of $\mu$. They also considered the grand mean, the mean of grouped means, and the MLE for estimating $\mu$. They calculated or simulated (by Monte Carlo methods) their biases and mean squared errors for variation in the magnitudes of the unknown parameters, the sample sizes and the number of groups. But they didn’t give any analytic results on the asymptotic distribution of these estimators.

1.1.5 Random intercept model

P. S. R. S. Rao and Kuranchie (1988) considered the special case of model (1.1) with $x_i^t \beta = \alpha + \beta x_i$. For estimation of variance components $\sigma_0^2$ and $\sigma_1^2$, they considered ANOVA estimators, least squares estimators, MINQU estimators with a priori values, ASR estimators, minimum mean square error estimators (MIMSQE), and maximum likelihood estimators (MLE). In their Monte Carlo simulation study, they discussed two-stage estimators of error variances and $\alpha$ and $\beta$ with these non-parametric procedures (the first five methods above). The two stages of estimation are as follows: (1) Starting with 1 as a priori values for both $\sigma_0^2$ and $\sigma_1^2$, calculate the WLS of $(\alpha, \beta)$. With this estimate of $(\alpha, \beta)$, compute the first-stage estimator of $\sigma_0^2$ and $\sigma_1^2$, and using the first-stage estimators of $\sigma_0^2$ and $\sigma_1^2$, obtain first-stage WLS estimator of $(\alpha, \beta)$. (2) From the first-stage WLS estimator of $(\alpha, \beta)$, obtain the second-stage estimator of $\sigma_0^2$ and $\sigma_1^2$. Then the second-stage WLS estimator of $(\alpha, \beta)$ is obtained using the second-stage weights. Using Monte–Carlo methods, they compared biases, variances and mean square error of these estimators. But no analytic results were presented. Further inferences, such as confidence intervals on the regression parameter, were not considered.
1.2 Introduction to Small Area Estimation

Most traditional large-scale sample surveys are typically designed to provide reliable estimates for large geographical regions and large subgroups of a population. For example, the U.S. National Health Interview Survey provides reliable estimates of health characteristics, such as mortality, disability and utilization of health services, for the nation and four broad geographical regions. The Canadian Labour Force Survey gives reliable estimates of unemployment rate for the nation and all provinces. However, estimates are also needed for small areas, such as counties and Health Service Areas (National Center for Health Statistics, 1979). Such estimates are increasingly being used in formulating policies and programs, in allocation of government funds, and in regional programs, etc.

Direct survey estimates for a small area, based on the data only from the sample units in the area, are likely to yield unacceptably large standard errors due to unduly small size of the sample in the area. Alternative estimates that borrow strength from related small areas are therefore needed to improve efficiency. Such estimates are based on either implicit or explicit models which provide a link to related small areas through supplementary data such as administrative records and recent census counts.

Most of the research on small area estimation has focused on cross-sectional data at a given point in time. Rao (1986) and Ghosh and Rao (1993), among others, have given an account of this research. Estimators proposed in the literature include (a) synthetic estimators (Gonzalez, 1973; Ericksen, 1974), and structure preserving estimators (Purcell and Kish, 1980); (b) sample-size dependent estimators (Drew et al., 1982; Särndal and Hidiroglou, 1989); (c) empirical Bayes estimators (Fay and Herriot, 1979; Ghosh and Lahiri, 1987), and two-stage estimators or empirical BLUP (best linear unbiased prediction) estimators (Prasad and Rao, 1986 and 1990; Battese et al., 1988); (d) hierarchical Bayes estimators (Datta and Ghosh, 1991).

Scott and Smith (1974), Jones (1980), Binder and Dick (1985) and others used time series methods to develop efficient estimators of aggregates (e.g., overall means)
from repeated surveys, by combining the direct survey estimates over time. Tiller (1992) used the Kalman filter to combine a current period state-wide estimate from the U.S. Current Population Survey with past estimates for the same state and auxiliary data from the Unemployment Insurance System and the Current Employment Statistics Payroll Survey. However, neither Scott and Smith (1974) nor Tiller (1992) considered small area estimation by combining time series and cross-sectional data.

In the rest of this section, we review some models in the literature. We start with the Fay–Herriot model for cross-sectional data in Section 1.2.1. Synthetic estimators, empirical Bayes or two-stage estimators, and hierarchical Bayes estimators of small area means are introduced. In Section 1.2.2, we review some models proposed in the econometrics literature that combine time series and cross-sectional data. Then we present small area models based on these ideas.

1.2.1 Cross-sectional models

We first obtain regression synthetic estimators, assuming a deterministic model on the small area means. We then allow uncertainty into the model and obtain empirical Bayes or two-stage estimators of small area means. We also consider the hierarchical Bayes estimator.

Regression synthetic estimators

A synthetic estimator refers to an estimator which is designed for a large area, but is applied to a small area under the assumption that the small area has the same characteristics as the large area (see Gonzalez, 1973). A synthetic estimator is simple to use and applicable to general sampling designs, and so has been used traditionally for small area estimation. But it can lead to serious bias if the assumption is not satisfied.

Let \( y_{iT} \) be the direct survey estimator of \( i \)-th small area mean for the current period \( T \), say \( \theta_{iT} \) (\( i = 1, \ldots, k \)). We assume that \( y_{iT} \) is design-unbiased for \( \theta_{iT} \), i.e.,

\[
y_{iT} = \theta_{iT} + e_{iT},
\]

(1.12)
where $e_{iT}$'s are sampling errors with $E(e_{iT}) = 0$, given $\theta_{iT}$. We assume that a vector of fixed concomitant variables $x_{iT} = (x_{iT1}, \ldots, x_{iT_p})'$ related to $\theta_{iT}$ is available for all $i$ at time $T$; for example, time varying supplementary data such as administrative records and other auxiliary data such as census data.

A simple regression synthetic estimator of the current small area mean $\theta_{iT}$, based solely on the cross-sectional data $\{(y_{iT}, x_{iT}), i = 1, \ldots, k\}$ for time $T$, is obtained by assuming the following deterministic model on $\theta_{iT}$:

$$\theta_{iT} = x'_{iT}\beta_T,$$  \hspace{1cm} (1.13)

where $\beta_T = (\beta_{T1}, \ldots, \beta_{Tp})'$ is the vector of regression coefficients. It is given by

$$\hat{\theta}_{iT}(\text{reg}) = x'_{iT}\hat{\beta}_T,$$  \hspace{1cm} (1.14)

where $\hat{\beta}_T$ is the ordinary least squares estimator of $\beta_T$ obtained from the combined model based on (1.12) and (1.13):

$$y_{iT} = x'_{iT}\beta_T + e_{iT}, \hspace{0.5cm} i = 1, \ldots, k$$

for time $T$.

Synthetic estimators like (1.14) could lead to substantial design biases since they give a zero weight to the direct estimator $y_{iT}$. On the other hand, empirical Bayes or two-stage estimators give "proper" weights to the survey estimator and the synthetic estimator. As a result they can lead to smaller biases relative to synthetic estimators.

**Two-stage estimation**

Following Fay and Herriot (1979), we can introduce uncertainty into the model (1.13) as follows:

$$\theta_{iT} = x'_{iT}\beta_T + v_{iT},$$  \hspace{1cm} (1.15)

where the error terms $v_{iT}$'s are independent random variables with mean 0 and unknown variance $\sigma^2_v$. For sampling errors, we assume that the $e_{iT}$'s are independent
normal variables with \( E(e_{iT}) = 0 \) and \( V(e_{iT}) = \sigma_{iT}^2 \), where \( \sigma_{iT}^2 \) is known. The combined model is given by

\[
y_{iT} = x_{iT}' \beta_T + u_{iT} + e_{iT}. \tag{1.16}
\]

A two-stage estimator (or variance components estimator) of \( \theta_{iT} \) is obtained in two steps. In the first step, the best linear unbiased prediction (BLUP) estimator under model (1.16), assuming \( \sigma_{iT}^2 \) is known, is obtained as follows, using Henderson’s (1950) general result for mixed linear models:

\[
\hat{\theta}_{iT}^H = \frac{\sigma_{iT}^2}{\sigma_{iT}^2 + \hat{\sigma}_{iT}^2} y_{iT} + \frac{\hat{\sigma}_{iT}^2}{\sigma_{iT}^2 + \hat{\sigma}_{iT}^2} x_{iT}' \hat{\beta}_T(\sigma_{iT}^2). \tag{1.17}
\]

where \( \hat{\beta}_T(\sigma_{iT}^2) \) is the weighted least squares estimator of \( \beta_T \) under the combined model (1.16). In the second step, a suitable estimator \( \hat{\sigma}_{iT}^2 \) is substituted for \( \sigma_{iT}^2 \) to obtained the so-called empirical BLUP (EBLUP) estimator. The EBLUP estimator can be written as a weighted sum of the direct estimator \( y_{iT} \) and the regression synthetic estimator \( \hat{\theta}_{iT}(reg) = x_{iT}' \hat{\beta}_T(\sigma_{iT}^2) \):

\[
\hat{\theta}_{iT}^H = w_{iT} y_{iT} + (1 - w_{iT}) \hat{\theta}_{iT}(reg), \tag{1.18}
\]

where \( y_{iT} = (y_{iT}, \ldots, y_{kT})' \), and \( w_{iT} = \hat{\sigma}_{iT}^2 / (\hat{\sigma}_{iT}^2 + \sigma_{iT}^2) \). This result holds without assuming normal distributions for \( u_{iT} \) and \( e_{iT} \).

Note that the ratio \( \hat{\sigma}_{iT}^2 / \sigma_{iT}^2 \) measures between small area variation relative to sampling variance. More weight, \( w_{iT} \), is given to the direct estimator \( y_{iT} \) as this ratio increases.

Prasad and Rao (1986 and 1990) used the following moment estimator of \( \sigma_{iT}^2 \):

\[
\hat{\sigma}_{iT}^2 = \max(\hat{\sigma}_{iT}^2, 0), \tag{1.19}
\]

\[
\hat{\sigma}_{iT}^2 = (k - p)^{-1} \left[ \sum_{i=1}^{k} \hat{a}_{iT}^2 - \sum_{i=1}^{k} \sigma_{iT}^2 (1 - x_{iT}'X_T^{-1}X_T^{-1}x_{iT}) \right],
\]

where \( \hat{a}_{iT} = y_{iT} - x_{iT}'(X_T^{-1}X_T)^{-1}X_T'y_{iT} \) and \( X_T = (x_{1T}, x_{2T}, \ldots, x_{kT}) \). More complicated estimators, such as the restricted maximum likelihood (REML) estimator, have also been employed in the literature (see, e.g., Cressie, 1992).
CHAPTER 1. INTRODUCTION

The estimation of MSE of the EBLUP estimator is considered in Prasad and Rao (1986 and 1990). Using a moment estimator $\hat{\sigma}_T^2$ of $\sigma_T^2$ in (1.19), and under normality of $v_i$, they proved under certain regularity conditions that the MSE of $\hat{\theta}_T^H$ can be approximated correct to term of order $O(k^{-1})$ by the following quantity:

$$MSE(\hat{\theta}_T^H) \approx g_{1i}(\sigma_T^2) + g_{2i}(\sigma_T^2) + g_{3i}(\sigma_T^2),$$

where

$$g_{1i}(\sigma_T^2) = \sigma_T^2 \sigma_T^2 \sigma_T^2 + \sigma_T^2 - 1,$$

$$g_{2i}(\sigma_T^2) = \sigma_T^2 \sigma_T^2 (\sigma_T^2 + \sigma_T^2 - 1)^{-2} x_i T (X' T V^{-1} X_T)^{-1} x_i T,$$

and

$$g_{3i}(\sigma_T^2) = \sigma_T^2 \sigma_T^2 (\sigma_T^2 + \sigma_T^2 - 1)^{-3} [2 k^{-1} \sigma_T^2 + 2 \sigma_T^2 \sum_{i=1}^{k} \sigma_T^2 / k + \sum_{i=1}^{k} \sigma_T^2 / k],$$

where $V = \sigma_T^2 J_k + \text{diag}(\sigma_T^2)$ with $J_k$ being a $k \times k$ matrix with 1 as every element. They also provided an estimator of the above MSE as

$$\text{mse}(\hat{\theta}_T^H) = g_{1i}(\sigma_T^2) + g_{2i}(\sigma_T^2) + 2 g_{3i}(\sigma_T^2),$$

(1.20)

correct to the same order of approximation. Lahiri and Rao (1992) showed that the above estimator of MSE above is also valid under moderate non-normality of the random effect, $v_i$. Thus the inference based on $\hat{\theta}_T^H$ and mse($\hat{\theta}_T^H$) is robust to nonnormality of the random effects.

Empirical Bayes Estimation

In the empirical Bayes (EB) approach, $y_i T$ is assumed to be normal distributed with mean $\theta_T$ and variance $\sigma_T^2$ conditional on $\theta_i T$, and $\theta_T$ is assumed to be normal distributed with mean $x_i T \beta_T$ and variance $\sigma_T^2$. Hence, the posterior distribution of $\theta_T$ given $y_i T$, $\beta_T$ and $\sigma_T^2$ is normal with mean $\theta_T^B$ and variance $g_{12}(\sigma_T^2)$, where

$$\theta_T^B = E(\theta_T | y_i T, \beta_T, \sigma_T^2) = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_T^2} y_i T + \frac{\sigma_T^2}{\sigma_T^2 + \sigma_T^2} x_i T \beta_T.$$  

(1.21)

Under quadratic loss, $\theta_T^B$ is the Bayes estimator of $\theta_T$. Since $\sigma_T^2$ and $\beta_T$ are unknown, estimators $\hat{\theta}_T^B$ and $\hat{\beta}_T$ are obtained using ML, REML or method of
moments as in the two-step estimation. The estimated posterior distribution is \( N(\hat{\theta}_{EB}^{T}, g_{1i}(\hat{\sigma}_{\nu T}^{2})) \), where \( \hat{\theta}_{EB}^{T} \) is identical to \( \hat{\theta}_{EB}^{T} \). A naive EB approach uses \( \hat{\theta}_{EB}^{T} \) as the estimator of \( \theta_{iT} \) and measure its uncertainty by the estimated posterior variance

\[
\text{Var}(\theta_{iT} | y_{iT}, \hat{\beta}_{T}, \hat{\sigma}_{\nu T}^{2}) = g_{1i}(\hat{\sigma}_{\nu T}^{2}).
\]

Although this EB estimator \( \hat{\theta}_{EB}^{T} \) is approximately equal to the true posterior mean \( E(\theta_{iT} | y_{T}) \), \( g_{1i}(\hat{\sigma}_{\nu T}^{2}) \) underestimates the true posterior variance \( \text{Var}(\theta_{iT} | y_{T}) \) since the variation due to estimating \( \sigma_{\nu T}^{2} \) and \( \beta_{T} \) is not taken into account. Two methods of accounting for this underestimation can be found in Laird and Louis (1987) and Kass and Steffy (1989).

**Hierarchical Bayes estimation**

We now apply the hierarchical Bayes (HB) approach to model (1.16). We assume the same conditional distributions on \( y_{iT} \) and \( \theta_{iT} \) as those in the EB approach, i.e., \( y_{iT} | \theta_{iT} \sim N(0, \sigma_{\nu T}^{2}) \), and \( \theta_{iT} | \beta_{T}, \sigma_{\nu T}^{2} \sim N(x'_{iT}\beta_{T}, \sigma_{\nu T}^{2}) \). Further we assume prior distributions for \( \beta_{T} \) and \( \sigma_{\nu T}^{2} \). The inference are then based on the posterior distribution of \( \theta_{iT} \) given observations \( y_{T} \).

By assuming that the prior distribution of \( \beta_{T} \) is a uniform over \( \mathbb{R}^{p} \), we can show that the posterior distribution of \( \theta_{iT} \) given \( y_{T} \) and \( \sigma_{\nu T}^{2} \) is normal distribution with mean equal to the BLUP \( \hat{\theta}_{iT}^{H} \) and variance equal to \( g_{1i}(\sigma_{\nu T}^{2}) + g_{2i}(\sigma_{\nu T}^{2}) \). Hence when \( \sigma_{\nu T}^{2} \) is assumed to be known, the HB and the BLUP approach lead to identical inferences.

To take the uncertainty about \( \sigma_{\nu T}^{2} \) into account, we need to calculate the posterior distribution of \( \theta_{iT} \) given the observations \( y_{T} \), by assuming a suitable prior on \( \sigma_{\nu T}^{2} \). The HB estimator of \( \theta_{iT} \) is then given by

\[
\hat{\theta}_{iT}^{HB} = E(\theta_{iT} | y_{T}) = E_{\sigma_{\nu T}^{2}}(\hat{\theta}_{iT}^{H}).
\]  

A measurement of uncertainty associated with \( \hat{\theta}_{iT}^{HB} \) is given by the posterior variance

\[
\text{Var}(\theta_{iT} | y_{T}) = E_{\sigma_{\nu T}^{2}}[g_{1i}(\sigma_{\nu T}^{2}) + g_{2i}(\sigma_{\nu T}^{2})] + V_{\sigma_{\nu T}^{2}}(\hat{\theta}_{iT}^{H}),
\]
where $E_{\sigma^2}$ and $V_{\sigma^2}$ respectively denote the expectation and variance with respect to the posterior distribution of $\sigma^2_T$ given $y_T$. Numerical evaluations of (1.22) and (1.23) involve one-dimensional integration.

An excellent application of hierarchical Bayes approach to small area estimation can be seen in Datta and Ghosh (1991).

The three methods, EBLUP, EB and HB, lead to comparable estimators of the parameter of interest. But EB approach leads to underestimation of uncertainty. The estimator of MSE associated with the EBLUP estimator is accurate only with large number of small areas. The HB measure of uncertainty is applicable even for a small number or moderate number of small areas, but requires suitable priors for model parameters. Also intensive computation are involved in this approach. For more complex models, Gibbs sampler, which will be introduced in Section 7.1.2, is needed.

Fay and Herriot (1979) used estimators of the form (1.18) to estimate per capita income for small areas (with population less than 500 or 1000). They presented empirical evidence that (1.18) leads to smaller average error than either the direct survey estimate (based on a 20 percent sample from the 1970 U.S. Census of Population and Housing) or the county average. Fay (1987) and Datta, Fay and Ghosh (1991) extended the Fay-Herriot model to multiple characteristics of interest, and derived empirical Bayes and hierarchical Bayes estimators of small area means.

There are many other small area models dealing with cross-sectional data, e.g., Battese, Harter and Fuller (1988), Stukel (1991), Fuller and Harter (1987). MacGibbon and Tomberlin (1989) studied logistic regression models with random small area effects for binary data using empirical Bayes methods.

### 1.2.2 Cross-sectional and time series models

The methods in Section 1.2.1 use only cross-sectional data for the current period $T$. As a result, they do not exploit the information in the data at other time points. Cross-sectional and time series models are needed to take advantage of data at other time points. We consider one such model here, and using this model we
extend the Fay-Herriot approach to time series of direct small area estimators \{y_{it}\}, in conjunction with supplementary data \{x_{it}\}, i = 1, \ldots, k; t = 1, \ldots, T. We assume that

\[ y_{it} = \theta_{it} + e_{it}, \quad i = 1, \ldots, k; \quad t = 1, \ldots, T, \]  

(1.24)

where the \(e_{it}\)'s are sampling errors with \(E(e_{it}) = 0\), given the small area means \(\theta_{it}\).

Previously proposed models

Extensive econometrics literature exists on modelling and estimating relationships that combine time series and cross-sectional data (see Judge et al., 1985, Chapter 13), but sampling errors are seldom taken into account. We now consider some of these models on \(\theta_{it}\):

(I) \[ \theta_{it} = x_{it}'\beta + v_i + \epsilon_{it}, \]

where \(x_{it} = (x_{i1}, \ldots, x_{ip})'\) is the vector of fixed concomitant variables for \(i\)-th area at time \(t\), \(\beta = (\beta_1, \ldots, \beta_p)'\) is a vector of regression parameters, the \(v_i\)'s are fixed small area effects, and the \(\epsilon_{it}\)'s are independent normal variables with mean 0 and variance \(\sigma^2\), abbreviated \(\epsilon_{it} \sim N(0, \sigma^2)\).

(II) \[ \theta_{it} = x_{it}'\beta + v_i + \epsilon_{it}, \]

where \(v_i \sim N(0, \sigma_v^2)\), \(\epsilon_{it} \sim N(0, \sigma^2)\) and \{\(v_i\)\} and \{\(\epsilon_{it}\)\} are independent. Here the \(v_i\)'s are random small area effects.

(III) \[ \theta_{it} = x_{it}'\beta + v_i + u_t + \epsilon_{it}, \]

where \(v_i \sim N(0, \sigma_v^2)\), \(u_t \sim N(0, \sigma_u^2)\), \(\epsilon_{it} \sim N(0, \sigma^2)\) and \{\(v_i\)\}, \{\(u_t\)\}, \{\(\epsilon_{it}\)\} are mutually independent. Here the \(v_i\)'s and the \(u_t\)'s are random small area effects and random time effects respectively.

(IV) \[ \theta_{it} = x_{it}'\beta + v_i + u_{it} \]

and

\[ u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1, \]

(1.25)

where \(v_i \sim N(0, \sigma_v^2)\), \(\epsilon_{it} \sim N(0, \sigma^2)\) and \{\(v_i\)\}, \{\(\epsilon_{it}\)\} are independent (Anderson and Hsiao, 1981). Here the \(v_i\)'s are random small area effects and the \(u_{it}\)'s follow an
CHAPTER 1. INTRODUCTION

AR(1) process. Model (1.25) may be rewritten as a distributed lag model:

\[ \theta_{it} = \rho \theta_{i,t-1} + (x_{it} - \rho x_{i,t-1})' \beta + (1 - \rho) v_i + \epsilon_{it}. \]  \hspace{1cm} (1.25a)

The alternative form (1.25a) relates \( \theta_{it} \) to the previous period mean \( \theta_{i,t-1} \), the values of the auxiliary variables for the time points \( t \) and \( t - 1 \), and the small area effect \( v_i \). The special case of \( \rho = 0 \) which gives a nested error variance components model has also been studied in the econometrics literature.

More complex models than (1.25) may be formulated by assuming an ARMA process for the \( u_{it} \)'s instead of the simple AR(1) process, for example, Binder and Dick (1989) and Tiller (1989) used ARMA models but the resulting efficiency gains relative to (1.27) are unlikely to be significant. Similarly, random slopes \( \{ \beta_{it} \} \) obeying an autoregressive process may be used in place of constant slopes \( \beta \) (Pfeffermann and Burck, 1990; Singh, Mantel and Thomas, 1991), but empirical evidence seems to suggest that the resulting efficiency gains relative to (1.27) are likely to be small (Singh, Mantel and Thomas, 1991).

The \( \theta_{it} \)'s are related to direct survey estimators \( y_{it} \) through (1.24) and (1.25). Choudhry and Rao (1989) treated the composite error \( w_{it} = e_{it} + u_{it} \) as an AR(1) process and assumed \( \theta_{it} = x_{it}' \beta + v_i \), i.e., time effects on true area means are only through \( x_{it} \) unlike (1.25). The resulting nested error regression model may be written as

\[ y_{it} = x_{it}' \beta + v_i + w_{it} \]
\[ w_{it} = \hat{\rho} w_{i,t-1} + \tilde{\epsilon}_{it}, \hspace{0.5cm} |\hat{\rho}| < 1, \] \hspace{1cm} (1.26)

where \( \tilde{\epsilon}_{it} \overset{iid}{\sim} N(0, \sigma^2) \). They obtained a two-stage estimator of small area mean \( \theta_{it} \) under (1.26), and evaluated its efficiency relative to two synthetic estimators and the direct estimator, \( y_{it} \), using monthly survey estimates of unemployment for census divisions (small areas) from the Canadian Labour Force Survey in conjunction with monthly administrative counts from the Unemployment Insurance System and monthly survey estimates of population in labour force as auxiliary variables.
CHAPTER 1. INTRODUCTION

Proposed model

Model (1.26) does not explicitly depend on the sampling error, but it is less realistic than the combined model based on (1.24) and (1.25), viz.,

\[ y_{it} = \theta_{it} + e_{it} = x_{it}' \beta + u_{it} + e_{it} \]
\[ u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1. \tag{1.27} \]

Following Fay and Herriot (1979), we assume that the sampling errors \( e_{it} \) are normally distributed with zero mean and block diagonal covariance matrix \( \Sigma \) with known blocks, \( \Sigma_i \), where \( \Sigma_i \) is a \( T \times T \) matrix, i.e., independent of sample errors across small areas. The covariance matrices \( \Sigma_i \) may be obtained by using the method of generalized variance functions (see Wolter, 1985, Chapter 5) and prior estimates of panel correlations (see Lee, 1990); a direct estimator of \( \Sigma_i \) is likely to be unstable due to small sample sizes in the areas. Model (1.27) provides an extension of the Fay–Herriot model to cross-sectional data and time series data.

We focus on the extended Fay–Herriot model (1.27), and obtain a two-stage estimator and hierarchical Bayes estimator of the current small area means, \( \theta_{i,t} \), in Chapters 6 and 7.

1.3 Thesis Outline

In this section, we present an outline of this dissertation. It is divided into two subsections. In Section 1.3.1, we give an outline of Chapters 2 to 5, dealing with inferences for one and two-fold nested-error regression models. In Section 1.3.2, we give an outline of Chapters 6 and 7, dealing with small area estimation.

1.3.1 Outline for Part I

In Chapters 2 to 4, we concentrate on the general model (1.1). Denote

\[ y_{n \times 1} = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{k1}, \ldots, y_{kn_k})', \]
\[ x_{n \times p} = (x_{11} l_{n_1}', x_{21} l_{n_2}', \ldots, x_{k1} l_{n_k}'), \]
\[ \mathbf{e}_{n \times 1} = (e_{11}, \ldots, e_{1n_1}, \ldots, e_{k1}, \ldots, e_{kn_k})', \]

where \( l_i \) is a \( t \)-vector with all entries equal to one. A matrix form of model (1.1) is then given by

\[ \mathbf{y} = \mathbf{x}' \beta + \mathbf{e}. \tag{1.28} \]

The dispersion matrix of the error \( \mathbf{e} \) is

\[ \mathbf{V} = \text{Var}(\mathbf{e}) = \text{block diag}(\sigma_0^2 I_{n_1} + \sigma_1^2 I_{n_1} l_1', \ldots, \sigma_k^2 I_{n_k} + \sigma_k^2 I_{n_k} l_k'). \]

Without loss of generality, we assume \( \mathbf{X} \) is of full column rank (so is the matrix \( \mathbf{X} \) defined below). The primary interest is the estimation of \( \beta \), and \( \sigma_0^2 \).

In the literature, a WLS estimator of \( \beta \) is generally used for models with unequal (heteroscedastic) error variances, e.g., Fuller and Rao (1978), Shao (1989a). When \( \sigma_0^2 \) and \( \sigma_i^2 \) are known, the WLS estimator of \( \beta \) is given by

\[ \hat{\beta}_w = \left( \mathbf{x}' \mathbf{V}^{-1} \mathbf{x} \right)^{-1} \mathbf{x}' \mathbf{V}^{-1} \mathbf{y}. \tag{1.29} \]

Due to the special structure of our design matrix, with replicates within each group, \( \hat{\beta}_w \) reduces to

\[ \tilde{\hat{\beta}}_w = \left( \mathbf{x}' \mathbf{W}^{-1} \mathbf{x} \right)^{-1} \mathbf{x}' \mathbf{W}^{-1} \mathbf{y}, \tag{1.30} \]

where

\[ \mathbf{x} = (x_1, \ldots, x_k)', \quad \mathbf{W} = \text{diag}(w_i), \quad w_i = \sigma_0^2 + n_i^{-1} \sigma_i^2, \]

\[ \mathbf{y} = (\tilde{y}_1, \ldots, \tilde{y}_k)', \quad \tilde{y}_i = \sum_{i=1}^{n_i} y_{ij}/n_i, \]

and \( \text{diag}(a_i) \) denotes a \( k \times k \) diagonal matrix with \( a_1, \ldots, a_k \) as diagonal elements.

Since \( \sigma_0^2 \) and \( \sigma_i^2 \) are usually unknown, a two-step WLS estimator is used in practice:

\[ \hat{\beta}_w = \left( \mathbf{x}' \hat{\mathbf{V}}^{-1} \mathbf{x} \right)^{-1} \mathbf{x}' \hat{\mathbf{V}}^{-1} \mathbf{y} = \left( \mathbf{x}' \hat{\mathbf{W}}^{-1} \mathbf{x} \right)^{-1} \mathbf{x}' \hat{\mathbf{W}}^{-1} \mathbf{y}, \tag{1.31} \]

where \( \hat{\mathbf{V}} \) and \( \hat{\mathbf{W}} \) are estimators of \( \mathbf{V} \) and \( \mathbf{W} \) obtained by substituting suitable estimators into \( \mathbf{V} \) and \( \mathbf{W} \) for \( \sigma_0^2 \) and \( \sigma_i^2 \).

For estimating \( \sigma_0^2 \) and \( \sigma_i^2 \), we introduce in Section 2.1 a weighted least squares (WLS) procedure, viz., given a priori values for \( \sigma_0^2 \) and \( \sigma_i^2 \), the WLS estimator of
$\beta$ using these prior values is calculated. Then $\sigma^2_1$ and $\sigma^2_2$ are estimated by equating the components of the weighted sum of squares of residuals to their expectations. The analysis of variance (ANOVA) estimation is seen as its special case. It is called least squares estimator (LSE) in P. S. R. S. Rao and Kuranchie (1988). The minimum norm quadratic unbiased (MINQU) estimators with a priori values for $\sigma^2_1$ and $\sigma^2_2$ are derived in Section 2.2. The average of squared residual (ASR) estimators are introduced in Section 2.3. In Section 2.4, approximate MINQU (AMINQU) estimators are obtained from the MINQU estimation equations (with a priori values) in Section 2.2, by ignoring the lower order terms under certain conditions. It turns out that our estimators after truncation guarantee positive weights for the WLS estimator of regression parameters, unlike the ANOVA or MINQU estimators. After we present the model assumptions and some preliminary results in Section 5.2.1, the asymptotic normality of our AMINQUE estimators of $\sigma^2_2$ are established in Section 2.5.2. In Section 2.5.3, an estimator of the asymptotic variance of our AMINQU estimators is obtained. A confidence interval for $\sigma^2_2$ is then constructed.

Using the AMINQUE estimators of $\sigma^2_1$ and $\sigma^2_2$, we construct a two-stage WLS estimator of $\beta$ in Chapter 3. We present some preliminary results in Section 3.1. Then we establish the asymptotic normality of the WLS estimator $\hat{\beta}_w$ of $\beta$ in Section 3.2. A consistent estimator of the asymptotic variance–covariance matrix of $\hat{\beta}_w$ is given in Section 3.3. In Section 3.4, we introduce a modification using the AMINQUE estimator of $\sigma^2_2$ and the within group sample variances as estimator of $\sigma^2_2$'s. The resulting WLS estimator of $\beta$ is also shown to be asymptotically normal. A consistent estimator of its asymptotic variance–covariance matrix is also obtained.

In Chapter 4, we present the jackknife estimation methodology for our model (1.1). Since the observations within a group are correlated, the usual delete–1 jackknife variance estimators, based on the assumption of independent observations, is inconsistent. On the other hand, deleting all observation in the same group at a time only makes use of between group independence and maintains the within–group correlation structure. Therefore, the delete–group jackknife technique is applied to our model. In Section 4.1, we review the classical jackknife methodology. We con-
CHAPTER 1. INTRODUCTION

consider the delete-group jackknife estimators of variance component $\sigma^2_r$ in Section 4.2. Delete-group jackknife variance estimators of $\sigma^2_r$ are constructed and proved to be consistent. In Section 4.3, we construct a delete-group jackknife variance-covariance estimator of the WLS estimator of $\beta$, and prove its consistency. The jackknife variance estimator for the alternate WLS estimator of $\beta$ is outlined in Section 4.4. In Section 4.5, Monte Carlo simulation results are presented to throw some light on the performance of our AMINQU estimators, jackknife estimators of $\sigma^2_r$, variance estimators for $\sigma^2_r$, the substitution and delete-group jackknife estimators of asymptotic covariance matrix of WLS estimators of regression parameters.

In Chapter 5, we extend our AMINQU estimation method to two-fold nested error regression models with unequal error variances. The approach is parallel to that in Chapters 2 and 3. We give a brief introduction to this model in Section 5.1. In Section 5.2, we apply MINQU, ASR, WLS, AMINQU estimation procedures to our model. Due to the model complexity, explicit expressions for AMINQU estimators of variance components for only a special set of a priori values are obtained. The properties of AMINQU estimators are discussed. We consider the estimation of the regression parameters in Section 5.3. Since AMINQU estimators do not guarantee positive weights, the approach in Section 3.4 is adopted. The asymptotic normality of the induced WLS estimator of regression parameter $\beta$ is established. A consistent estimator of its asymptotic variance-covariance matrix is also obtained.

1.3.2 Outline for Part II

The main purpose of Part II of this dissertation is to study small area estimation under a combined cross-sectional and time series model involving autocorrelated random effects and sampling errors. This model, given by (1.27), was discussed in Section 1.2.2. Further analysis of the model is presented in Chapters 6 and 7.

In Chapter 6, we focus on the case of small or moderate number of time points $T$ and relatively large number of small areas $k$. But even when $T$ is large it may be adequate to consider only a moderate number of time points, say 5-10, in estimating the current small area means, $\theta_{iT}$. We first derive the BLUP estimator of $\theta_{iT}$ for
model (1.27) by using general results in Henderson (1950). Then the estimators of variance components $\sigma^2_\rho$ and $\sigma^2$ are obtained using moment estimators assuming $\rho$ is known. With $\rho$ known, a two-stage or EBLUP estimator of $\theta_{iT}$ is obtained by substituting the estimator of $\sigma^2_\rho$ and $\sigma^2$ into BLUP. Extending the approach of Prasad and Rao (1986, 1990) for the Fay–Herriot model with cross-sectional data, we then obtain in Section 6.3 an estimator of MSE of the two-stage estimator, correct to a second-order term for a small or moderate number of time points, $T$, and a relatively large number of small areas, $k$. The case of $\rho$ unknown is discussed in Section 6.4. We present a method for estimating $\rho$, and construct the EBLUP estimator of $\theta_{iT}$ using the estimator of $\rho$ and corresponding estimators of $\sigma^2_\rho$ and $\sigma^2$. A second order approximation to MSE and estimator of the MSE for this EBLUP estimator is similarly derived. Our simulation results, however, indicated that this estimator of $\rho$ doesn't perform well in small samples. Therefore we suggest two alternative approaches in Section 6.4.2, and obtain EBLUP estimators of $\theta_{iT}$ and corresponding second order approximations to MSE and estimators of MSE.

In Chapter 7, we present the hierarchical Bayes (HB) analysis to model (1.27). We also discuss two cases: autocorrelation $\rho$ known and $\rho$ unknown. We propose a multistage HB model on the parameter $\theta = (\theta_{i1}, \ldots, \theta_{iT}, \ldots, \theta_{k1}, \ldots, \theta_{kT})'$ and other model parameters in each case. Two methods are used to obtain the posterior expectation and posterior covariance matrix of our parameter $\theta$. The first method is based on numerical integration, while the second method employs the recently developed Gibbs sampling (see Gelfand and Smith, 1990).
Part I

Inference for One and Two-Fold Nested-Error Regression Models
Chapter 2

Estimating Variance Components

In this chapter, asymptotic inference on the variance component \( \sigma_v^2 \) is considered for the general nested error regression model (1.1). Estimators of error variances \( \sigma_i^2 \) are also obtained. The estimators of \( \sigma_v^2 \) and \( \sigma_i^2 \) are needed for inference on \( \beta \), but inference on \( \sigma_v^2 \) is also of interest in practice.

There are many methods for estimating the variance components proposed in the literature. We review the weighted least squares (WLS) procedure (analysis of variance (ANOVA) method is a special case of this procedure), the minimum norm quadratic unbiased (MINQU) estimation with a priori values, and the average of squared residuals (ASR) procedure with respect to our model (1.1), in Sections 2.1–2.3. In Section 2.4, the so called approximate MINQU (AMINQU) estimators for \( \sigma_v^2 \) and \( \sigma_i^2 \) are derived, by dropping lower order terms in MINQU estimation equations with a priori values. The AMINQU estimators of \( \sigma_v^2 \) are biased, but they are asymptotically unbiased under certain conditions, as shown in Section 2.5.2. Asymptotic normality of these estimators is also established. A consistent estimator of the common asymptotic variance is obtained by the substitution method in Section 2.5.3. A method for constructing confidence intervals for \( \sigma_v^2 \) is also suggested.

2.1 WLS Estimators

The weighted least squares (WLS) procedure for estimating the variance components \( \sigma_v^2 \) and \( \sigma_i^2 \) is a generalization of the ANOVA approach. It is referred to as least
squares estimation (LSE) procedure in P. S. R. S. Rao and Kuranchie (1988). The method is as follows: Given a set of prior values \( \sigma^2_{10} \) and \( \sigma^2_{10} \), a WLS estimator of \( \beta \) is obtained from (1.31) with \( W \) obtained from a priori values. Then a particular types of weighted sum of squares of residuals is considered which can be written as a sum of quadratic terms, and then the estimators of \( \sigma^2_v \) and \( \sigma^2_i \) are obtained by equating these quadratic terms to their expectations. The details are given below.

Given a priori values \( \sigma^2_{10} \), \( \sigma^2_{10} \) for \( \sigma^2_v \) and \( \sigma^2_i \) respectively, we can construct the weights

\[
w_{i0} = \sigma^2_{10} + \sigma^2_{10}/n_i.
\]

Denote

\[
W_0 = \text{diag}(w_{i0}),
\]

and

\[
V_0 = \text{block diag}(\sigma^2_{10}I_{n_1} + \sigma^2_{10}1_{n_1}1'_{n_1}, \ldots, \sigma^2_{k0}I_{n_k} + \sigma^2_{k0}1_{n_k}1'_{n_k}).
\]

Then we can find \( \hat{\beta}_0 \) of \( \beta \) from (1.31):

\[
\hat{\beta}_0 = (X'W_0^{-1}X)^{-1}X'W_0^{-1}y.
\]

We next consider the following weighted sum of squares of residuals

\[
R^2_0 = (Y - X'\hat{\beta}_0)'V_0^{-1}(Y - X'\hat{\beta}_0).
\]

After simplification, it can be shown that

\[
R^2_0 = \sum_{i=1}^{k} \sigma^{-2}_{10} \left[ \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 \right] + \left[ \sum_{i=1}^{k} w_{i0}^{-1} (\hat{y}_i - x'_i\hat{\beta}_0)^2 \right],
\]

which is a sum of \( k + 1 \) quadratic forms in \( Y \) (given in the square brackets). By equating all these terms to their expectations, WLS estimators of \( \sigma^2_v \) and \( \sigma^2_i \) can be found from the following expressions:

\[
\hat{\sigma}^2_i = s^2_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 \tag{2.1}
\]
\[
\hat{\sigma}^2_v \sum_{i=1}^{k} w_{i0}^{-1} A_i + \sum_{i=1}^{k} n_i^{-1} w_{i0}^{-1} \hat{\sigma}^2_i A_i = \sum_{i=1}^{k} w_{i0}^{-1} (\hat{y}_i - x'_i\hat{\beta}_0)^2, \tag{2.2}
\]
where

\[ A_i = 1 - w_{i0}^{-1} x_i' \left( \sum_{i=1}^{k} w_{i0}^{-1} x_i x_i' \right)^{-1} x_i. \]

The advantages of this procedure are: (1) \( \hat{\sigma}_e^2 \) and \( \hat{\sigma}_0^2 \) are unbiased, (2) \( \hat{\sigma}_e^2 \) is easily evaluated from (2.1), (3) as will be seen, the ANOVA procedure is a special case of this method. However, \( \hat{\sigma}_0^2 \) may take negative values. We cannot therefore guarantee positive weights for the WLS estimation of \( \beta \).

**ANOVA estimation.** The analysis of variance (ANOVA) estimator can be obtained from the WLS procedure with \( \sigma_{i0}^2 = 0 \) and \( \sigma_{i0}^2 = 1 \) for all \( i \). Hence \( \hat{\sigma}_e^2 = s_i^2 \) as in (2.1), and \( \hat{\sigma}_0^2 \) can be solved from

\[ \hat{\sigma}_0^2 \sum_{i=1}^{k} n_i A_i + \sum_{i=1}^{k} \hat{\sigma}_e^2 A_i = \sum_{i=1}^{k} n_i (\bar{y}_i - x_i' \hat{\beta}_0)^2, \quad (2.3) \]

where

\[ A_i = 1 - n_i x_i' \left( \sum_{i=1}^{k} n_i^{-1} x_i x_i' \right)^{-1} x_i. \]

We now give two examples of ANOVA estimator.

**Example 1.** If \( x_i' \beta = \mu \), our model is reduced to that considered in P. S. R. S. Rao, Kaplan and Cochran (1981). Since \( \hat{\beta}_0 = \hat{\mu} = \sum_{i=1}^{k} n_i \bar{y}_i / n = \bar{y} \), and \( \sum_{i=1}^{k} n_i x_i x_i' = n \), \( \hat{\sigma}_0^2 \) is given from (2.3) as

\[ \hat{\sigma}_0^2 (n - \sum_{i=1}^{k} n_i^2 / n) + \sum_{i=1}^{k} \hat{\sigma}_e^2 (1 - n_i / n) = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2, \]

where \( \hat{\sigma}_e^2 = s_i^2 \).

**Example 2.** If \( x_i' \beta = \alpha + \beta x_i \), our model is reduced to that of P. S. R. S. Rao and Kuranchie (1988). The ANOVA estimator of \( \sigma_e^2 \) is given by \( \hat{\sigma}_e^2 = s_i^2 \). And the estimator of \( \sigma_0^2 \), from (2.2), is given by

\[ \hat{\sigma}_0^2 \sum_{i=1}^{k} n_i (1 - q_i) + \sum_{i=1}^{k} (1 - q_i) \hat{\sigma}_e^2 = \sum_{i=1}^{k} n_i [\bar{y}_i - \bar{y} - \hat{\beta}_0 (x_i - \bar{x})]^2, \]

where \( q_i = (n_i / n) + n_i (x_i - \bar{x})^2 / [\sum_{i=1}^{k} n_i (x_i - \bar{x})^2] \), and \( \hat{\beta}_0 = [\sum_{i=1}^{k} n_i (x_i - \bar{x}) \bar{y}_i] / [\sum_{i=1}^{k} n_i (x_i - \bar{x})^2] \). (Note that \( x_i \) and \( \beta \) are scalars in this case.)

**USS estimator.** It is noted that the procedure with the unweighted sums of squares (USS) is also a special case of our WLS procedure. The USS estimators of \( \sigma_e^2 \) and \( \sigma_0^2 \) can be obtained from (2.1) and (2.2) with \( \sigma_{i0}^2 = 1 \) and \( \sigma_{i0}^2 = 0 \) for all \( i \).
2.2 MINQUE Procedure with a Priori Values

In Section 1.1.2, the idea of minimum norm quadratic unbiased (MINQU) estimation was introduced under model (1.3). For our model (1.28), MINQU estimation with a priori values is usually used. This idea was introduced in C. R. Rao (1971, 1972), which is quite similar to that of MINQU estimation. We now apply this method to our model.

In model (1.28), the error term $\mathcal{E}$ can be written as a weighted sum

$$\mathcal{E} = U_1\xi_1 + \cdots + U_k\xi_k + U_v\xi_v,$$

where $U_i' = (0 : I_n : 0)_{n \times n}$, $\xi_i$ is an $n_i$-vector with mean 0 and dispersion matrix $\sigma_i^2 I_n$, $U_v = \text{block diag}_i(1_{n_i})_{n \times k}$, and $\xi_v$ is a $k$-vector with mean 0 and dispersion matrix $\sigma_v^2 I_k$. For a set of a priori values $\sigma_i^2, \sigma_k^2, \ldots, \sigma_v^2$, denote

$$V_0 = \sigma_i^2 F_1 + \cdots + \sigma_k^2 F_k + \sigma_v^2 F_v,$$

where $F_v = U_v U_v'$, and $F_i = U_i U_i'$.

Parameters of interest are of the form $p' \theta = p_1 \sigma_i^2 + \cdots + p_k \sigma_k^2 + p_v \sigma_v^2$ for specified $p' = (p_1, \ldots, p_k, p_v)$ where $\theta' = (\sigma_i^2, \ldots, \sigma_k^2, \sigma_v^2)$. Similar to MINQUE in Section 1.1.2, the estimator of $p' \theta$ under MINQUE with a priori values is a quadratic form $\mathcal{Y} A \mathcal{Y}$ where $A$ is obtained by minimizing $tr(A V_0 A V_0)$ subject to the condition of invariance, $A \mathcal{X} = 0$, and the conditions of unbiasedness, $tr(AX_i) = p_i$, $i = 1, \ldots, k$, and $tr(AF_v) = p_v$ (Note that $tr(A)$ denotes the trace of matrix $A$). C. R. Rao (1971, 1972) gave the solution, which is quite similar to the MINQUE solution in (1.10). C. P. Rao and Kleffe (1988) (see p.96, Chapter 5) extended it to estimating $\theta = (\sigma_i^2, \ldots, \sigma_k^2, \sigma_v^2)'$ jointly. The estimator of $\theta$ by MINQUE with a priori values is the solution to the following system of linear equations:

$$H_{UI} \hat{\theta} = q_{UI}, \quad (2.4)$$

where

$$H_{UI} = \begin{pmatrix} (tr(M V_0 M)^+ F_i (M V_0 M)^+ F_j)_{k \times k} & \mu \\ \mu' & tr(M V_0 M)^+ F_v (M V_0 M)^+ F_v \end{pmatrix}.$$
\[ q_{ui} = (Q_1, \ldots, Q_k, Q_v)^	op, \]

with
\[ M = I_n - X(X'X)^{-1}X', \]
\[ \mu = (tr(MV_0M)^+F_1(MV_0M)^+F_v, \ldots, tr(MV_0M)^+F_k(MV_0M)^+F_v)_{k 	imes 1}'. \]
\[ Q_v = \mathcal{Y}(MV_0M)^+F_v(MV_0M)^+\mathcal{Y}, \]
\[ Q_i = \mathcal{Y}(MV_0M)^+F_i(MV_0M)^+\mathcal{Y}, \]

and \( A^+ \) as the Moore-Penrose generalized inverse (g-inverse) of \( A \) (see C. R. Rao, 1973). Equation (2.4) is usually referred to as the MINQUE equations.

Using the following formula (a special case of formula 1.3.4 in C. R. Rao and Kleffe (1988), and it can be verified directly)
\[ (MV_0M)^+ = V_0^{-1} - V_0^{-1}X(X'V_0^{-1}X)^{-1}X'V_0^{-1}, \quad (2.5) \]
in the right hand side of (2.4), \( Q_v \) and \( Q_i \)'s can be simplified as
\[ Q_v = \frac{1}{\sigma^2} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + n_i^{-1}w_{io}^{-2}(\bar{y}_i - \hat{\beta}_0'x_i)^2, \quad (2.6) \]
\[ Q_i = \sum_{i=1}^{k} w_{io}^{-2}(\bar{y}_i - \hat{\beta}_0'x_i)^2, \quad (2.7) \]

where
\[ w_{io} = \sigma^2 + n_i^{-1}\sigma^2, \]

and
\[ \hat{\beta}_0 = (X'W_0^{-1}X)^{-1}X'W_0^{-1}y \]
\[ W_0 = \text{diag}_i(w_{io}). \quad (2.8) \]

However, for the left hand side of (2.4), except in very simple cases, further simplification seems to be difficult. Therefore, an explicit expression for the estimator of \( \theta \) under MINQUE with a priori values is usually not available. P. S. R. S. Rao, Kaplan and Cochran (1981) have given an expression for the common mean case,
CHAPTER 2. ESTIMATING VARIANCE COMPONENTS

i.e., \( x'_i \beta = \mu \). This makes it difficult to study the asymptotic properties of these estimators, let alone the properties of the WLS estimator of \( \beta \) using weights constructed from these variance component estimators. Also, this procedure may lead to negative estimators for \( \sigma^2 \) and \( \sigma_i^2 \) (especially for the former parameter). As a result, these estimators cannot be used in constructing a WLS estimator of \( \beta \) without some arbitrary modification to guarantee positive weights.

2.3 ASR Estimator

Let \( \sigma^2_{i0}, \sigma^2_{i1}, \ldots, \sigma^2_{ik} \) be a set of a priori values of \( \sigma^2_i, \sigma^2_i, \ldots, \sigma^2_i \) respectively. Following the general approach of average of squared residuals (ASR) in P. S. R. S. Rao (1977) and P. S. R. S. Rao and Chaubey (1978), we obtained ASR estimators for \( \sigma^2_v \) and \( \sigma^2_i \) as

\[
\hat{\sigma}_v^2 = (\sigma^4_{i0}/k)Q_v,
\]

\[
\hat{\sigma}_i^2 = (\sigma^4_{i0}/n_i)Q_i,
\]

where \( Q_v \) and \( Q_i \)'s are the same as those in (2.6) and (2.7). These estimators are clearly non-negative. But, unless some conditions are satisfied (e.g., \( n_i \)'s are quite large, which is not the case we discuss here), the bias can be quite large. However, empirical results have shown that the MSE of the ASR estimator is usually smaller than the variance of ANOVA and MINQU estimators (P. S. R. S. Rao, Kaplan and Cochran, 1981). Asymptotic properties of the WLS estimator of \( \beta \) using weights constructed from the ASR estimator are unknown.

2.4 AMIQUE Procedure

To avoid the difficulties with the estimators from the MINQU estimation with a priori values, the approximate MINQU (AMINU) estimation procedure is proposed in this section.

Note that in the right hand side of formula (2.5), the first term \( V_0^{-1} \) is of higher order than the second term. To be exact, under certain conditions (which will be
stated explicitly in Section 2.5) on the group size \( n_i \), the design matrix, and bounded prior values, we have

\[
(MV_0M)^+ = V_0^{-1} + O(k^{-1}).
\]

Therefore, \( H_{UI} \) in (2.4) is approximately equal to \( H^*_U \) in the sense that

\[
H_{UI} = H^*_U + O(k^{-1}),
\]

where

\[
H^*_U = \begin{pmatrix}
    (\text{tr}(V_0^{-1}F_iV_0^{-1}F_j))_{k\times k} & \text{col}_i\left(\text{tr}(V_0^{-1}F_iV_0^{-1}F_v)\right) \\
    \text{col}_i'\left(\text{tr}(V_0^{-1}F_iV_0^{-1}F_v)\right) & \text{tr}(V_0^{-1}F_vV_0^{-1}F_v)
\end{pmatrix},
\]

where \( \text{col}_i(a_i) \) denotes a \( k \)-vector with \( a_i \) as \( i \)-th element, \( \text{col}_i'(a_i) = (\text{col}_i(a_i))' \).

Evaluating the traces in the above matrix, we obtain that

\[
H^*_U = \begin{pmatrix}
    \text{diag}(n_i^{-2}w_{i0}^{-2}d_i^{-1})_{k\times k} & \text{col}_i(n_i^{-2}w_{i0}^{-2}) \\
    \text{col}_i'(n_i^{-2}w_{i0}^{-2}) & \sum_{i=1}^k w_{i0}^{-2}
\end{pmatrix},
\]

where \( w_{i0} = \sigma_{i0}^2 + \sigma_{i0}^2/n_i \) and

\[
d_i = \frac{1}{1 + (n_i - 1)(1 + \rho_{i0}n_i)^2}.
\]  

(2.10)

with

\[
\rho_{i0} = \sigma_{i0}^2/\sigma_{i0}^2.
\]

Now the linear system (2.4) is reduced to the following simpler linear system

\[
H^*_U\tilde{\theta} = q_{UI},
\]  

(2.11)

with

\[
\tilde{\theta} = (\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_k^2, \tilde{\sigma}_v^2)'.
\]

The matrix \( H^*_U \) is non-singular and its inverse can be found by applying the following lemma in matrix theory (see C. R. Rao, 1973, p.33):

**Lemma 2.1** Let

\[
D = \begin{pmatrix}
    A & B \\
    B' & C
\end{pmatrix},
\]
be a non-singular matrix. If \( C \) and \( H = A - B'C'B \) are invertible, then

\[
D^{-1} = \begin{pmatrix}
H^{-1} & -H^{-1}BC^{-1} \\
-BC^{-1}H^{-1} & C^{-1} - B'H^{-1}BC^{-1}
\end{pmatrix}.
\]

Let \( A = \text{diag}(n_i^{-2}w_{i0}^{-2}d_i^{-1}) \), \( B = \text{col}(n_i^{-1}w_{i0}^{-2}) \) and \( C = \sum_{i=1}^{k} w_{i0}^{-2} \). Then we have

\[
(H_U^{-1})^{-1} = \begin{pmatrix}
\text{diag}(n_i^2w_{i0}^2d_i) + \nu^{-1}\text{col}(n_id_i)\text{col}'(n_id_i) & -\nu^{-1}\text{col}(n_id_i) \\
-\nu^{-1}\text{col}'(n_id_i) & \nu^{-1}
\end{pmatrix},
\]

where

\[
\nu = \sum_{i=1}^{k} c_i, \quad \text{and} \quad c_i = n_i^2(n_i - 1)\sigma_{i0}^{-4}d_i.
\]  \hspace{1cm} (2.12)

The solution to linear system (2.11) is

\[
\hat{\theta} = (H_U^{-1})^{-1}q_{u_1}.
\]

Hence, our untruncated AMINQUE estimators of \( \sigma_v^2 \) and \( \sigma_i^2 \) are obtained as

\[
\hat{\sigma}_v^2 = \frac{1}{\sum_i c_i} (Q_v - \sum_{i=1}^{k} n_id_iQ_i), \quad \hspace{1cm} (2.13)
\]

\[
\hat{\sigma}_i^2 = n_i^2w_{i0}^2d_iQ_i - n_id_i\hat{\sigma}_v^2. \quad \hspace{1cm} (2.14)
\]

Like the MINQUE estimators, \( \hat{\sigma}_v^2 \) and \( \hat{\sigma}_i^2 \) may also take negative values which is beyond the parameter space for \( \sigma_v^2 \) and \( \sigma_i^2 \). We propose the following adjustment to obtain a nonnegative estimator \( \hat{\sigma}_v^2 \) of \( \sigma_v^2 \):

\[
\hat{\sigma}_v^2 = \max(0, \hat{\sigma}_v^2). \quad \hspace{1cm} (2.15)
\]

An estimator of \( \sigma_i^2 \) is obtained correspondingly as

\[
\hat{\sigma}_i^2 = n_i^2w_{i0}^2d_iQ_i - n_id_i\hat{\sigma}_v^2. \quad \hspace{1cm} (2.16)
\]

Note that \( \hat{\sigma}_i^2 \) may still take a negative value. We will discuss this later in the remarks.

**Remark 1.** We discussed our AMINQUE procedure for the case of replicates within groups. The reasons for doing so are: (1) In practical situations, replication is often employed. (2) This case is simpler to illustrate. However, the AMINQUE estimators
for the case of no replications can be similarly obtained, although they are more complicated.

**Remark 2.** We saw that (2.13) and (2.14) may lead to negative values. The same holds for other unbiased estimation procedures. Generally, there are non-negative unbiased quadratic estimators for the random error variance \( \sigma^2_r \), e.g., the within group sample variance. But it can be shown that there does not exist a non-negative unbiased quadratic estimator for \( \sigma^2_v \) (La Motte, 1973). Kleffe and Rao (1986) have shown that even asymptotically consistent nonnegative quadratic estimators for \( \sigma^2_v \) do not exist.

**Remark 3.** The modified estimator \( \hat{\sigma}^2_v \) may take negative values, but this is not important since estimation of \( \sigma^2_v \) is not our major concern. Also, as the group size, \( n_i \), is small, there is no really good estimator for \( \sigma^2_v \) except when some parametric assumption is imposed. Despite the possible negativeness of \( \hat{\sigma}^2_v \), we still get positive weights using (2.15) and (2.16):

\[
\hat{w}_i = \hat{\sigma}^2_v + n_i^{-1} \hat{\sigma}^2_i = (1 - d_i) \hat{\sigma}^2_v + n_i w_{ii} d_i Q_i > 0.
\]  

(2.17)

The weights \( \hat{w}_i \) can be applied directly to construct \( \hat{\mathbf{W}} \) in (1.31) for the WLS estimator of \( \beta \). Other adjustments are not required, like those discussed in C. R. Rao (1984), P. S. R. S. Rao, Kaplan and Cochran (1981). This is an advantage of our approach.

**Remark 4.** When some of the \( \sigma^2_i \) are equal, the corresponding \( F_i \) should be pooled together before calculating pooled \( Q_i \). For the purpose of estimating regression coefficient \( \beta \), such pooling seems unnecessary as shown in the simulation results later in Chapter 4. The asymptotic theory for regression coefficients given in Chapters 3 and 4 does not depend on the distinctness of \( \sigma^2_i \).

**Remark 5.** For the special case \( \sigma^2_v = 0 \), our model reduced to model (1.2). Since we do not have the parameter \( \sigma^2_v \), we should exclude it while applying our AMINQUE procedure. Let \( \sigma^2_0 = 1 \) for all \( i \). From (2.4), (2.5)–(2.11), the AMINQUE estimator of \( \sigma^2_i \) can be obtained as

\[
\hat{\sigma}^2_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - x'_i \hat{\beta}_0)^2,
\]
which is identical to the estimator of $\sigma_*^2$ considered by Rao (1973), Fuller and Rao (1978), Shao (1989a, 1989b). Their estimator can therefore be seen as an application of AMINQUE to their model.

**Remark 6.** The following differences between MINQUE and AMINQUE may be noted:

1. MINQUE gives unbiased estimators while AMINQUE gives biased estimators. However, as shown in the Monte Carlo study, the variance of the MINQU estimators is larger than the MSE of AMINQU estimators.

2. AMINQUE gives explicit forms for the estimators of all variance components, $\sigma_*^2, \sigma_1^2, \ldots, \sigma_q^2$, unlike the MINQUE, while the calculation of MINQU estimators is complicated for general design matrices, since it involves inverting $\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X}$. Even for the common mean case (P. S. R. S. Rao, Kaplan and Cochran, 1981), the expressions for MINQU estimators are quite complicated.

### 2.5 Asymptotic Properties of $\hat{\sigma}_v^2$

It is often impractical to obtain more than 4–5 replicates in each group (see Jacquez et al, 1968; Fuller and Rao, 1978 and Shao, 1989a). Therefore we consider the case of bounded $n_i$. With bounded $n_i$ it is impossible to obtained a consistent estimator for $\sigma_*^2$ unless $\sigma_i^2$ is somehow structurally explicit, e.g., a smooth function of the design points, regression coefficients, or some auxiliary parameters. We establish the asymptotic consistency and asymptotic normality of $\hat{\sigma}_v^2$ and $\hat{\sigma}_*^2$ in Section 2.5.2, after we present the model assumptions, some definitions and preliminary results in Section 2.5.1. In Section 2.5.3, consistent estimators of the asymptotic variance of AMINQU estimator are obtained by the substitution method. These results are important since they enable us to construct confidence intervals for $\sigma_*^2$ which attain the desired level asymptotically.
2.5.1 The model assumptions and preliminary results

For proving the consistency of our AMINQU estimator of $\sigma^2_0$, we impose the following conditions on model (1.1):

(A) There are positive constants $\sigma_0^2, \sigma^2_\infty, c_\infty$, and a positive integer $n_\infty$ such that $\sigma_0^2 \leq \sigma_i^2 \leq \sigma^2_\infty, 2 \leq n_i \leq n_\infty$ and $\|x_i\| \leq c_\infty$ for all $i$, where $\|x\| = (x'x)^{1/2}$ is the Euclidean norm of $x$.

(B) There is a positive constant $c_0$ such that

$$c_0 \leq k^{-1} (\text{the minimum eigenvalue of } X'X).$$

(C) Both $v_i$ and $e_{ij}$ are symmetrically distributed about zero, and there is a positive constant $r_1$ such that

$$E |v_i|^{4+2\delta} \leq r_1 \quad \text{and} \quad E|e_{ij}|^{2+\delta} \leq r_1 \quad \text{for all } i, j$$

for some positive constant $\delta$.

Assumption (B) together with the boundedness of $\|x_i\|$ implies that

$$\lambda_0 I_p \leq k^{-1} X'X \leq \lambda_\infty I_p$$

for all $k$ and some positive constants $\lambda_0, \lambda_\infty$.

The symmetry of the distributions in condition (C) implies that $Ev_i = 0, Ev^3_i = 0, Ee_{ij} = 0$, and

$$E \frac{\hat{e}_i}{[r_2 + r_3 \sum_{j=1}^n (e_{ij} - \bar{e}_i)^2 + r_4 \bar{e}^2_i]^s} = 0,$$

for any positive constants $r_2, r_3, r_4$, and $s$, where

$$\hat{e}_i = v_i + \bar{e}_i, \quad \bar{e}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} e_{ij}.$$

The property (2.20) is used in Chapter 3.

By the independence of $v_i, e_{ij}$, and condition (A)

$$\text{Var}(\hat{e}_i) = \sigma_0^2 + \sigma^2_i/n_i \leq C$$
for some positive constant $C$. From (2.18), for the $\delta$ given in condition (C) by $C_\tau$-inequality (see Lemma 2.2)

$$E|\varepsilon_i|^2 \leq C_1$$

(2.21)

for some positive constant $C_1$.

For the prior values $\sigma^2_0$, $\sigma^2_{10}, \ldots, \sigma^2_{k0}$, we are able to choose them so that they are bounded away from both 0 and $\infty$. Therefore, from condition (A), $c_i$'s in (2.12) and $d_i$'s in (2.10) satisfy that

$$0 < c_0^* \leq c_i \leq c_\infty^* < +\infty, \quad 0 < d_0^* \leq d_i < \frac{1}{2}$$

(2.22)

for some positive constants $c_0^*$, $c_\infty^*$, $d_0^*$.

From (2.19), $W_0 = \text{diag}(w_i)$ in (2.9) satisfies that

$$\lambda_0^* I_p \leq k^{-1} X' W_0^{-1} X \leq \lambda_\infty^* I_p$$

(2.23)

for some positive constants $\lambda_0^*$, $\lambda_\infty^*$.

When we consider the limiting properties as $k \to \infty$, a subscript should be attached to $\varphi^2$, the design $X$ (or $X$), the estimators of $\beta$ and other quantities. For simplicity of notation, however, subscript $k$ is omitted except when confusion can arise.

Note that in the following context, all $\delta$'s refers to the one in condition (C) except in some lemmas and their proofs where a $\delta$ is mentioned in the assumptions. Also lower case letters and upper case letters have totally different meaning in the context. Lower case letters $c$'s and $d$'s refer in general to fixed constant mentioned in this section, while $C$, $C_1$ or $C_2$ appear in many places of this paper. These upper case letters mean some positive constants, and may assume different values at different places.

We now present some definitions and preliminary results.

**Definition 2.1** Let $\{\xi_k\}$ be a sequence of random variables. $\{\xi_k\}$ is said to converge to a random variable $\xi$ in distribution, denoted as $\xi_k \to_d \xi$, if the cumulative distribution function (cdf) of $\xi_k$ converges to the cdf of $\xi$ at almost all points where the latter is continuous.
Definition 2.2 Let \( \{\xi_k\} \) and \( \{\eta_k\} \) be two sequences of random variables. Then \( \{\xi_k\} \) is said to converge to 0 in probability if and only if for any \( \delta > 0 \), \( P(|\xi_k| \geq \delta) \to 0 \) as \( k \to \infty \). It is denoted as \( \xi_k \to_p 0 \) or \( \xi_k = \sigma_p(1) \) or \( \xi_k \) is of order \( o_p(1) \). Also denote \( \xi_k = o_p(\eta_k) \) (\( \xi_k \) is of order \( o_p(\eta_k) \)) if \( \xi_k/\eta_k = o_p(1) \).

Definition 2.3 Let \( \{\xi_k\} \) and \( \{\eta_k\} \) be two sequences of random variables. Then \( \{\xi_k\} \) is said to be bounded in probability if and only if for any \( \delta > 0 \), there exist an \( M_\delta > 0 \) and a positive integer \( K_\delta \) such that \( P(|\xi_k| > M_\delta) < \delta \) for all \( k > K_\delta \). It is denoted as \( \xi_k = O_p(1) \) or \( \xi_k \) is of order \( O_p(1) \). And \( \xi_k = O_p(\eta_k) \) (\( \xi_k \) is of order \( o_p(\eta_k) \)) if and only if \( \xi_k/\eta_k = O_p(1) \).

We have the following results on \( o_p, O_p \) (see e.g., Bishop et al., 1975, §14.2):

\[
O_p(1) \cdot o_p(1) = o_p(1), \quad O_p(1) \cdot O_p(1) = O_p(1).
\]

(2.24)

The following lemma is the well known \( C_r \)-inequality. It will be used very frequently in this dissertation.

Lemma 2.2 \( (C_r\text{-inequality}) \) Let \( \xi \) and \( \eta \) be two random variables, and \( r \) be a positive real number. Then

\[
E(|\xi + \eta|^r) \leq 2^r [E(|\xi|^r) + E(|\eta|^r)].
\]

where \( s = \min(1, r) \).

The following result, a generalization of Chebychev Inequality, is an easy tool for proving convergence in probability. The proof is very simple and hence omitted.

Lemma 2.3 Suppose \( \{\xi_k\} \) is a sequence of random variables with expectations \( E(\xi_k) = \mu_k < \infty \), and finite \( (1+\epsilon)\text{-central moments for some } \epsilon \geq 0 \), i.e. \( m_{1+\epsilon} = E(|\xi_k - \mu_k|^{1+\epsilon}) < \infty \). Then for any positive constant \( M > 0 \),

\[
P(|\xi_k - \mu_k| \geq M) \leq m_{1+\epsilon}/M^{1+\epsilon},
\]

i.e., \( \xi_k - \mu_k = O_p(1) \).

Remark. From the above result, it is easy to see that if \( m_{1+\epsilon} \to 0 \) as \( k \to +\infty \), \( \xi_k - \mu_k \to_p 0 \).
CHAPTER 2. ESTIMATING VARIANCE COMPONENTS

The following lemma is a slight generalization of the classical law of large numbers. It is presented here without proof.

Lemma 2.4 Let \( \{\xi_i\} \) be a sequence of independent random variables with \( E\xi_i = \mu_i \), and \( \{a_{ki} : 1 \leq i \leq k\} \) be a double array which satisfies \( \max_i |a_{ki}| \leq M/k \) for some positive constant \( M \). If for a positive \( \delta < 1 \), \( k^{-(1+\delta)} \sum_i E|\xi_i|^{1+\delta} \to 0 \), then \( \sum_i a_{ni}(\xi_i - \mu_i) \to_p 0 \). \( \square \)

Remark. In the special case of \( a_{ki} = k^{-1} \), we have

\[
k^{-1} \sum_i (\xi_i - \mu_i) \to_p 0, \quad \text{as} \quad k \to +\infty \tag{2.25}
\]

under the assumptions of Lemma 2.4.

Now we present some lemmas directly related to our model.

Lemma 2.5 Suppose that conditions (A) and (C) hold. Then for any non-random sequence \( \{z_i\} \) such that \( |z_i| \leq z_\infty \)

\[
k^{-1/2} \sum_{i=1}^{k} z_i \bar{\xi}_i x_i = O_p(1) .
\]

Hence

\[
k^{-1} \sum_{i=1}^{k} z_i \bar{\xi}_i x_i = o_p(1) .
\]

Proof. The \( t \)-th element of \( k^{-1/2} \sum_{i=1}^{k} z_i \bar{\xi}_i x_i \) is

\[
k^{-1/2} \sum_{i=1}^{k} z_i x_i \bar{\xi}_i .
\]

Since

\[
\text{var}(k^{-1/2} \sum_{i=1}^{k} z_i x_i \bar{\xi}_i) = k^{-1} \sum_{i=1}^{k} z_i^2 x_i^2 (\sigma_i^2 + n_i^{-1} \sigma_i^2) \leq C
\]

for some constant \( C > 0 \), the result follows from Lemma 2.3 with \( \epsilon = 1 \). \( \square \)

Lemma 2.6 Suppose that conditions (A), (B) and (C) hold. Then

\[
\hat{\beta}_0 - \beta = O_p(k^{-1/2}) .
\]
CHAPTER 2. ESTIMATING VARIANCE COMPONENTS

Proof. From (2.8)

\[ \hat{\beta}_0 - \beta = (X'W_0^{-1}X)^{-1}X'W_0^{-1}\varepsilon , \]

where \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)' \). The result follows from (2.23) (viz., \( k(X'W_0^{-1}X)^{-1} = O_p(1) \)) and Lemma 2.5 since

\[ k^{1/2}(\hat{\beta}_0 - \beta) = k(X'W_0^{-1}X)^{-1}k^{-1/2}\sum_{i}(\sigma_{i0}^2 + n_i^{-1}\sigma_{io}^2)^{-1}\varepsilon_i x_i . \] \( \square \)

Lemma 2.7 Suppose that conditions (A), (B) and (C) hold. Let \( \{\eta_i\} \) be a sequence of independent random variable such that \( E|\eta_i|^{1+\delta} \leq M \) for some constant \( M > 0 \) and \( \delta > 0 \). Then for any non-negative integer \( m \)

\[ \sum_{i=1}^{k}(\eta_i - E\eta_i)[(\hat{\beta}_0 - \beta)'x_i]^m = o_p(k^{-m/2+1}) . \] \( (2.26) \)

Hence

\[ \sum_{i=1}^{k}\eta_i[(\hat{\beta}_0 - \beta)'x_i]^m = O_p(k^{-m/2+1}) . \] \( (2.27) \)

Proof. Since (2.27) is implied by (2.26), we prove (2.26) only. Note that

\[ |\sum_{i=1}^{k}[\eta_i - E(\eta_i)][(\hat{\beta}_0 - \beta)'x_i]^m| \]
\[ \leq k \sum_{t_1=1}^{p} \ldots \sum_{t_m=1}^{p} k^{-1} \sum_{i=1}^{k} x_{it_1} \ldots x_{it_m}[\eta_i - E(\eta_i)] \prod_{i=1}^{m} |\hat{\beta}_{oi} - \beta_{ti}| \]

where \( \hat{\beta}_{oi} \) is the \( t_i \)-th element of \( \hat{\beta}_0 \). Since \( |\hat{\beta}_{oi} - \beta_{ti}| = O_p(k^{-1/2}) \) from Lemma 2.6, (2.26) follows from

\[ k^{-1} \sum_{i=1}^{k}(\eta_i - E\eta_i)x_{it_1} \ldots x_{it_m} \rightarrow_p 0 , \]

which is implied by Lemma 2.4. \( \square \)

2.5.2 The asymptotic normality of AMINQUE

Now we turn to the main result of this section. We establish the asymptotic normality of \( \hat{\sigma}_0^2 \) first. Then the asymptotic normality of \( \hat{\sigma}_0^2 \), the truncation of \( \sigma_0^2 \) at 0, is just a corollary.
Substituting (2.6), (2.7) into (2.13), we have

\[
\hat{\sigma}_v^2 = \frac{1}{\sum_{i=1}^k c_i} \sum_{i=1}^k c_i \left[ (\hat{y}_i - \hat{\beta}_0 x_i)^2 - \frac{\sum_j (y_{ij} - y_i)^2}{n_i^2 - n_i} \right].
\]

Further noting that \( y_{ij} = x'_i \beta + v_i + e_{ij} \), we can express \( \hat{\sigma}_v^2 \) as

\[
\hat{\sigma}_v^2 = \frac{\sum_i c_i Z_i}{\sum_i c_i} - 2R_{1k} + R_{2k} = \hat{\sigma}_{vp}^2 - 2R_{1k} + R_{2k}
\]

where

\[
Z_i = v_i^2 + 2v_i \hat{e}_i + \sum_j \sum_{i \neq j} e_{ij} e_{il} / (n_i^2 - n_i),
\]

\[
\hat{\sigma}_{vp}^2 = (\sum_{i=1}^k c_i)^{-1} \sum_{i=1}^k c_i Z_i,
\]

\[
R_{1k} = (\sum_{i=1}^k c_i)^{-1} \sum_{i=1}^k c_i \hat{e}_i (\hat{\beta}_0 - \beta)' x_i,
\]

\[
R_{2k} = (\sum_{i=1}^k c_i)^{-1} \sum_{i=1}^k c_i \left[ (\hat{\beta}_0 - \beta)' x_i \right]^2.
\]

We have the following theorem on the asymptotic normality of \( \hat{\sigma}_v^2 \).

**Theorem 2.1** Suppose that assumptions (A), (B) and (C) hold. Then

\[
\frac{\sum_i c_i}{\sqrt{\sum_i c_i^2 \text{var}(Z_i)}} (\hat{\sigma}_v^2 - \sigma_v^2) \to_d N(0, 1) \quad \text{as } k \to +\infty
\]

where

\[
\text{var}(Z_i) = M_4 - \sigma_v^4 + 4\sigma_v^2 \sigma_v^2 / n_i + \sigma_v^4 / (n_i^2 - n_i), \quad \text{with } M_4 = E(v_1^4).
\]

**Proof.** We show first that

\[
\frac{1}{\sqrt{\sum_i c_i^2 \text{var}(Z_i)}} \sum_i c_i (Z_i - \sigma_v^2) \to_d N(0, 1) \quad \text{as } k \to +\infty.
\]

Note that \( EZ_i = \sigma_v^2 \). By using \( C_r \)-inequality repeatedly

\[
E|Z_i - \sigma_v^2|^{2+6} = E \left| v_i^2 + 2v_i \hat{e}_i + \sum_j \sum_{i \neq j} e_{ij} e_{il} / (n_i^2 - n_i) - \sigma_v^2 \right|^{2+6}
\]

\[
\leq 4^{2+6} \left\{ E|v_i|^{4+26} + 2^{2+6} E|v_i|^{2+6} E|\hat{e}_i|^{2+6} + E|\sum_j \sum_{i \neq j} e_{ij} e_{il}|^{2+6} / (n_i^2 - n_i)^{2+6} + \sigma_v^{4+26} \right\}
\]

\[
\leq 4^{2+6} \left\{ E|v_i|^{4+26} + 2^{2+6} E|v_i|^{2+6} \sum_{j=1}^n E|e_{ij}|^{2+6} + \sum_j \sum_{i \neq j} E|e_{ij}|^{2+6} E|e_{il}|^{2+6} + \sigma_v^{4+26} \right\}.
\]
CHAPTER 2. ESTIMATING VARIANCE COMPONENTS

Then from (2.18),
\[ E|Z_i - \sigma^2_i|^{2+\delta} \leq C \]  
(2.33)

for some positive constant C. From (2.22)
\[ E|c_i(Z_i - \sigma^2_i)|^{2+\delta} \leq C_1, \]
for some constant C_1 by C_r-Inequality. Moreover,
\[ \sqrt{\sum_i c_i^2 \text{var}(Z_i)} \geq c_0 \sqrt{\sum_i 4\sigma^2 \sigma^2_i/n_i} \geq \sqrt{k}/C_2 \]  
(2.34)

for some constant C_2 > 0. Therefore as \( k \to +\infty \)
\[ \left( \sum_i c_i^2 \text{var}(Z_i) \right)^{-\frac{(2+\delta)/2}{k}} \sum_i E|c_i(Z_i - \sigma^2_i)|^{2+\delta} \leq C_1 k \cdot k^{-(1+\delta/2)} C_2^{2+\delta} \to 0. \]

Hence Linderberg’s condition is satisfied and thus (2.32) holds.

Next, note that \( \sum_{i=1}^k c_i \geq c_0^* k \) from (2.22). We have \( R_{1k} = \rho_p(k^{-1/2}) \) and \( R_{2k} = \rho_p(k^{-1}) \), since
\[ \sqrt{k}R_{1k} = \frac{k}{\sum_i c_i} \cdot k^{-1/2} \sum_i c_i \hat{e}_i (\hat{\theta}_0 - \beta)' x_i = k^{-1/2} \rho_p(k^{-1/2+1}) = \rho_p(1), \]
and
\[ \sqrt{k}R_{2k} = \frac{k}{\sum_i c_i} \cdot k^{-1/2} \sum_i c_i [(\hat{\theta}_0 - \beta)' x_i]^2 = k^{-1/2} \rho_p(k^{-2/2+1}) = \rho_p(k^{-1/2}), \]
by applying Lemma 2.7. Also, from (2.34),
\[ \sum_{i=1}^k c_i / \left[ \sqrt{\sum_{i=1}^k c_i^2 \text{var}(Z_i)} \right] \leq k c_0^* / \sqrt{k} = C_3 k^{1/2}. \]
for some constant C_3. Thus
\[ \frac{\sqrt{\sum_i c_i}}{\sqrt{\sum_i c_i^2 \text{var}(Z_i)}} (-2R_{1k} + R_{2k}) = \rho_p(1). \]  
(2.35)

Finally, from (2.28), (2.32) and (2.35), by Slutsky’s theorem (see, e.g., Bickel and Doksum, 1977, A.14.9),
\[ \frac{\sum_i c_i}{\sqrt{\sum_i c_i^2 \text{var}(Z_i)}} (\hat{\sigma}^2_0 - \sigma^2_0) = \frac{\sum_i c_i (Z_i - \sigma^2_i)}{\sqrt{\sum_i c_i^2 \text{var}(Z_i)}} + \frac{\sum_i c_i}{\sqrt{\sum_i c_i^2 \text{var}(Z_i)}} (-2R_{1k} + R_{2k}) \]
\[ \to_d N(0, 1) \quad \text{as} \quad k \to +\infty. \]
This completes the proof. 

From the above theorem, the asymptotic variance of $\hat{\sigma}^2_v$ is

$$V_a = \text{var}(\hat{\sigma}^2_v) = \frac{\sum c_i^2 \text{var}(Z_i)}{(\sum c_i)^2}$$

$$= \frac{\sum c_i^2 (M_4 - \sigma^4_v + 4\sigma^2_v \sigma^2_x / n_i + \sigma^4_x / (n_i^2 - n_i))}{(\sum c_i)^2}.$$ 

It can be seen that $V_a$ goes to zero under the conditions of Theorem 2.1 as $k \to +\infty$. From the remark immediately after Lemma 2.3, we have the following corollary.

**Corollary 2.1** Under the assumptions of Theorem 2.1, $\hat{\sigma}^2_v \rightarrow_p \sigma^2_v$.

From the above corollary, for any $\varepsilon > 0$,

$$P\{|\hat{\sigma}^2_v - \sigma^2_v| > \varepsilon\} \to 0$$

as $k \to +\infty$. Hence, when $\sigma^2_v > 0$,

$$P\{\hat{\sigma}^2_v \leq 0\} \to 0$$

as $k \to +\infty$. Therefore

$$P\{\hat{\sigma}^2_v = \hat{\sigma}^2_v\} \to 1,$$

as $k \to +\infty$, or

$$\hat{\sigma}^2_v - \hat{\sigma}^2_v \rightarrow_p 0.$$ 

The above argument, combined with Theorem 2.1 gives the following corollary.

**Corollary 2.2** Assume $\sigma^2_v > 0$. Then under the assumptions of Theorem 2.1,

$$\hat{\sigma}^2_v = \max(0, \hat{\sigma}^2_v) \rightarrow_p \sigma^2_v,$$

and

$$\frac{\sum c_i}{\sqrt{\sum c_i^2 \text{var}(Z_i)}} (\hat{\sigma}^2_v - \sigma^2_v) \rightarrow_d N(0, 1)$$

Results in the last corollary imply that $V_a$ is also the asymptotic variance of $\hat{\sigma}^2_v$.

**Remark.** The replication of $x_i$'s within each group is not critical in the proof of Theorem 2.1. When the design points within group are not duplicates, asymptotic results can be established similarly.
2.5.3 Estimation of asymptotic variance

In the previous subsection, asymptotic normality of the AMINQU estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_v^2$ were established. In this subsection, we provide a consistent estimator of their common asymptotic variance, $V_a$, by the substitution method.

An estimator of $V_a$ by the substitution method is given as follows:

$$
\hat{V}_a = \frac{1}{\left(\sum c_i\right)^2} \sum_i c_i^2 \left[ (\bar{y}_i - \hat{\beta}_0 x_i)^2 - \frac{\sum_j (y_{ij} - \bar{y}_i)^2}{(n_i^2 - n_i)} - \hat{\sigma}^2_v \right]^2. \tag{2.36}
$$

As shown in the following theorem, $\hat{V}_a$ is consistent for $V_a$.

**Theorem 2.2** Under the assumption of Theorem 2.1, as $k \to +\infty$,

$$
k(\hat{V}_a - V_a) \to_p 0.
$$

**Proof.** Note that

$$
k(\hat{V}_a - V_a) = \frac{k}{\left(\sum c_i\right)^2} \sum_i c_i^2 \left[ (Z_i - \sigma^2_v)^2 - \text{var}(Z_i) \right]
- 2(\hat{\sigma}_v^2 - \sigma_v^2) \frac{k^2}{\left(\sum c_i\right)^2} \left[ k^{-1} \sum_i c_i^2 (Z_i - \sigma^2_v) \right] + \frac{k}{\left(\sum c_i\right)^2} \sum_i c_i^2
+ \frac{k}{\left(\sum c_i\right)^2} \sum_i c_i^2 \left\{ 4\tilde{e}_i^2((\hat{\beta}_0 - \beta)' x_i)^2 + [(\hat{\beta}_0 - \beta)' x_i]^4 
- 4Z_i\tilde{e}_i(\hat{\beta}_0 - \beta)' x_i + 2Z_i[(\hat{\beta}_0 - \beta)' x_i]^2 
+ \tilde{\sigma}_v^2 [4\tilde{e}_i(\hat{\beta}_0 - \beta)' x_i - 2((\hat{\beta}_0 - \beta)' x_i)^2] - 4\tilde{e}_i[(\hat{\beta}_0 - \beta)' x_i]^3 \right\}.
$$

In the right hand side of the last equality, the third term is clearly of the order $o_p(1)$; from (2.33), the first two terms are of the order $o_p(1)$ by applying (2.25); and from (2.18), (2.33), $C_r$-inequality and Lemma 2.7, the fourth term is also of the order $o_p(1)$. Therefore the proof is complete. □

**Remark.** When $\sigma_v^2 > 0$, with a similar proof, we can prove the consistency of the following estimator $\hat{V}_a$ of $V_a$:

$$
\hat{V}_a = \frac{1}{\left(\sum c_i\right)^2} \sum_i c_i^2 \left[ (\bar{y}_i - \hat{\beta}_0 x_i)^2 - \frac{\sum_j (y_{ij} - \bar{y}_i)^2}{(n_i^2 - n_i)} - \hat{\sigma}^2_v \right]^2. \tag{2.37}
$$
CHAPTER 2. ESTIMATING VARIANCE COMPONENTS

Corollary 2.3 Suppose conditions (A), (B) and (C) hold. If \( \sigma_v^2 > 0 \), then we have

\[
k(\hat{V}_a - V_a) \rightarrow_p 0.
\]

Note that \( \hat{V}_a \) and \( \hat{V}_a \) are the same when \( \hat{\sigma}_v^2 = \hat{\sigma}_v^2 \), but when the true value of \( \sigma_v^2 \) is small or close to zero, \( \hat{V}_a \) performs poorly.

Combining Theorem 2.1, Corollary 2.2 and Theorem 2.2, we have the following corollaries.

Corollary 2.4 Under the assumptions of Theorem 2.1,

\[
\hat{V}_a^{-1/2}(\hat{\sigma}_v^2 - \sigma_v^2) \rightarrow_d N(0, 1).
\]  

(2.38)

Corollary 2.5 Under the assumptions of Corollary 2.3, we have

\[
\hat{V}_a^{-1/2}(\hat{\sigma}_v^2 - \sigma_v^2) \rightarrow_d N(0, 1).
\]  

(2.39)

Theoretically, both (2.38) and (2.39) can be used as pivotals (asymptotically) for constructing confidence intervals for \( \sigma_v^2 \) when we have prior knowledge that \( \sigma_v^2 > 0 \). However, (2.38) may be preferred since \( \hat{\sigma}_v^2 \) requires no truncation. Also in the case of moderate \( k \), it is better to treat \( \hat{V}_a^{-1/2}(\hat{\sigma}_v^2 - \sigma_v^2) \) as a \( t \)-variate with \( k - 1 \) degrees of freedom. Hence, a confidence interval of level \( 1 - \alpha \) for \( \sigma_v^2 \) may be taken as

\[
(\hat{\sigma}_v^2 - t_{1-\alpha/2,k-1}\hat{V}_a^{1/2}, \hat{\sigma}_v^2 + t_{1-\alpha/2,k-1}\hat{V}_a^{1/2})
\]  

(2.40)

where \( t_{1-\alpha/2,k-1} \) is the \( 1 - \alpha/2 \) percentile of a student \( t \)-variate with \( k - 1 \) degrees of freedom.
Chapter 3

Estimating Regression Coefficients

In Chapter 2, we obtained estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_i^2$ of $\sigma_v^2$ and $\sigma_i^2$ through AMINQUE procedure. We found that for the model with replicates within each group, $\hat{\sigma}_v^2 + \hat{\sigma}_i^2/n_i$ is always positive. This allows us to construct weights

$$\hat{w}_i = \hat{\sigma}_v^2 + \hat{\sigma}_i^2/n_i, \quad \hat{W} = \text{diag}(\hat{w}_i).$$

From (1.31), the regression vector $\beta$ can then be estimated by

$$\hat{\beta}_w = (X'\hat{W}^{-1}X)^{-1}X'\hat{W}^{-1}y.$$

In Section 3.1, some notation is introduced. Some lemmas are also presented. The asymptotic normality of $\hat{\beta}_w$ is established in Section 3.2. A consistent estimator of the asymptotic variance-covariance matrix is obtained by the substitution method in Section 3.3. In Section 3.4 an alternative estimator is obtained using an alternative set of weights. A consistent estimator of the asymptotic variance-covariance matrix of this regression estimator is also provided.

Throughout the following sections, we assume that $\sigma_v^2 > 0$. This condition is required for the consistency of variance estimator $\hat{\sigma}_v^2$ as shown in Section 2.5.2. When $\sigma_v^2 = 0$, our model reduces to the model considered in Fuller and Rao (1978), Shao (1989a, 1989b), Shao and Chen (1991). They established the asymptotic normality of their WLS estimator of regression coefficients under somewhat stronger conditions on the moments of the error $e_{ij}$'s.

46
3.1 Notations and Some Preliminary Results

In preparation for the main result in Section 3.2, we introduce the following notation. Let

\[
\begin{align*}
\hat{w}_i &= \hat{\sigma}_\nu^2 + n_i^{-1} \hat{\sigma}_i^2 \\
&= (1 - d_i) \hat{\sigma}_\nu^2 + d_i \left[ n_i^{-1} (1 + \rho_{i0} n_i) \sum_j (y_{ij} - \bar{y}_i)^2 + (\bar{y}_i - \hat{\beta}_0 x_i)^2 \right], \\
\hat{w}_{ii} &= (1 - d_i) \sigma_\nu^2 + d_i \left[ n_i^{-1} (1 + \rho_{i0} n_i) \sum_j (y_{ij} - \bar{y}_i)^2 + (\bar{y}_i - \hat{\beta}_0 x_i)^2 \right], \\
\hat{u}_i &= (1 - d_i) \sigma_\nu^2 + d_i \left[ n_i^{-1} (1 + \rho_{i0} n_i) \sum_j (e_{ij} - \hat{e}_i)^2 + \hat{\epsilon}_i^2 \right],
\end{align*}
\]

and

\[
\tau_{1i} = E \hat{u}_i^{-1}, \quad \tau_{2i} = E \hat{u}_i^{-1} \hat{e}_i^2, \quad \tau_{3i} = E \hat{u}_i^{-2} \hat{e}_i^2, \quad (3.1)
\]

where \(d_i\)'s are defined in (2.10) and \(\rho_{i0} = \sigma_{i0}^2 / \sigma_{i0}^2\).

Remark. When \(\sigma_\nu^2 > \hat{\nu}\), from (2.22), we have

\[
u_i^{-1} \leq 2/\sigma_\nu^2. \quad (3.2)
\]

Hence \(\tau_{1i}, \tau_{2i}\) and \(\tau_{3i}\) all exist from assumption (2.18).

Now, some preliminary results are established for proving the main result in Section 3.2.

Lemma 3.1 Suppose that condition (2.18) holds. If \(\sigma_\nu^2 > 0\), then for any nonrandom sequence \(\{z_i\}\) satisfying \(|z_i| \leq z_\infty\) for all \(i\),

\[
k^{-1} \sum_{i=1}^k z_i \left[ \nu_i^{-m} \hat{e}_i^r - E(\nu_i^{-m} \hat{e}_i^r) \right] \to_p 0
\]

for \(r = 0, 1, 2\) and any nonnegative integer \(m\).

Proof. From (3.2), we have for \(r = 0, 1, 2\)

\[
E|z_i \nu_i^{-m} \hat{e}_i^r|^{1+\delta/2} \leq C_1 E|\hat{e}_i^r|^{1+\delta/2} \leq C
\]

for some positive constants \(C\) and \(C_1\). The result follows from Lemma 2.4. \(\square\)
Lemma 3.2 Suppose that assumptions (A), (B) and (C) hold. Assume $\sigma_o^2 > 0$. Let $\zeta_1, \zeta_2$ be two non-negative random variables, $\{\eta_i\}$ be a sequence of random variables with $E|\eta_i| \leq M$ for some positive constant $M$ for all $i$. Then for non-negative integers $m, r$ and $s$\[ k^{-3/2} \sum_i \frac{\eta_i}{[(1 - d_i)\sigma_o^2 + \zeta_1][(1 - d_i)\sigma_o^2 + \zeta_2]^m[(1 - d_i)\sigma_o^2 + \zeta_2]^r} \to_P 0, \]as $k \to +\infty$.

Proof. With $\sigma_o^2 > 0$, from Corollary 2.2, we have $\hat{\sigma}_o^2 \to_P \sigma_o^2$. Therefore\[ P(\hat{\sigma}_o^2 \leq \sigma_o^2/2) \to 0 \quad \text{as } k \to +\infty. \]

For any $\delta_1 > 0$,
\begin{align*}
P\left\{k^{-3/2} & \left| \sum_i \frac{\eta_i}{[(1 - d_i)\sigma_o^2 + \zeta_1][(1 - d_i)\sigma_o^2 + \zeta_2]^m[(1 - d_i)\sigma_o^2 + \zeta_2]^r} \right| \geq \delta_1 \right\} \\
& \leq P\{\hat{\sigma}_o^2 \leq \sigma_o^2/2\} \\
& \quad + P\left\{k^{-3/2} \sum_i \frac{|\eta_i|}{[(1 - d_i)\sigma_o^2/2][(1 - d_i)\sigma_o^2]^m} \geq \delta_1, \quad \hat{\sigma}_o^2 > \sigma_o^2/2 \right\} \\
& \leq P\{\hat{\sigma}_o^2 \leq \sigma_o^2/2\} + P\{Ck^{-3/2} \sum_i |\eta_i| \geq \delta_1\} \\
& \leq P\{\hat{\sigma}_o^2 \leq \sigma_o^2/2\} + Ck^{-3/2} \sum_i E|\eta_i|/\delta_1 \quad \text{(by Lemma 2.3)} \\
& \leq Ck^{-1/2} \to 0
\end{align*}
as $k \to +\infty$, where $C$ and $C_1$ are some positive constants. The proof is complete. \hfill \Box

Lemma 3.3 Suppose that assumptions (A), (B) and (C) hold. If $\sigma_o^2 > 0$, then for any non-random sequence $\{z_i\}$ satisfying $|z_i| \leq z_\infty$\[ k^{-1} \sum_{i=1}^k z_i w_i^{-1} \hat{\omega}_i^{-1} e_i = o_p(1). \]

Proof. From the definition of $\hat{\omega}_i$, $u_i$ and $w_i$,
\begin{align*}
\hat{\omega}_i - w_i &= (1 - d_i)(\sigma^2 - \sigma_o^2), \\
u_i - w_i &= d_i \left(2\hat{\epsilon}_i^2 - (\hat{\beta}_0 - \beta \gamma) x_i \right) (\hat{\beta}_0 - \beta \gamma) x_i.
\end{align*}
CHAPTER 3. ESTIMATING REGRESSION COEFFICIENTS

Note that for any non-zero constants \( u_1, u_2 \),

\[
  u_2^{-1} = u_1^{-1} - u_1^{-1}u_2^{-1}(u_1 - u_2). \tag{3.5}
\]

and

\[
  u_2^{-2} = u_1^{-2} + (u_1^{-2}u_2^{-1} + u_1^{-1}u_2^{-2})(u_1 - u_2).
\]

Using the last four equalities

\[
k^{-1} \sum_{i=1}^{k} z_i w_{ii}^{-1} \hat{\varepsilon}_i = \begin{align*}
  &k^{-1} \sum_{i=1}^{k} z_i w_{ii}^{-2} \hat{\varepsilon}_i - k^{-1} \sum_{i=1}^{k} z_i w_{1i}^{-1} w_{ii}^{-1} (\hat{\omega}_i - \omega_i) \hat{\varepsilon}_i \\
  &k^{-1} \sum_{i=1}^{k} z_i u_i^{-2} \hat{\varepsilon}_i + k^{-1} \sum_{i=1}^{k} z_i (u_i^{-2} w_{ii}^{-1} + u_i^{-1} w_{ii}^{-2}) \hat{\varepsilon}_i \\
  &\left(2 \hat{\varepsilon}_i - (\hat{\beta}_0 - \beta)' x_i \right) \\
  &- k^{1/2}(\hat{\sigma}_v^2 - \sigma_v^2) k^{-3/2} \sum_{i=1}^{k} z_i (1 - d_i) w_{ii}^{-2} \hat{\omega}_i^{-1} \hat{\varepsilon}_i.
\end{align*}
\]

In the right hand side of the last equality, the first term is \( o_p(1) \) from Lemma 3.1 since \( E u_i^{-2} \hat{\varepsilon}_i = 0 \) from (2.20); the second term is \( o_p(1) \) by Lemma 2.7 and Lemma 3.2; the third term is also \( o_p(1) \) by Corollary 2.2 and Lemma 3.2. The proof is complete. \( \Box \)

**Lemma 3.4** Suppose that assumptions (A), (B) and (C) hold. Assume \( \sigma_v^2 > 0 \). Let

\[
  H = \text{diag}_i(z_i \hat{\varepsilon}_i^{-1}),
\]

\[
  G = \text{diag}_i(z_i \tau_{ii}),
\]

where \( z_i \)'s satisfy \( z_0 \leq z_i \leq z_\infty < +\infty \). Then

\[
k^{-1} X' H X - k^{-1} X' G X \to_p 0.
\]

**Proof.** The \((t, s)\)-elements of \( k^{-1} X' H X \) and \( k^{-1} X' G X \) are respectively

\[
k^{-1} \sum_{i=1}^{k} z_i \hat{\omega}_i^{-1} x_{it} x_{is} \quad \text{and} \quad k^{-1} \sum_{i=1}^{k} z_i \tau_{ii} x_{it} x_{is}.
\]
Using the identity (3.5) twice, their difference can be decomposed as

\[ k^{-1} \sum_{i=1}^{k} z_i \hat{\omega}_i^{-1} x_{ii} x_{is} - k^{-1} \sum_{i=1}^{k} z_i \tau_{ii} x_{ii} x_{is} \]

\[ = k^{-1} \sum_{i=1}^{k} z_i x_{ii} x_{is} (u_i^{-1} - \tau_{ii}) \]

\[ + k^{-1} \sum_{i=1}^{k} z_i x_{ii} x_{is} u_i^{-1} w_{ii}^{-1} (2 \hat{\epsilon}_i - (\hat{\beta}_0 - \beta)' x_i) (\hat{\beta}_0 - \beta)' x_i \]

\[ - (\hat{\sigma}_v^2 - \sigma_v^2) k^{-1} \sum_{i=1}^{k} (1 - d_i) z_i x_{ii} x_{is} w_{ii}^{-1} \hat{\omega}_i^{-1}. \]

These three terms on the right hand side of the last equality are of order \( o_p(1) \) from Lemma 3.1 and through arguments similar to the proof of Lemma 3.3. The proof is complete. \( \square \)

### 3.2 Asymptotic Distribution of \( \hat{\beta}_w \)

Note that \( \hat{\omega}_i \)'s are even functions of \( \epsilon_{ij} \), or it remains unchanged if \( v_i \) is replaced by \( -v_i \) and \( e_{ij} \) is replaced by \( -e_{ij} \). Hence \( \hat{\beta}_w \) is unbiased (see Kakwani, 1967). Further, we obtain the asymptotic distribution of \( \hat{\beta}_w \) in the following Theorem 3.1.

Denote

\[ \Sigma_k = X' D_3 X + 2 X' D_2 W_0^{-1} X (X' W_0^{-1} X)^{-1} X' D_0 D_3 X \]

\[ + 2 X' D_2 D_0 X (X' W_0^{-1} X)^{-1} X' W_0^{-1} D_2 X \]

\[ + 4 X' D_3 D_0 X (X' W_0^{-1} X)^{-1} X' W_0^{-1} D W_0^{-1} X (X' W_0^{-1} X)^{-1} X' D_0 D_3 X, \]

where

\[ D_0 = \text{diag}(d_i), \quad D = \text{diag}(\sigma_v^2 + n_i^{-1} \sigma_i^2), \quad D_t = \text{diag}(\tau_{ti}) \quad (t = 1, 2, 3), \]

and \( \tau_{ti} \)'s are defined in (3.1).

The following theorem establishes the asymptotic distribution of the WLS estimator \( \hat{\beta}_w \).
CHAPTER 3. ESTIMATING REGRESSION COEFFICIENTS

\textbf{Theorem 3.1} Suppose that assumptions (A), (B) and (C) hold. If $\sigma_i^2 > 0$, then

\[ V_k^{-1/2}(\hat{\beta} - \beta) \rightarrow_d N(0, I_p) \quad (3.6) \]

where $V_k^{-1/2} = (V_k^{1/2})^{-1}$ and $V_k^{1/2}$ is a square root of

\[ V_k = (X'D_1X)^{-1}\Sigma_k(X'D_1X)^{-1}. \]

\textbf{Proof. } We first show that

\[ \Sigma_k^{-1/2}T_k \rightarrow_d N(0, I_p), \quad (3.7) \]

where

\[ T_k = X'\hat{\omega}_1^-\varepsilon + 2(X'D_3D_0X)(X'W_0^{-1}X)^{-1}X'W_0^{-1}\varepsilon \]

with

\[ \hat{\omega}_1 = \text{diag}(u_i) \quad \text{and} \quad \varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)' . \]

From (2.20), $ET_k = 0$. Then

\[ \text{Var}(T_k) = E(T_kT_k') \]

\[ = E(X'\hat{\omega}_1^-\varepsilon\varepsilon'\hat{\omega}_1^-X) + 2E[X'\hat{\omega}_1^-\varepsilon\varepsilon'W_0^{-1}X(X'W_0^{-1}X)^{-1}X'D_0D_3X] \]

\[ + 2E\left[X'D_3D_0X(X'W_0^{-1}X)^{-1}X'W_0^{-1}\varepsilon\varepsilon'\hat{\omega}_1^-X\right] \]

\[ + 4E\left[X'D_3D_0X(X'W_0^{-1}X)^{-1}X'W_0^{-1}\varepsilon\varepsilon'W_0^{-1}X(X'W_0^{-1}X)^{-1}X'D_0D_3X\right] \]

\[ = \Sigma_k , \]

since

\[ \hat{\omega}_1^-\varepsilon\varepsilon'\hat{\omega}_1^- = \begin{pmatrix} \frac{\varepsilon_i\varepsilon_i'}{u_iu_i} \end{pmatrix}_{(k \times k)} , \]

and

\[ \varepsilon\varepsilon'\hat{\omega}_1^- = \begin{pmatrix} \frac{\varepsilon_i\varepsilon_i'}{u_i} \end{pmatrix}_{(k \times k)} , \]

with $(a_{it})_{(k \times k)}$ denoting a $k \times k$ matrix with $a_{it}$ as $(i, t)$-element.

Let $l$ be a fixed non-zero $p$-vector and $\lambda_k = \Sigma_k^{-1/2}l/(l'\Sigma_k^{-1}l)^{1/2}$. Then $\|\lambda_k\| = 1$, and

\[ d_k^2 = \text{var}(\lambda_k'T_k) = (l'\Sigma_k^{-1}l)^{-1}l'\Sigma_k^{-1/2}\Sigma_k\Sigma_k^{-1/2}l = (l'\Sigma_k^{-1}l)^{-1}l'l . \]
CHAPTER 3. ESTIMATING REGRESSION COEFFICIENTS

Let \( \lambda_k \) and \( \eta_k \) be the \( s \)-th element of \( \lambda_k \) and \( \eta_k = 2(X' W_0^{-1} X)^{-1} X' D_0 D_3 X \lambda_k \), respectively. Then

\[
\lambda_k^* T_k = \sum_{i=1}^{k} \sum_{s=1}^{p} \left[ x_{is} \lambda_k u_i^{-1} \varepsilon_i + x_{is} \eta_k w_i^{-1} \varepsilon_i \right].
\]

From assumption (A) and (2.22) (due to choice of \( \sigma_0^2 \), \( \sigma_v^2 \))

\[
\| \eta_k \|^2 = 4 \lambda_k^* X' D_3 D_0 X (X' W_0^{-1} X)^{-2} X' D_0 D_3 X \lambda_k \leq C_1
\]

where \( C_1 \) is some positive constant, and

\[
E| x_{is} \lambda_k u_i^{-1} \varepsilon_i + x_{is} \eta_k w_i^{-1} \varepsilon_i |^{2+\delta} \leq C_2 E \left\{ \left[ (1 - d_i)^{-1} \sigma_v^{-2} + 1 \right] \| \varepsilon_i \|^{2+\delta} \right\} \leq C_3,
\]

for some positive constants \( C_2 \) and \( C_3 \). Hence

\[
d_k^{-(2+\delta)} \sum_{i} \sum_{s} E| x_{is} \lambda_k u_i^{-1} \varepsilon_i + x_{is} \eta_k w_i^{-1} \varepsilon_i |^{2+\delta} \leq C_3 p k d_k^{-(2+\delta)}.
\]

From assumptions (A), (B) and (C), there are positive constants \( C_0^*, C_1^* \) such that

\[
C_0^* I_p \leq k^{-1} \Sigma_k \leq C_1^* I_p. \tag{3.8}
\]

Hence

\[
k d_k^{-(2+\delta)} \leq C_0^{-(1-\delta/2)} k^{-\delta/2} \to 0,
\]

and the Linderberg's condition holds. Therefore

\[
d_k^{-(2+\delta)} \lambda_k^* T_k = l' \Sigma_k^{-1/2} T_k / (l'L)^{1/2} \to_d N(0, 1).
\]

Since \( l \) is arbitrary, the proof of (3.7) is complete.

Next, we show that

\[
X' \hat{W}^{-1} \varepsilon - T_k = o_p(k^{1/2}). \tag{3.9}
\]

The \( t \)-th element of \( X' \hat{W}^{-1} \varepsilon \) is \( \sum_{i=1}^{k} x_{it} \hat{w}_i^{-1} \varepsilon_i \). By using (3.5) repeatedly

\[
\sum_{i=1}^{k} x_{it} \hat{w}_i^{-1} \varepsilon_i
\]

\[
= \sum_{i=1}^{k} x_{it} u_i \varepsilon_i + 2 \sum_{i=1}^{k} d_t \tau_{3i} x_{it} (\hat{\beta}_0 - \beta)' x_i + 2 \sum_{i=1}^{k} d_t x_{it} (u^{-2} \varepsilon_i^2 - \tau_{3i}) (\hat{\beta}_0 - \beta)' x_i
\]

\[
- \sum_{i=1}^{k} d_t u_i^{-1} w_{ii}^{-1} x_{it} \varepsilon_i ((\hat{\beta}_0 - \beta)' x_i)^2 - \sum_{i=1}^{k} (1 - d_i) x_{it} w_{ii}^{-1} \hat{w}_i^{-1} \varepsilon_i (\delta_v^2 - \sigma_v^2)
\]

\[
+ 2 \sum_{i=1}^{k} d_t^2 x_{it} u_i^{-2} w_{ii}^{-1} \varepsilon_i^2 ((\hat{\beta}_0 - \beta)' x_i)^2 (2 \varepsilon_i - (\hat{\beta}_0 - \beta)' x_i).
\]
CHAPTER 3. ESTIMATING REGRESSION COEFFICIENTS

On the right hand side of the last equality, the first two terms are exactly the \( t \)-th element of \( T_k \); the third, fourth, and fifth terms are of order \( o_p(k^{1/2}) \) from Lemma 2.7 (note: for the fifth term \( |u_i^{-2} \varepsilon_i^3| \leq |d_i^{-1} u_i^{-1} \varepsilon_i| \)); the last term is of the order \( o_p(k^{1/2}) \) from Lemma 3.3 and \( k^{1/2}(\hat{\sigma}_v^2 - \sigma_v^2) = O_p(1) \). Hence (3.9) is proved.

Finally, from (3.7) and (3.9)

\[
\Sigma_k^{-1/2} X' \hat{W}^{-1} \varepsilon \rightarrow_d N(0, I_p). \tag{3.10}
\]

Since \( \hat{\beta}_v - \beta = (X' \hat{W}^{-1} X)^{-1} X' \hat{W}^{-1} \varepsilon \),

\[
\Sigma_k^{-1/2}(X'D_1 X)(\hat{\beta}_v - \beta) = \Sigma_k^{-1/2}(X'D_1 X)(X' \hat{W}^{-1} X)^{-1} \Sigma_k^{1/2} \Sigma_k^{-1/2} X' \hat{W}^{-1} \varepsilon. \tag{3.11}
\]

Note that

\[
\Sigma_k^{-1/2}(X'D_1 X)(X' \hat{W}^{-1} X)^{-1} \Sigma_k^{1/2} = \Sigma_k^{-1/2}(X'D_1 X) \left[(X' \hat{W}^{-1} X)^{-1} - (X'D_1 X)^{-1}\right] \Sigma_k^{1/2} + I_p.
\]

From Lemma 3.4, \( k[(X' \hat{W}^{-1} X)^{-1} - (X'D_1 X)^{-1}] \rightarrow_p 0 \). From (3.8) and the fact \( X'D_1 X \leq 2\sigma_v^{-2} X'X \),

\[
\Sigma_k^{-1/2}(X'D_1 X)(X' \hat{W}^{-1} X)^{-1} \Sigma_k^{1/2} \rightarrow I_p.
\]

Thus from (3.10) and (3.11)

\[
\Sigma_k^{-1/2}(X'D_1 X)(\hat{\beta}_v - \beta) \rightarrow_d N(0, I_p). \tag{3.12}
\]

We have shown (3.6) when \((X'D_1 X)^{-1} \Sigma_k^{1/2}\) is taken as a square root of \(V_k\). For any arbitrary square root \(V_k^{1/2}\), (3.6) follows from (3.12) by using the same argument as in Fahrmeir and Kaufmann (1985, p349). This completes the proof of the theorem.

\( \square \)

In practice, we need to estimate some function of \( \beta \), say, \( \theta = g(\beta) \), where \( g \) is a function on \( R^p \). We use the natural estimator

\[
\hat{\theta} = g(\hat{\beta}_v).
\]

Then we have the following results about \( \hat{\theta} \).
Theorem 3.2 Assume \( g \) is a continuously differentiable function from \( \mathbb{R}^p \to \mathbb{R}^q \) and \( \theta = g(\beta) \). Under conditions in Theorem 3.1,
\[
V_\theta^{-1/2}(\hat{\theta} - \theta) \to N(0, I_q),
\]
where
\[
V_\theta = \Delta g(\beta) V_\delta [\Delta g(\beta)]',
\]
with \( \Delta g(\beta) \) denoting the gradient of \( g \) at \( \beta \).

The proof of Theorem 3.2 follows standard arguments and is omitted.

3.3 Substitution Estimator of \( V_k \)

From the structure of \( V_k \), we can estimate it piece by piece. We get an estimator of \( X' D_1 X \), and an estimator of \( \Sigma_k \) (also piece by piece). Thus an estimator of \( V_k \) is given by
\[
\hat{V}_k = (X' \hat{W}^{-1} X)^{-1} \hat{\Sigma}_k (X' \hat{W}^{-1} X)^{-1}
\]
with
\[
\hat{\Sigma}_k = \hat{X}' \hat{D}_3 X + 2X' \hat{D}_2 W_0^{-1} X (X' W_0^{-1} X)^{-1} X' D_0 \hat{D}_3 X
+ 2X' \hat{D}_3 D_0 X (X' W_0^{-1} X)^{-1} X' W_0^{-1} \hat{D}_2 X
+ 4X' \hat{D}_3 D_0 X (X' W_0^{-1} X)^{-1} X' W_0^{-1} \hat{D} W_0^{-1} X (X' W_0^{-1} X)^{-1} X' D_0 \hat{D}_3 X,
\]
and
\[
\hat{D} = \text{diag}_i \left((\tilde{y}_i - \hat{\beta}_0^i x_i)^2\right),
\hat{D}_2 = \text{diag}_i \left(\hat{w}_i^{-1}(\tilde{y}_i - \hat{\beta}_0^i x_i)^2\right),
\hat{D}_3 = \text{diag}_i \left(\hat{w}_i^{-2}(\tilde{y}_i - \hat{\beta}_0^i x_i)^2\right).
\]

As shown in the following theorem, \( \hat{V}_k \) is consistent.

Theorem 3.3 Under the assumptions in Theorem 3.1, we have
\[
k(\hat{V}_k - V_k) \to_p 0.
\]
CHAPTER 3. ESTIMATING REGRESSION COEFFICIENTS

Proof. We prove first that

\[ k^{-1} (\hat{\Sigma}_k - \Sigma_k) \rightarrow_p 0. \]  

(3.13)

Analogous to Lemma 3.4, by using Lemma 2.7 and Lemma 3.3, we can show

\[ k^{-1} (X' \hat{D}_3 X - X' D_3 X) \rightarrow_p 0, \]

\[ k^{-1} (X' W_0^{-1} \hat{D}_2 X - X' W_0^{-1} D_2 X) \rightarrow_p 0, \]

\[ k^{-1} (X' D_0 \hat{D}_3 X - X' D_0 D_3 X) \rightarrow_p 0, \]

\[ k^{-1} (X' W_0^{-1} \hat{D} W_0^{-1} X - X' W_0^{-1} D W_0^{-1} X) \rightarrow_p 0. \]

(3.14)

(3.13) follows immediately (Note \( k(X' W_0^{-1} X)^{-1} = O_p(1) \)). Moreover

\[
k(\hat{V}_k - V_k)
= k[(X' \hat{W}^{-1} X)^{-1} - (X' D_1 X)^{-1}] \hat{\Sigma}_k [(X' \hat{W}^{-1} X)^{-1} - (X' D_1 X)^{-1}]
+ k[(X' \hat{W}^{-1} X)^{-1} - (X' D_1 X)^{-1}] \hat{\Sigma} (X' D_1 X)^{-1}
+ k(X' D_1 X)^{-1} \hat{\Sigma}_k [(X' \hat{W}^{-1} X)^{-1} - (X' D_1 X)^{-1}]
+ k(X' D_1 X)^{-1} (\hat{\Sigma}_k - \Sigma_k) (X' D_1 X)^{-1}.
\]

From Lemma 3.4, \( k[(X' \hat{W}^{-1} X)^{-1} - (X' D_1 X)^{-1}] = o_p(1) \). From (3.13) and (3.8), \( k^{-1} \hat{\Sigma}_k = O_p(1) \). Clearly \( k(X' D_1 X)^{-1} = O_p(1) \). This completes the proof.

Similarly for the parameter \( \theta \) in Section 3.2. The following variance estimator \( \hat{V}_\theta \) is a consistent estimator of \( V_\theta \).

**Theorem 3.4** Suppose the conditions in Theorem 3.2 hold. Then

\[ k(\hat{V}_\theta - V_\theta) \rightarrow_p 0. \]

where

\[ \hat{V}_\theta = \Delta g(\hat{\theta}_\omega) \hat{V}_k [\Delta g(\hat{\theta}_\omega)]'. \]

The proof of this theorem is simple and is omitted.

Combining the results in Theorem 3.1 and Theorem 3.3, we have the following corollary.
Corollary 3.1 Under the assumptions in Theorem 3.1, we have
\[ \hat{V}_k^{-1/2}(\hat{\beta}_w - \beta) \rightarrow_d N(0, I_p) \] (3.15)
as \( k \rightarrow \infty \).

Similarly, from Theorem 3.2 and Theorem 3.4, we have the following corollary.

Corollary 3.2 Under the assumptions in Theorem 3.2, we have
\[ \hat{V}_\theta^{-1/2}(\hat{\theta} - \theta) \rightarrow_d N(0, I_q) \]
as \( k \rightarrow \infty \).

Using the result in Corollary 3.1 we can construct a simultaneous confidence region for the regression coefficients in the form
\[ \{ \beta : (\hat{\beta}_w - \beta)'\hat{V}_k^{-1}(\hat{\beta}_w - \beta) \leq C \} \] (3.16)
where \( C \) is some constant corresponding to the required confidence level. Similarly we can obtain a confidence region for \( \theta = g(\beta) \) from Corollary 3.2, in particular, a confidence interval on \( \theta = \beta^t \), the \( t \)-th component of \( \beta \). The WLS estimator is \( \hat{\theta} = \hat{\beta}_w^t \), the \( t \)-th element of \( \hat{\beta}_w \). Its asymptotic variance is \( \hat{V}_\theta = \hat{V}_{k,tt} \), the \((t,t)\)-element of \( \hat{V}_k \). Hence we can use the interval
\[ (\hat{\theta} - t_{1-\alpha/2,k-1}\hat{V}_\theta^{1/2}, \hat{\theta} + t_{1-\alpha/2,k-1}\hat{V}_\theta^{1/2}) \] (3.17)
as a \( 1 - \alpha \) confidence interval of \( \theta \) asymptotically, where \( t_{1-\alpha/2,k-1} \) is the \( 1 - \alpha/2 \) percentile of a student \( t \)-variate with \( k - 1 \) degrees of freedom.

3.4 Another Estimator of \( \beta \)

Using the idea behind the ASR-SSE type variance estimator (see P. S. R. S. Rao, Kaplan, and Cochran, 1981), we propose another set of estimators of weights \( w_i \). From the proof of Theorem 3.1, we see that we can obtain a consistent estimator of \( \beta \) with any invariant estimator of \( \sigma^2 \). One classical estimator of \( \sigma^2 \) is \( s^2 \) =
CHAPTER 3. ESTIMATING REGRESSION COEFFICIENTS

\((n_i - 1)^{-1} \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2\) when \(n_i \geq 2\) for all \(i\). With \(\hat{w}_{si} = \hat{\sigma}_i^2 + s_i^2/n_i\) as weight we can compute the weighted least squares estimator of \(\beta\).

If \(n_i\) is large, \(s_i^2 \approx \hat{\sigma}_i^2\) in probability. But with small \(n_i\) (2 \(\leq n_i \leq 5\)), \(\hat{\sigma}_i^2\) is a weighted average of \(s_i^2\) and adjusted sum of squares of residual \(\sum_{j=1}^{n_i} (y_{ij} - x'_i \hat{\beta})^2 - \hat{\sigma}_i^2\). Another reason for using \(s_i^2\) is to guarantee positive weights for the WLS estimator of \(\beta\). Since \(s_i^2\) is always positive, \(\hat{\sigma}_i^2 + s_i^2/n_i\) is always positive. Using \(s_i^2\), as will be seen in Chapter 5, is helpful for the two-fold nested-error regression model in Chapter 5.

When the true \(\sigma_\epsilon^2\) is small or close to 0, \(\hat{\sigma}_i^2\) is close to 0. Our model is then close to model (1.2). In this case this WLS estimator is thus not efficient with small \(n_i\), as shown empirically in the literature. A simulation result confirmed this conclusion.

The alternative estimator of \(\beta\) can be obtained from (1.31) as

\[ \hat{\beta}_{sw} = (X' \hat{W}_s^{-1} X)^{-1} X' \hat{W}_s^{-1} y \]

with

\[ \hat{W}_s = \text{diag}(\hat{w}_{si}). \]

Denote

\[ u_{si} = \sigma_\epsilon^2 + n_i^{-1} s_i^2, \]
\[ \tau_{si} = E(u_{si}^{-1}), \]
\[ \tau_{s_i} = E(u_{si}^{-2} \hat{\epsilon}_i^2), \]
\[ D_4 = \text{diag}(\tau_{si}), \]
\[ D_5 = \text{diag}(\tau_{s_i}). \]

Clearly \(\tau_{si}\) and \(\tau_{s_i}\) are finite since \(u_{si}^{-1} \leq \sigma_\epsilon^{-2}\) given \(\sigma_\epsilon^2 > 0\).

Analogous to Lemma 3.1–Lemma 3.3, we have following results.

**Lemma 3.5** Suppose that conditions (A), (B) and (C) hold. If \(\sigma_\epsilon^2 > 0\), then

(i) For any non-random sequence \(\{z_i\}\) such that \(|z_i| \leq z_\infty\) for all \(i\), as \(k \to +\infty\)

\[ k^{-1} \sum_{i=1}^{k} z_i \left[u_{si}^{-m} \hat{\epsilon}_i^m - E(u_{si}^{-m} \hat{\epsilon}_i^m)\right] \to_p 0 \]
for $r = 0, 1, 2$ and any nonnegative integer $m$.

(ii) Let $\{\eta_i\}$ be a sequence of random variable with $E|\eta_i| \leq M$ for some positive constant $M$ for all $i$. Then for non-negative integers $m$, $r$ as $k \to +\infty$,

$$k^{-3/2} \sum_i \hat{w}^{-r}_{si} u^{-m}_{si} \eta_i \to_p 0.$$

(iii)

$$k^{-1}(X'\hat{W}^{-1}_{si}X - X'D_4X) \to_p 0 \quad \text{as} \quad k \to +\infty. \quad \Box$$

The proofs of these results are parallel to the proofs of Lemma 3.1–Lemma 3.3, and are omitted.

Analogous to Theorem 3.1, we have the following theorem on the asymptotic normality of $\hat{\beta}$.

**Theorem 3.5** Suppose that conditions (A), (B) and (C) hold. If $\sigma^2 > 0$, then

$$V^{-1/2}_{sk}(\hat{\beta}_{s1} - \beta) \to_d N(0, I_p) \quad \text{as} \quad k \to +\infty$$

where $V^{-1/2}_{sk} = (V^{-1/2}_{sk})^{-1}$ and $V^{1/2}_{sk}$ is a square root of

$$V_{sk} = (X'D_4X)^{-1}X'D_5X(X'D_4X)^{-1}.$$ 

**Proof.** Let

$$\hat{W}_{s1} = \text{diag}(u_{s1}).$$

Then

$$\hat{W}^{-1}_{s1} = \text{col}_i (u^{-1}_{si} \xi_i),$$

$$\hat{W}^{-1}_{s1} \xi \hat{W}^{-1}_{s1} = \left( \begin{array}{c} \xi_i \xi_i \\ u_{s1} u_{s1} \end{array} \right)_{(s, k, \xi)}.$$ 

From (2.20)

$$EX'\hat{W}^{-1}_{s1} \xi = 0.$$ 

Thus

$$\text{Var}(X'\hat{W}^{-1}_{s1} \xi) = E[X'\hat{W}^{-1}_{s1} \xi \hat{W}^{-1}_{s1} X] = X'D_5X.$$
Along the lines of the proof of (3.7) and (3.9), we can prove similarly by using Lemma 3.5 that
\[(X' D_5 X)^{-1/2} X' \hat{W}_s^{-1} \epsilon \rightarrow_d N(0, I_p),\]
\[X' \hat{W}_s^{-1} \epsilon - X' \hat{W}_{s1}^{-1} \epsilon = o_p(k^{1/2}).\]

Hence
\[(X' D_5 X)^{-1/2} X' \hat{W}_s^{-1} \epsilon \rightarrow_d N(0, I_p) \quad \text{as} \quad k \rightarrow +\infty.\]

Then, similar to the proof of (3.12),
\[(X' D_5 X)^{-1/2} (X' D_4 X)(\hat{\beta}_{sw} - \beta) \rightarrow_d N(0, I_p) \quad \text{as} \quad k \rightarrow +\infty.\]

Similar argument to the last paragraph of the proof of Theorem 3.1 completes this proof. The details are omitted. \(\Box\)

Similar to the estimator of \(V_k\) in section 3.3, we have a consistent estimator of \(V_{sk}\) as
\[\hat{V}_{sk} = (X' \hat{W}_s^{-1} X)^{-1} (X' \hat{D}_5 X) (X' \hat{W}_s^{-1} X)^{-1}\]

with
\[\hat{D}_5 = \text{diag}_i \left( \hat{w}_{si}^{-2}(\hat{y}_i - \hat{\beta}_0 x_i)^2 \right).\]

**Theorem 3.6** Suppose conditions (A), (B) and (C) hold. Then

\[k(\hat{V}_{sk} - V_{sk}) \rightarrow_p 0 \quad \text{as} \quad k \rightarrow +\infty.\]

The proof of the theorem is similar to that of Theorem 3.3, and is omitted for simplicity. \(\Box\)

In the same fashion as those results and comments after Theorem 3.3, the confidence intervals associated with \(\hat{\beta}_{sw}\) and variance estimator \(\hat{V}_{sk}\) can also be constructed. The details are omitted for simplicity.
Chapter 4

Jackknife Estimators

In model (1.1), the observations within a group are correlated. The usual delete-1 jackknife variance estimators, based on the assumption of independent observations, is therefore inconsistent. On the other hand, deleting all observations in the same group at a time only makes use of between group independence and maintains the within-group structure. Therefore, the delete-group jackknife technique is applied to our model in this chapter.

The following sections are organized as follows. In Section 4.1, we give a brief review of jackknife method and its application to special cases of our model (1.1).

In Section 4.2, jackknife estimation of $\sigma^2$ is discussed. Jackknifing the untruncated AMINQUE estimator $\hat{\sigma}^2$ is proposed. A jackknife variance estimator of $\hat{\sigma}^2$ (and $\hat{\sigma}^2_0$) is also constructed and its consistency is established. The jackknife estimator and variance estimator are used to construct a pivotal for the confidence interval on $\sigma^2$. We can also obtain a jackknife estimator of $\sigma^2_0$ by jackknifing $\hat{\sigma}^2_0$, and corresponding variance estimators. They are consistent if we have the prior knowledge $\sigma^2_0 > 0$.

In Section 4.3, the jackknife method is applied to the estimation of regression parameters. A procedure similar to Miller (1974b) is used to construct a jackknife estimator of the covariance matrix of $\hat{\beta}_w$. The consistency of this jackknife estimator
is established. The jackknife estimator of covariance matrix is used to construct a confidence region for the regression parameter $\beta$.

In Section 4.4, we outline the jackknife method to construct a consistent estimator of $V_{nk}$, the asymptotic variance–covariance matrix of the estimator of regression parameters in Section 3.4. Results similar to those in Section 4.3 are obtained.

Finally, in Section 4.5, Monte Carlo simulation results are presented on the proposed jackknife variance estimators and associated confidence intervals on the regression parameters.

### 4.1 Review of Jackknife Methodology

The method of jackknife was reviewed in Miller (1974a). Its idea was first proposed in Quenouille (1949) and later extended in Quenouille (1956) as a method for reducing the bias of an estimator. Suppose $Y_1, Y_2, \ldots, Y_n$ is a sequence of independent identically distributed random variables and $T_n(Y_1, Y_2, \ldots, Y_n)$ is an estimator of a parameter of interest, $\theta$. If $T_n$ has the following property

$$E[T_n(Y_1, Y_2, \ldots, Y_n)] = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots,$$

then pseudo–values $Q_i = nT_n - (n - 1)T_{n-1,i}(Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)$ would be biased to order $1/n^2$ only. A symmetric estimator is then constructed by averaging $Q_i$ over $i = 1, 2, \ldots, n$:

$$T_{nJ} = \frac{1}{n} \sum_{i=1}^{n} Q_i = nT_n - \frac{n-1}{n} \sum_{i=1}^{n} T_{n-1,i}.$$

Tukey (1958) proposed that the $Q_i$'s may be considered as an approximately i.i.d. random sample (they are approximately unbiased for $\theta$), and use

$$v_j = \frac{1}{n(n-1)} \sum_{i=1}^{n} (Q_i - \bar{Q})^2$$

as the variance estimator of $T_{nJ}$. Further, he suggested the use of

$$t^* = \frac{T_{nJ} - \theta}{\sqrt{v_j}}$$
as a pivotal quantity for obtaining a robust confidence interval on $\theta$, by treating it as a $t$-variate with $n - 1$ degrees of freedom. Miller (1964) and Miller (1968) present situations where Tukey's method is asymptotically valid.

For non-iid case of $Y_1, Y_2, \ldots, Y_n$, some modified jackknife methods are also used. Arvesen (1969) proved the consistency of jackknifing U-statistics when $Y_1, Y_2, \ldots, Y_n$ are independent (but not necessarily identically distributed) and have a symmetric kernel $f^*(Y_{k_1}, \ldots, Y_{k_m})$ such that $E[f^*(Y_{k_1}, \ldots, Y_{k_m})] = \eta$ for all $k_1, \ldots, k_m$. Further he proposed a jackknife scheme for jackknifing variance components in a balanced one-way random effect model, $y_{ij} = \mu + v_i + e_{ij}$. Arvesen and Layard (1975) extended this result to the unbalanced one-way random effect model. Their jackknife methods are actually equivalent to the delete-group jackknife method which will be discussed later.

Miller (1974b), Hinkley (1977) and Wu (1986) considered the classic regression model (1.2). The parameter of interest is $\theta = f(\beta)$, where $f$ is a real-valued function. Let $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$ be the ordinary least square (OLS) estimator of $\beta$, and $\tilde{\beta}_{OLS}^{(i)}$ be the OLS estimator after the observation $(y_i, x_i)$ is removed. Denote $\hat{\theta} = f(\hat{\beta}_{OLS})$ and $\tilde{\theta}_i = f(\tilde{\beta}_{OLS}^{(i)})$. Then the pseudo-values are defined as

$$Q_i = k\hat{\theta} - (k - 1)\tilde{\theta}_i.$$

A jackknife estimator of $\theta$ is

$$\hat{\theta}_J = \frac{1}{k} \sum_{i=1}^{k} \tilde{\theta}_i,$$

and a jackknife variance estimator is defined as

$$v_J = \frac{k - 1}{k} \sum_{i=1}^{k} (\tilde{\theta}_i - \hat{\theta}_J)^2.$$

Miller (1974b) established consistency of the variance estimator and the asymptotic normality of the jackknife estimator $\hat{\theta}_J$.

Hinkley (1977) and Wu (1986) proposed weighted jackknife methods for the regression model.

For model (1.3) with $x'_i\beta = \mu$, Rao (1980) considered a delete-group (i.e., the observations in one group are all removed from estimation at one time) jackknife
estimator for the common mean \( \mu \). An extensive Monte Carlo simulation study under a variety of settings showed that the jackknife variance estimator performs pretty well for most cases.

For model (1.3), Shao (1989b) proved that the usual delete-one jackknife variance estimator is inconsistent. He also proposed a consistent variance estimator by obtaining a jackknife estimator for each term of the asymptotic covariance matrix. Under the same model, without assuming independence within groups, Shao and Rao (1993) constructed a weighted delete-group jackknife variance estimator of both OLS and WLS estimator of \( \theta = f(\beta) \) for some real-valued function \( f \). They proved that these jackknife variance estimators are consistent.

### 4.2 Jackknife Estimators and Variance Estimators for \( \sigma_v^2 \)

In Section 2.4 we obtained the estimator \( \hat{\sigma}_v^2 \) of \( \sigma_v^2 \) through the approximate MINQUE approach which, from (2.28), has the form

\[
\hat{\sigma}_v^2 = \hat{\sigma}_{vp}^2 - 2R_{1k} + R_{2k}.
\]

We proved the asymptotic normality of \( \hat{\sigma}_v^2 \) in Theorem 2.1, and so obtained the asymptotic variance \( V_\theta \) of \( \hat{\sigma}_v^2 \). In Section 2.5.3 we obtained a consistent estimator \( \hat{V}_\theta \) of \( V_\theta \) by the substitution method. We know that in small samples, AMINQUE estimators \( \hat{\sigma}_v^2 \) (and \( \hat{\sigma}_0^2 \)) are biased. In this section, we consider jackknife estimators of \( \sigma_v^2 \), hoping it would reduce the bias. Also, we are going to obtain jackknife variance estimators for \( \hat{\sigma}_v^2 \). First we consider jackknifing \( \hat{\sigma}_v^2 \) by deleting one group at a time. The procedure is as follows.

Let \( \hat{\sigma}_{v(-t)}^2 \) denote the estimator of \( \sigma_v^2 \) through the procedure in Section 2.4 with observations in the \( t \)-th group (or at design point \( x_t \)) removed. Analogous to (2.28), \( \hat{\sigma}_{v(-t)}^2 \) can be written as

\[
\hat{\sigma}_{v(-t)}^2 = \hat{\sigma}_{vp(-t)}^2 - 2R_{1k(-t)} + R_{2k(-t)}
\]
with
\[
\hat{\sigma}^2_{v(-t)} = \left( \sum_{i \neq t} c_i \right)^{-1} \sum_{i \neq t} c_i Z_i,
\]
\[
R_{1k(-t)} = \left( \sum_{i \neq t} c_i \right)^{-1} \sum_{i \neq t} c_i \xi_i (\hat{\beta}_{0(-t)} - \beta)' x_i,
\]
\[
R_{2k(-t)} = \left( \sum_{i \neq t} c_i \right)^{-1} \sum_{i \neq t} c_i \left[ (\hat{\beta}_{0(-t)} - \beta)' x_i \right]^2,
\]
where
\[
\hat{\beta}_{0(-t)} = \left( X'_{(-t)} W_{0(-t)}^{-1} X_{(-t)} \right)^{-1} X'_{(-t)} W_{0(-t)}^{-1} \bar{y}_{(-t)},
\]  
(4.1)

$X_{(-t)}$ is the design matrix without design point $x_t$, $W_{0(-t)} = \text{diag}(w_{t0}, \ldots, w_{(t-1)0}, w_{(t+1)0}, \ldots, w_{k0})$, and $\bar{y}_{(-t)} = (\bar{y}_1, \ldots, \bar{y}_{t-1}, \bar{y}_{t+1}, \ldots, \bar{y}_k)'$.

The pseudo-values are defined as
\[
Q_t = k \hat{\sigma}^2_v - (k - 1) \hat{\sigma}^2_{v(-t)}.
\]

A jackknife estimator of $\sigma^2_v$ is then given as
\[
\hat{Q} = k^{-1} \sum_{t=1}^k Q_t = \hat{\sigma}^2_{v(J)}.
\]

And a jackknife estimator of variance of $\hat{\sigma}^2_{v(J)}$ or $\hat{\sigma}^2$ is given as
\[
\hat{V}_{jt} = (k(k - 1))^{-1} \sum_{t=1}^k (Q_t - \hat{Q})^2 = k^{-1}(k - 1) \sum_{t=1}^k \left[ \hat{\sigma}^2_{v(-t)} - k^{-1} \sum_t \hat{\sigma}^2_{v(-t)} \right]^2.
\]  
(4.2)

We shall establish the asymptotic normality of $\hat{\sigma}^2_{v(J)}$ and consistency of $\hat{V}_{jt}$ in this section. First the following lemmas are needed for proving these results.

**Lemma 4.1** Let $\{\xi_i\}$ be a sequence of random variables with $E|\xi_i|^{1+\delta} < M$ for all $i$ for some $\delta > 0$ and $M > 0$. Then
\[
k^{-1} \max_i |\xi_i| \rightarrow_p 0 \quad \text{as} \quad k \rightarrow +\infty.
\]

**Proof.** For any $\delta_1 > 0$,
\[
P\{k^{-1} \max_i |\xi_i| \geq \delta_1 \} \leq \sum_i P\{k^{-1}|\xi_i| \geq \delta_1 \}
\leq \sum_i k^{-(1+\delta)} E|\xi_i|^{1+\delta} / \delta_1^{1+\delta}
\leq Ck^{-\delta} \rightarrow 0 \quad \text{as} \quad k \rightarrow \infty.
where $C$ is some positive constant. □

In the following lemmas, for any vector $\xi = (\xi_1(t), \ldots, \xi_p(t))'$, we denote
\[ \max_t |\xi| = (\max_t |\xi_1|, \ldots, \max_t |\xi_p|)' . \]

**Lemma 4.2** Suppose that assumptions (A), (B) and (C) hold. Then
\[ \max_t |\hat{\beta}_0 - \hat{\beta}_{0(-t)}| = o_p(k^{-1/2}) . \] (4.3)

**Proof.** From a standard matrix result,
\[ (X'_{(-t)} W_{0(-t)}^{-1} X'_{(-t)})^{-1} = (X'W_0^{-1} X - w_{00}^{-1} x_t x'_t)^{-1} \]
\[ = (X'W_0^{-1} X)^{-1} + w_{00}^{-1} h_{0t}^{-1} (X'W_0^{-1} X)^{-1} x_t x'_t (X'W_0^{-1} X)^{-1} \] (4.4)

with
\[ h_{0t} = 1 - w_{00}^{-1} x'_t (X'W_0^{-1} X)^{-1} x_t . \] (4.5)

Applying identity (4.4) in (4.1), we get
\[ \hat{\beta}_{0(-t)} = (X'W_0^{-1} X - w_{00}^{-1} x_t x'_t)^{-1} (X'W_0^{-1} y - w_{00}^{-1} x_t \tilde{y}_t) \]
\[ = \hat{\beta}_0 - w_{00}^{-1} h_{0t}^{-1} (X'W_0^{-1} X)^{-1} x_t (\tilde{y}_t - x'_t \hat{\beta}_0) . \] (4.6)

Or
\[ \hat{\beta}_0 - \hat{\beta}_{0(-t)} = w_{00}^{-1} h_{0t}^{-1} (X'W_0^{-1} X)^{-1} x_t (\tilde{y}_t - x'_t \hat{\beta}_0) . \] (4.7)

In (4.5), from the choice of prior values and condition (A),
\[ \max_t |h_{0t}| \to 1 \text{ as } k \to +\infty . \]

Also due to the fact $k(X'W_0^{-1} X)^{-1} = O_p(1)$, (4.3) follows from
\[ k^{-1/2} \max_t |\tilde{y}_t - x'_t \hat{\beta}_0| = o_p(1) . \] (4.8)

Now
\[ |\tilde{y}_t - x'_t \hat{\beta}_0| = |\tilde{\epsilon}_t - (\hat{\beta}_0 - \beta)' x_t| \leq |\tilde{\epsilon}_t| + |(\hat{\beta}_0 - \beta)' x_t| . \]

Since $E|\tilde{\epsilon}_t|^{2+\delta}$ is bounded, from Lemma 4.1
\[ k^{-1} \max_t |\tilde{\epsilon}_t|^2 = o_p(1) , \]
or
\[ k^{-1/2} \max_t |\tilde{\varepsilon}_t| = o_p(1). \]  \hspace{1cm} (4.9)

Since \( \hat{\beta}_0 - \beta = O_p(k^{-1/2}) \) and it is independent of \( t \),
\[ k^{-1/2} \max_t |(\hat{\beta}_0 - \beta)'x_t| = o_p(1). \]  \hspace{1cm} (4.10)

Equation (4.8) follows from (4.9),(4.10). Hence the proof is complete. \( \square \)

**Remark.** From Lemma 2.6 and (4.3),
\[ \max_t |\hat{\beta}_{0(-t)} - \beta| = O_p(k^{-1/2}). \]  \hspace{1cm} (4.11)

**Lemma 4.3** Suppose that conditions (A), (B) and (C) hold. Then as \( k \to +\infty \),
\[ k \max_t |R_{1k} - R_{1k(-t)}| \to_p 0, \]  \hspace{1cm} (4.12)
\[ k \max_t |R_{2k} - R_{2k(-t)}| \to_p 0, \]  \hspace{1cm} (4.13)
\[ k^{1/2} \max_t |\hat{\sigma}^2_0 - \hat{\sigma}^2_{0(-t)}| \to_p 0. \]  \hspace{1cm} (4.14)

**Proof.** Since the proof of (4.13) is similar to that of (4.12), the proof of (4.13) is omitted. Note that
\[ k \max_t |R_{1k} - R_{1k(-t)}| \leq \max_t \left( \frac{k}{\nu - c_t} \right) \left[ \max_t \left| \sum_i c_i \tilde{\varepsilon}_i (\hat{\beta}_0 - \hat{\beta}_{0(-t)})'x_i \right| \right. \]  \hspace{1cm} (4.15)
\[ \left. + \max_t \left| c_t/\nu \sum_i c_i \tilde{\varepsilon}_i (\hat{\beta}_0 - \beta)'x_i \right| \right] + \max_t \left| c_t \tilde{\varepsilon}_i (\hat{\beta}_{0(-t)} - \beta)'x_i \right|, \]
where \( \nu = \sum_{i=1}^k c_i \). Since \( \max_t k/(\nu - c_t) \leq C \) for some positive constant \( C \), we need only consider the three terms in the square bracket of the right hand side of inequality (4.15). Clearly
\[ \max_t \left| c_t/\nu \sum_i c_i \tilde{\varepsilon}_i (\hat{\beta}_0 - \beta)'x_i \right| = o_p(1). \]

From (4.3) and Lemma 2.5 \( (k^{-1/2} \sum_i c_i \tilde{\varepsilon}_i x_{it} = O_p(1) \) for any \( t \))
\[ \max_t \left| \sum_i c_i \tilde{\varepsilon}_i (\hat{\beta}_0 - \hat{\beta}_{0(-t)})'x_i \right| \]
\[ \leq \sum_{t=1}^p \max_t \left| k^{1/2}(\hat{\beta}_0 - \hat{\beta}_{0(-t)})' \right| \left| k^{-1/2} \sum_i c_i \tilde{\varepsilon}_i x_{it} \right| = o_p(1), \]
where $\hat{\beta}_{0(t)}$ and $\hat{\beta}_{0(-t)}$ are the $l$-th elements of $\hat{\beta}_0$ and $\hat{\beta}_{0(-t)}$ respectively. From (4.11) and (4.9)
\[
\max_t |c_t \varepsilon_t (\hat{\beta}_{0(t)} - \beta)' x_t| \leq \max_t |k^{1/2}(\hat{\beta}_{0(t)} - \beta)' x_t| \leq \max_t |k^{-1/2} c_t \varepsilon_t x_t| = o_p(1).
\]
Therefore (4.12) is proved.

Since
\[
k^{1/2} \max_t |\hat{\sigma}_{v(t)}^2 - \hat{\sigma}_v^2| \leq k^{1/2} \max_t |\hat{\sigma}_{vp(t)}^2 - \hat{\sigma}_{vp}^2| + 2k^{1/2} \max_t |R_{1k} - R_{1k(-t)}| + k^{1/2} \max_t |R_{2k} - R_{2k(-t)}|.
\]
(4.14) follows from
\[
k^{1/2} \max_t |\hat{\sigma}_{vp(t)}^2 - \hat{\sigma}_{vp}^2| = o_p(1).
\]
In fact,
\[
k^{1/2} \max_t |\hat{\sigma}_{vp(t)}^2 - \hat{\sigma}_{vp}^2| \\
\leq \max_t [k(\nu - c_t)^{-1}] \max_t |k^{-1/2} c_t Z_t| + \max_t [k^{1/2} c_t (\nu - c_t)^{-1/2}] |\nu^{-1} \sum_t c_t Z_t|
\]
which is of the order $o_p(1)$ from the boundedness of $c_t$, Lemma 4.1 and (2.33). This completes the proof. □

**Remark.** As consequences of Lemma 4.3,
\[
k \sum_t [R_{1k} - R_{1k(-t)}]^2 \leq k^2 \max_t |R_{1k} - R_{1k(-t)}|^2 = o_p(1), \tag{4.16}
\]
\[
k \sum_t [R_{2k} - R_{2k(-t)}]^2 \leq k^2 \max_t |R_{2k} - R_{2k(-t)}|^2 = o_p(1). \tag{4.17}
\]

Now we have the following result on the asymptotic normality of the jackknife variance estimator $\hat{\sigma}_v(J)$.

**Theorem 4.1** Suppose that conditions (A), (B) and (C) hold, then
\[
V_a^{-1/2}(\hat{\sigma}_v^2 - \sigma_v^2) \rightarrow_d N(0,1) \quad \text{as} \quad k \rightarrow +\infty. \tag{4.18}
\]

**Proof.** Since
\[
\hat{\sigma}_v(J) - \sigma_v^2 = \hat{\sigma}_v^2 - \sigma_v^2 + \frac{k-1}{k} \sum_{t=1}^k (\hat{\sigma}_v^2 - \hat{\sigma}_{v(-t)}^2),
\]
the result follows from
\[ k^{1/2} \sum_t (\hat{\sigma}^2 - \hat{\sigma}^2_{(t)}) \rightarrow_p 0 \quad \text{as} \quad k \rightarrow +\infty. \] (4.19)

by Theorem 2.1. Now
\[ k^{1/2} \sum_t (\hat{\sigma}^2 - \hat{\sigma}^2_{(t)}) = k^{1/2} \sum_t (\hat{\sigma}^2_{\nu} - \hat{\sigma}^2_{\nu(t)}) - 2k^{1/2} \sum_t (R_{1k} - R_{1k(t)}) + k^{1/2} \sum_t (R_{2k} - R_{2k(t)}). \] (4.20)

Since \( \sum_i c_i |c_i - c_t| [(\nu - c_t)(\nu - c_t)] \leq C/k \) for any \( i \) for some positive constant \( C \), from Lemma 2.4

\[ k^{1/2} \sum_t (\hat{\sigma}^2_{\nu} - \hat{\sigma}^2_{\nu(t)}) = k^{1/2} \sum_t \left( \frac{1}{\nu - c_i} - \nu^{-1} \sum_i c_i \right) c_i Z_i \]
\[ = k^{-1/2} \nu^{-1} \sum_t \left[ \sum_i c_i (c_i - c_t) \right] c_i Z_i = o_p(1). \] (4.21)

For the second term in the right hand side of (4.20)
\[ k^{1/2} \sum_{t=1}^k (R_{1k} - R_{1k(t)}) \] (4.22)
\[ = k^{1/2} \sum_t (\nu - c_t)^{-1} \sum_i c_i \xi_i (\hat{\beta}_o - \hat{\beta}_o(t))'x_i + k^{1/2} \sum_t c_t (\nu - c_t)^{-1} \xi_i (\hat{\beta}_o(t) - \beta)'x_i \\
- k^{1/2} \nu^{-1} \sum_t c_t (\nu - c_t)^{-1} \sum_i c_i \xi_i (\hat{\beta}_0 - \beta)'x_i. \]

From (4.11), by a proof similar to that of Lemma 2.7, the second term of (4.22) is

\[ k^{1/2} \sum_t c_t (\nu - c_t)^{-1} \xi_i (\hat{\beta}_o(t) - \beta)'x_i = o_p(1). \]

By the boundedness of \( c_t \)'s, and Lemma 2.7, the third term of (4.22) is

\[ k^{1/2} \nu^{-1} \left( \sum_t c_t (\nu - c_t)^{-1} \right) \sum_i c_i \xi_i (\hat{\beta}_0 - \beta)'x_i = o_p(1). \]

From Lemma 2.4, and by a proof similar to that of Lemma 2.7 again

\[ R_{3k} \triangleq \sum_t (\nu - c_t)^{-1} w_{01}^{-1} h_{0i}^{-1} x_i \left( \xi_i - (\hat{\beta}_0 - \beta)'x_i \right) = o_p(1). \]
Then from (4.6), the first term in (4.22) is
\[ k^{1/2} \sum_{t} (\nu - c_t)^{-1} \sum_{i} c_i \hat{\epsilon}_i (\hat{\lambda}_0 - \hat{\lambda}_{(i-t)})' x_i \]
\[ = \left( k^{-1/2} \sum_{i} c_i x_i \right)' (k(X'W_0^{-1}X)^{-1}) R_{ik} = o_p(1). \]

From the arguments above, we have
\[ k^{1/2} \sum_{t} (R_{1k} - R_{1k(-t)}) = o_p(1). \tag{4.23} \]

Similarly, we can prove
\[ k^{1/2} \sum_{t} (R_{2k} - R_{2k(-t)}) = o_p(1). \tag{4.24} \]

Hence (4.19) follows from (4.20), (4.21), (4.23) and (4.24). The proof of the theorem is complete. \(\square\)

We know from Section 2.5.2 that \(V_a\) is the asymptotic variance of \(\hat{\sigma}_v^2\) and of \(\hat{\sigma}_v^2\) when \(\sigma_v^2 > 0\). The result in the above theorem implies that \(V_a\) is also the asymptotic variance of \(\hat{\sigma}_v^2(\mathcal{J})\). The following theorem shows that the jackknife variance estimator \(\hat{V}_{Jt}\) (defined in (4.2)) is consistent for \(V_a\).

**Theorem 4.2** Suppose that conditions (A), (B) and (C) hold. Then
\[ k(\hat{V}_{Jt} - V_a) \to_p 0 \quad \text{as} \quad k \to +\infty. \tag{4.25} \]

**Proof.** We have
\[ \hat{V}_{Jt} = (k(k - 1))^{-1} \sum_{t=1}^{k} (Q_t - \bar{Q})^2 \]
\[ = (k - 1) k^{-1} \frac{1}{2} \sum_{t=1}^{k} \left[ \frac{1}{2} \left( \hat{\sigma}_v^2 - \hat{\sigma}_v^2(-t) \right) - k^{-1} \sum_{i} \left( \hat{\sigma}_v^2 - \hat{\sigma}_v^2(-t) \right) \right]^2 \]
\[ = (k - 1) k^{-1} \sum_{t=1}^{k} \left( \hat{\sigma}_v^2 - \hat{\sigma}_v^2(-t) \right)^2 - (k - 1) k^{-2} \left[ \sum_{i} \left( \hat{\sigma}_v^2 - \hat{\sigma}_v^2(-t) \right) \right]^2 \]
\[ = (k - 1) k^{-1} \sum_{t=1}^{k} \left( \hat{\sigma}_v^2 - \hat{\sigma}_v^2(-t) \right)^2 + o_p(k^{-2}). \]
The last equality follows from (4.19). Hence

\[ k(\hat{V}_{jt} - V_a) \]

\[ = k \left[ \frac{k-1}{k} \sum_{i=1}^{k}(\hat{\sigma}_v^2 - \hat{\sigma}_{v(-t)}^2) - V_a \right] + o_p(1) \]

\[ = k \left[ \frac{k-1}{k} \sum_{i=1}^{k}(\hat{\sigma}_{vp}^2 - \hat{\sigma}_{vp(-t)}^2) - V_a \right] \]

\[ + \frac{k-1}{k} \left[ 4k \sum_{i=1}^{k}(R_{1k} - R_{1k(-t)})^2 + k \sum_{i=1}^{k}(R_{2k} - R_{2k(-t)})^2 \right] \]

\[ + (k-1)k^{-2}(\text{c.p. terms}) + o_p(1) \]

\[ = k \left[ \frac{k-1}{k} \sum_{i=1}^{k}(\hat{\sigma}_{vp}^2 - \hat{\sigma}_{vp(-t)}^2) - V_a \right] + (k-1)k^{-2}(\text{c.p. terms}) + o_p(1) \]

where "c.p. terms" means "cross product terms". The last equality holds due to (4.16) and (4.17). It remains to prove that

\[ k \left[ \sum_{i=1}^{k}(\hat{\sigma}_{vp}^2 - \hat{\sigma}_{vp(-t)}^2) - V_a \right] = o_p(1) , \quad (4.26) \]

since \( k^{-1}(\text{c.p. terms}) = o_p(1) \) from (4.26), (4.16), (4.17) by using the Cauchy–Schwarz inequality.

After some simplification,

\[ k \left[ \sum_{i=1}^{k}(\hat{\sigma}_{vp}^2 - \hat{\sigma}_{vp(-t)}^2) - V_a \right] \]

\[ = \frac{k^2}{\nu^2} k^{-1} \sum_{i=1}^{k} \left[ (Z_t - \sigma_v^2)^2 - \nu \text{var}(Z_i) \right] + \frac{k^2}{\nu^2} k^{-1} \sum_{i=1}^{k} \frac{(2\nu - c_i)c_i^2(Z_t - \sigma_v^2)^2}{(\nu - c_i)^2} \]

\[ + k(\hat{\sigma}_{vp}^2 - \sigma_v^2)^2 \sum_{i=1}^{k} \frac{c_i^2}{(\nu - c_i)^2} - 2(\hat{\sigma}_{vp}^2 - \sigma_v^2)k \sum_{i=1}^{k} \frac{c_i^2(Z_t - \sigma_v^2)}{(\nu - c_i)^2} . \]

Note that \( k^2/\nu^2 \) is bounded. Due to (2.33), the first term and the second term are of order \( o_p(1) \) from Lemma 2.4. The third term is of order \( o_p(1) \) since \( \hat{\sigma}_{vp}^2 \rightarrow_p \sigma_v^2 \), and \( k \sum_{i=1}^{k} c_i^2/(\nu - c_i)^2 \) is bounded. The fourth term is \( o_p(1) \) since \( k^{1/2}(\hat{\sigma}_{vp}^2 - \sigma_v^2) = O_p(1) \) from Theorem 2.1 and

\[ k^{1/2} \sum_{i=1}^{k} \frac{c_i^2(Z_t - \sigma_v^2)}{(\nu - c_i)^2} = o_p(1) \]

from Lemma 2.4. The proof is complete. \( \Box \)
Instead of jackknifing \( \hat{\sigma}_v^2 \), we can also jackknife \( \hat{\sigma}_v^2 = \max(\hat{\sigma}_v^2, 0) \), and get another estimator of \( \sigma_v^2 \) and a variance estimator for \( V_a \).

Let

\[
\hat{\sigma}_{v(-t)}^2 = \max(0, \hat{\sigma}_{v(-t)}^2).
\]

Hence a jackknife estimator of \( \sigma_v^2 \) is given as

\[
\hat{\sigma}_{v(J)}^2 = k^{-1} \sum_{t=1}^{k} \left[ k \hat{\sigma}_v^2 - (k - 1) \hat{\sigma}_{v(-t)}^2 \right].
\]

and a jackknife variance estimator is given as

\[
\hat{V}_{Jh} = k^{-1}(k - 1) \sum_{t=1}^{k} \left[ \hat{\sigma}_{v(-t)}^2 - k^{-1} \sum_{t=1}^{k} \hat{\sigma}_{v(-t)}^2 \right]^2.
\]

We will see these estimators have the same properties as \( \hat{\sigma}_{v(J)}^2 \) and \( V_a \) theoretically when \( \sigma_v^2 > 0 \). To prove this result, we need the following lemma.

Lemma 4.4 Suppose that conditions (A), (B) and (C) hold. If \( \sigma_v^2 > 0 \), then

\[
P\{ \bigcup_t (\hat{\sigma}_{v(-t)}^2 \leq 0) \} \to 0 \quad \text{as} \quad k \to +\infty. \tag{4.27}
\]

Proof. Note that

\[
P\{ \bigcup_t (\hat{\sigma}_{v(-t)}^2 \leq 0) \}
= P\left( \bigcup_t (\hat{\sigma}_{v(-t)}^2 \leq 0) \right) \cap (\hat{\sigma}_v^2 < \sigma_v^2/2) + P\left( \bigcup_t (\hat{\sigma}_{v(-t)}^2 \leq 0) \right) \cap (\hat{\sigma}_v^2 \geq \sigma_v^2/2)
\leq P\{ \hat{\sigma}_v^2 < \sigma_v^2/2 \} + P\left( \bigcup_t (\hat{\sigma}_{v(-t)}^2 \leq 0) \right) \cap (\hat{\sigma}_v^2 \geq \sigma_v^2/2)
\leq P\{ \hat{\sigma}_v^2 < \sigma_v^2/2 \} + P\left( \sum_t (|\hat{\sigma}_{v(-t)}^2 - \hat{\sigma}_v^2| \geq \sigma_v^2/2) \right)
= P\{ \hat{\sigma}_v^2 < \sigma_v^2/2 \} + P\left( \max_t |\hat{\sigma}_{v(-t)}^2 - \hat{\sigma}_v^2| \geq \sigma_v^2/2 \right).
\]

Hence (4.27) follows from (4.14) and Corollary 2.1. \( \square \)

Define the set \( A_k \) as

\[
A_k = \bigcup_t (\hat{\sigma}_{v(-t)}^2 \leq 0) \cup (\hat{\sigma}_{v(J)}^2 \leq 0).
\]

Then from Lemma 4.4 and Corollary 2.1,

\[
P(A_k) \leq P\{ \bigcup_t (\hat{\sigma}_{v(-t)}^2 \leq 0) \} + P\{ \hat{\sigma}_{v(J)}^2 \leq 0 \} \to 0 \tag{4.28}
\]

as \( k \to \infty \). Now we have the following two results.
Theorem 4.3 Assume that the conditions of Theorem 4.1 hold. If \( \sigma_v^2 > 0 \), then
\[
V_a^{-1/2}(\hat{\sigma}_v^2(J) - \sigma_v^2) \rightarrow_d N(0, 1) \quad as \quad k \rightarrow +\infty.
\]

Proof. Let \( A_k^c \) be the complement of \( A_k \). Then the result follows from (4.28) and the fact that \( \hat{\sigma}_v^2(J) \) and \( \hat{\sigma}_v^2(J) \) are equal on \( A_k^c \). \( \square \)

Theorem 4.4 Under the conditions of Theorem 2.4,
\[
k(\hat{V}_{JH} - V_a) \rightarrow_p 0 \quad as \quad k \rightarrow +\infty.
\]

The proof is similar to that of Theorem 4.3 and hence is omitted.

Remark 1. Our simulation results indicate that both \( \hat{\sigma}_v^2(J) \) and \( \hat{\sigma}_v^2(J) \) reduce the bias in \( \hat{\sigma}_v^2 \) and \( \hat{\sigma}_v^2 \). The estimator \( \hat{\sigma}_v^2(J) \) is found to be almost unbiased.

Remark 2. From Theorems 4.1 and 4.2, we can use the following quantity as a pivotal for constructing confidence intervals for \( \sigma_v^2 \):
\[
\hat{V}_{Ji}^{-1/2}(\hat{\sigma}_v^2 - \sigma_v^2).
\]

Hence, a confidence interval of level \( 1 - \alpha \) for \( \sigma_v^2 \) may be taken as
\[
(\hat{\sigma}_v^2(J) - t_{1-\alpha/2, k-1} \hat{V}_{Ji}^{1/2}, \quad \hat{\sigma}_v^2(J) + t_{1-\alpha/2, k-1} \hat{V}_{Ji}^{1/2}), \quad (4.29)
\]
where \( t_{1-\alpha/2, k-1} \) is the \( 1 - \alpha/2 \) percentile of a student \( t \)-variate with \( k - 1 \) degrees of freedom. Theoretically, when we feel certain that \( \sigma_v^2 > 0 \), we can also construct a confidence interval of level \( 1 - \alpha \) for \( \sigma_v^2 \) as follows
\[
(\hat{\sigma}_v^2(J) - t_{1-\alpha/2, k-1} \hat{V}_{Jh}^{1/2}, \quad \hat{\sigma}_v^2(J) + t_{1-\alpha/2, k-1} \hat{V}_{Jh}^{1/2}). \quad (4.30)
\]
But the latter approach leads to severe undercoverage as found in our simulation study in Section 4.5. Actually, when \( \sigma_v^2 \) is small or close to 0, \( \hat{V}_{Jh} \) tends to be too small due to the truncation of AMINQU estimator. It is biased for estimating \( V_a \) in this case.
4.3 A Jackknife Variance Estimator for $\hat{\beta}_w$

As pointed out in Shao and Rao (1993), when the observations within a group are correlated, neither does the usual delete-one jackknife method provide a consistent estimator for $V_k$, nor does the modified jackknife method of Shao (1989b) by jackknifing each term in the expression of $V_k$. They provide a consistent jackknife estimator of $V_k$ by the delete-group jackknife method. The associated inferences on $\beta$ are therefore robust against possible within-group error independence. In other words, their variance estimator is consistent if the observations are obtained from our model (1.1). But their variance estimator does not take into account the special structure of model (1.1), and so may be less efficient. Here, with our model structure in mind, we obtain a consistent variance estimator using the delete-group jackknife method. As shown in the simulation study in Section 4.5, our variance estimator is more efficient than their variance estimator.

After deleting observations in the $t$-th group at design point $x_t$, we recalculate the AMINQU estimators of $\sigma^2_\epsilon$ and $\sigma^2_\eta$, and then we get the corresponding weights $\tilde{w}_{i(-t)}$ for $w_i(i \neq t)$. Using these weights we estimate $\beta$, and denote the estimator by $\hat{\beta}_{w(-t)}$. Then

$$\hat{\beta}_{w(-t)} = (X'_{(-t)}\tilde{W}_{(-t)}X_{(-t)})^{-1}X'_{(-t)}\tilde{W}_{(-t)}y_{(-t)}$$

(4.31)

where $X_{(-t)}$ and $y_{(-t)}$ are defined in Section 4.2, and

$$\tilde{W}_{(-t)} = \text{diag}_{i \neq t}(\tilde{w}_{i(-t)})$$

with

$$\tilde{w}_{i(-t)} = (1 - d_i)\hat{\sigma}^2_{\epsilon(-t)} + d_i \left[ n_{i}^{-1} (1 + \rho_{io}n_i)^2 \sum_j (y_{ij} - \bar{y}_i)^2 + (y_i - \hat{\beta}'_o x_i)^2 \right]$$

$$= (1 - d_i)\hat{\sigma}^2_{\epsilon(-t)}$$

$$+ d_i \left[ n_{i}^{-1} (1 + \rho_{io}n_i)^2 \sum_j (e_{ij} - \bar{\epsilon}_i)^2 + (\hat{\epsilon}_i - (\hat{\beta}'_{o(-t)} - \beta)' x_i)^2 \right]$$

We define the jackknife estimator of $V_k$ as

$$\hat{V}_j = k^{-1}(k - 1) \sum_t (\hat{\beta}_{w(-t)} - \hat{\beta}_{w}) (\hat{\beta}_{w(-t)} - \hat{\beta}_{w})'$$.
As shown in the following theorem, \( \hat{V}_k \) is consistent for \( V_k \).

**Theorem 4.5** Suppose that the conditions (A), (B) and (C) in Section 2.5.1 hold. Also assume \( \sigma_k^2 > 0 \). Then

\[
k(\hat{V}_k - V_k) \to_p 0 \quad \text{as} \quad k \to \infty
\]

Before proving the theorem, we present the following lemmas.

**Lemma 4.5** Let \( u_i \) be as defined in Section 3.1. Suppose that the assumptions in Theorem 4.5 hold. Then

\[
k^{1/2} \max_{i \leq k} |\hat{w}_i^{-1} u_i - 1| = O_p(1), \quad (4.32)
\]

\[
k^{1/2} \max_{i \leq k, i \neq t} |\hat{w}_i^{-1} u_i - 1| = O_p(1), \quad (4.33)
\]

\[
k^{1/2} \max_{i \leq k, i \neq t} |\hat{w}_i^{-1} u_i - \hat{w}_{i(-t)}^{-1} u_i| = o_p(1). \quad (4.34)
\]

**Proof.** Note that

\[
k^{1/2} |\hat{w}_i u_i^{-1} - 1|
\]

\[
\begin{align*}
&\leq (1 - d_i) k^{1/2} |\hat{\sigma}^2_{\hat{\varepsilon}} - \sigma^2_{\varepsilon} u_i^{-1} | + d_i k^{1/2} (\hat{\beta}_0 - \beta)' x_i [\hat{\sigma}^2_{\varepsilon} - (\hat{\beta}_0 - \beta)' x_i] u_i^{-1} | \\
&\leq \sigma^2_{\varepsilon} k^{1/2} |\hat{\sigma}^2_{\hat{\varepsilon}} - \sigma^2_{\varepsilon} | + d_i \zeta_i (d_i^{-1/2} + k^{-1/2} \zeta_i)
\end{align*}
\]

where \( \zeta_i = |k^{1/2} (\hat{\beta}_0 - \beta)' x_i|/u_i^{1/2} \). The result follows from

\[
\max_i \zeta_i \leq 2 \sigma_{\varepsilon}^{-1} |k^{1/2} (\hat{\beta}_0 - \beta)' x_i| = O_p(1)
\]

which is implied by Lemma 2.6. (4.33) and (4.34) can be proved similarly by using (4.8) and (4.3). The details are omitted. \( \square \)

**Lemma 4.6** Suppose that the assumptions in Theorem 4.5 hold. Then

\[
k^{3/2} \left[ (\sum_{i \neq t} \hat{w}_{i(-t)}^{-1} x_i x_i')^{-1} - (X' \hat{W}^{-1} X)^{-1} \right] \to_p 0 \quad \text{uniformly in } t
\]

**Proof.** From Lemma 3.4

\[
k(X' \hat{W}^{-1} X)^{-1} = k(\sum_i \hat{w}_i^{-1} x_i x_i')^{-1} = O_p(1). \quad (4.35)
\]
By Lemma 4.5

\[ k( \sum_{i \neq t} \hat{\omega}^{-1}_{i(-t)} x_i x'_i )^{-1} = O_p(1) \quad \text{uniformly in } t, \]

\[ k( \sum_{i \neq t} \hat{\omega}^{-1}_i x_i x'_i )^{-1} = O_p(1) \quad \text{uniformly in } t. \]

Note that

\[
( \sum_{i \neq t} \hat{\omega}^{-1}_{i(-t)} x_i x'_i )^{-1} - ( \sum_{i} \hat{\omega}^{-1}_i x_i x'_i )^{-1} \\
= \left[ ( \sum_{i \neq t} \hat{\omega}^{-1}_{i(-t)} x_i x'_i )^{-1} - ( \sum_{i \neq t} \hat{\omega}^{-1}_i x_i x'_i )^{-1} \right] \\
+ \left[ ( \sum_{i \neq t} \hat{\omega}^{-1}_{i(-t)} x_i x'_i )^{-1} - ( \sum_{i} \hat{\omega}^{-1}_i x_i x'_i )^{-1} \right] \\
= ( \sum_{i \neq t} \hat{\omega}^{-1}_{i(-t)} x_i x'_i )^{-1} ( \sum_{i \neq t} (\hat{\omega}_i - \hat{\omega}_{i(-t)}) x_i x'_i ) ( \sum_{i \neq t} \hat{\omega}^{-1}_i x_i x'_i )^{-1} \\
+ h_t(X' \hat{W}^{-1} X)^{-1} \hat{\omega}^{-1}_t x_t x'_t (X' \hat{W}^{-1} X)^{-1}, \tag{4.36}
\]

where the second term of the right hand side of equality was obtained by applying matrix result (4.4), and

\[ h_t = 1 - \hat{\omega}^{-1}_t x'_t (X' \hat{W}^{-1} X)^{-1} x_t. \]

From (4.35), \( \max_t |h_t| = O(1) \). From condition (A) and \( \max_t \hat{\omega}^{-1}_t = O_p(1) \),

\[ \max_t |h_t| \hat{\omega}^{-1}_t x_t x'_t = O_p(1). \]

Hence the second term in the right hand side of equation (4.36) is of order \( o_p(k^{-1/2}). \)

Therefore the result follows from

\[ k^{-1/2} \sum_{i \neq t} (\hat{\omega}_i - \hat{\omega}_{i(-t)}) x_i x'_i = o_p(1) \quad \text{uniformly in } t. \]

From condition (A), we have

\[ \|k^{-1/2} \sum_{i \neq t} (\hat{\omega}_i - \hat{\omega}_{i(-t)}) x_i x'_i \| \leq C k^{1/2} \max_{i \leq k, i \leq t} |\hat{\omega}_i u_i - \hat{\omega}_{i(-t)} u_i| k^{-1} \sum_i \hat{\omega}_i^{-1} \]

for some positive constant \( C \). Hence the result follows from (4.34) and the fact \( k^{-1} \sum_i \hat{\omega}_i^{-1} = O_p(1). \)

**Lemma 4.7** Suppose that the assumptions in Theorem 4.5 hold. Then

\[
\hat{\beta}_{\omega(-t)} - \hat{\beta}_{\omega} = -(X'D_1 X)^{-1}(x_t u_t^{-1} \hat{\epsilon}_t + 2(X'D_2 D_0 X)(X'W_0^{-1} X)^{-1} w_0^{-1} x_t \hat{\epsilon}_t) \\
+ o_p(k^{-1}) \quad \text{uniformly in } t.
\]
CHAPTER 4. JACKKNIFE ESTIMATORS

Proof. From the definitions of $\hat{\beta}_w(-t)$ and $\hat{\beta}_w$,

$$
\hat{\beta}_w(-t) - \hat{\beta}_w = (\sum_{i \neq t} \hat{\omega}_i^{-1} x_i x_i')^{-1} \sum_{i \neq t} \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i - (\sum_{i \neq t} \hat{\omega}_i^{-1} x_i x_i')^{-1} \sum_{i \neq t} \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i
$$

$$
= (X'\hat{W}^{-1}X)^{-1} \left[ \sum_{i \neq t} \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i - \sum_i \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i \right]
$$

$$
+ \left[ (\sum_{i \neq t} \hat{\omega}_i^{-1} x_i x_i')^{-1} - (X'\hat{W}^{-1}X)^{-1} \right] \sum_{i \neq t} \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i
$$

$$
= (X'\hat{W}^{-1}X)^{-1} \left[ \sum_{i \neq t} \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i - \sum_i \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i \right] + o_p(k^{-1})
$$

where the last equality follows from Lemma 4.6. From Lemma 3.4

$$
k([(X'\hat{W}^{-1}X)^{-1} - (X'D_t^{-1}X)^{-1}] \rightarrow_p 0.
$$

Hence the result follows if

$$
\sum_{i \neq t} \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i - \sum_i \hat{\omega}_i^{-1} \hat{\epsilon}_i x_i
$$

$$
= -x_t u_t^{-1} \hat{\epsilon}_t - 2(X'D_t D_0 X)(X'W_0^{-1}X)^{-1} w_0^{-1} x_t \hat{\epsilon}_t + o_p(1) \text{ uniformly in } t.
$$

The left hand side of the previous expression is equal to

$$
\sum_{i \neq t} (\hat{\omega}_i^{-1} - \hat{\omega}_t^{-1}) \hat{\epsilon}_i x_i - \hat{\omega}_t^{-1} \hat{\epsilon}_t x_t.
$$

From (4.32)

$$
\hat{\omega}_t^{-1} \hat{\epsilon}_t x_t = u_t^{-1} \hat{\epsilon}_t x_t + o_p(1) \text{ uniformly in } t.
$$

Also from (4.7), by using Lemma 2.6,

$$
\hat{\beta}_0(-t) - \hat{\beta}_0 = -w_0^{-1}(X'W_0^{-1}X)^{-1} x_t \hat{\epsilon}_t + O_p(k^{3/2}) \text{ uniformly in } t.
$$

Hence it suffices to show

$$
\sum_{i \neq t} (\hat{\omega}_i^{-1} - \hat{\omega}_t^{-1}) \hat{\epsilon}_i x_i = 2(X'D_t D_0 X)(\hat{\beta}_0(-t) - \hat{\beta}_0) + o_p(1) \text{ uniformly in } t. \ (4.37)
$$

Since

$$
\hat{\omega}_i(-t) - \hat{\omega}_i = (1 - d_i)(\hat{\sigma}_i^2(-t) - \hat{\sigma}_i^2)
$$

$$
+ d_i \left[ x'_i (\hat{\beta}_0 - \hat{\beta}_0(-t)) x'_i (2\hat{\epsilon}_i - (\hat{\beta}_0 + \hat{\beta}_0(-t) - 2\beta) \right],
$$
we have

\[
\sum_{i \neq t} (\hat{\omega}^{-1}_{i(-t)} - \hat{\omega}^{-1}_t)\varepsilon_i x_i = \\
2\sum_{i \neq t} d_i \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t \varepsilon_i^2 x_i x_i' (\hat{\beta}_{0(-t)} - \hat{\beta}_0) \\
+ (\hat{\sigma}^2_v - \sigma^2_v) \sum_{i \neq t} (1 - d_i) \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t \varepsilon_i x_i \\
+ \sum_{i \neq t} d_i \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t \varepsilon_i x_i x_i' (\hat{\beta}_0 - \hat{\beta}_{0(-t)}) x_i' (\hat{\beta}_0 + \hat{\beta}_{0(-t)} - 2\beta) \\
\triangleq (I) + (II) + (III) + o_p(1).
\] (4.38)

We will show that terms (II) and (III) in (4.38) are of order $o_p(1)$. In the term (II),
$k^{1/2} \max_t |\hat{\sigma}_v^2 - \sigma_v^2| = o_p(1)$ from Lemma 4.3 and Lemma 4.4. Also, from the fact

\[
k^{-1/2} \sum_{i \neq t} u_i^{-2} \varepsilon_i x_i = o_p(1) \text{ uniformly in } t,
\]

and Lemma 4.5, we have

\[
k^{-1/2} \sum_{i \neq t} \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t \varepsilon_i x_i = o_p(1) \text{ uniformly in } t.
\]

Hence, term (II) is of order $o_p(1)$ uniformly in $t$.

For the term (III) in (4.38), from condition (A), Lemma 2.5, Lemma 4.2 and equation (4.11),

\[
\|\sum_{i \neq t} d_i \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t \varepsilon_i x_i x_i' (\hat{\beta}_0 - \hat{\beta}_{0(-t)}) x_i' (\hat{\beta}_0 + \hat{\beta}_{0(-t)} - 2\beta)\| \leq \sigma_o(k^{-1}) \sum_{i \neq t} \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t |\varepsilon_i|.
\]

Note that, from (4.32), (4.33)

\[
|k^{-1/2} \sum_{i \neq t} |\varepsilon_i| (u_i^2 - \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t)| \leq \left[(\max_{i \neq t} |u_i^2 - \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t| - 1) + 3(k^{1/2} \max_{i \leq k} |u_i \hat{\omega}^{-1}_i - 1|) + k^{1/2} \max_{i \neq t} |u_i \hat{\omega}^{-1}_{i(-t)} - 1|\right] \\
\times C k^{-1} \sum_{i \neq t} u_i^{-2} |\varepsilon_i| = o_p(1) \text{ uniformly in } t.
\]

Hence

\[
\sum_{i \neq t} \hat{\omega}^{-1}_{i(-t)} \hat{\omega}^{-1}_t |\varepsilon_i| = \sum_{i \neq t} u_i^{-2} |\varepsilon_i| + o_p(k^{1/2}) = O_p(1) \text{ uniformly in } t.
\]

Therefore term (III) is also of order $o_p(1)$ uniformly in $t$. 

CHAPTER 4. JACKKNIFE ESTIMATORS

Back to term (I) in (4.38). From Lemma 3.1, we have
\[ k^{-1}(X'D_0D_3X - \sum_{i=1}^{k} d_i x_i x_i' u_i^{-2} \tilde{e}_i^2) \rightarrow_p 0. \]

And from Lemma 4.1,
\[ k^{-1} \max_{t} \|d_t x_t x_t' u_t^{-2} \tilde{e}_t^2\| \rightarrow_p 0. \]

Hence
\[ k^{-1}(X'D_0D_3X - \sum_{i \neq t} d_i x_i x_i' u_i^{-2} \tilde{e}_i^2) \rightarrow_p 0 \text{ uniformly in } t. \]

Using argument similar to (4.39), we have
\[ k^{-1/2} \sum_{i \neq t} d_i x_i x_i' \tilde{e}_i^2(u_i^2 - \hat{w}_{i(i-1)}^{-1}) = O_p(1) \text{ uniformly in } t. \]

Hence, in term (I)
\[ k^{-1}\left(X'D_0D_3X - \sum_{i \neq t} d_i \hat{w}_{i(i-1)}^{-1}\hat{w}_{i(i-1)}^{-1}\tilde{e}_i^2 x_i x_i'\right) \rightarrow_p 0 \text{ uniformly in } t. \]

Therefore, (4.37) follows from (4.38). This completes the proof of the lemma. \( \square \)

Proof of Theorem 4.5. From Lemma 4.7,
\[ k(\hat{V}_j - V_k) = k(X'D_1X)^{-1}\left[k^{-1}\sum_{i}(u_i^{-2}\tilde{e}_i^2 - \tau_{3i})x_i x_i' + 2(X'D_3D_0X)(X'W_0^{-1}X)^{-1}(k^{-1}\sum_{i}\\tilde{w}_i^{-1}(u_i^{-1}\tilde{e}_i^2 - \tau_{3i})x_i x_i') + 2(k^{-1}\sum_{i}\\tilde{w}_i^{-1}(u_i^{-1}\tilde{e}_i^2 - \tau_{3i})x_i x_i')(X'W_0^{-1}X)^{-1}(X'D_3D_0X) + 4(X'D_3D_0X)(X'W_0^{-1}X)^{-1}(k^{-1}\sum_{i}\\tilde{w}_i^{-2} \tilde{e}_i^2 - w_i)x_i x_i')(X'W_0^{-1}X)^{-1}(X'D_0D_3X)\right] \times [k(X'D_1X)^{-1}] + o_p(1). \]

Hence, the result follows from (3.14), (2.23) and Lemmas 3.1. \( \square \)

Remark 1. We can construct a delete-group jackknife estimator of the asymptotic covariance matrix of a parameter \( \theta = g(\beta) \) for some function \( g \). The consistency of such a jackknife variance estimator can be established following standard arguments. The details are omitted for simplicity.

Remark 2. Combining results in Theorems 3.1 and 4.5, we can construct confidence intervals on \( \beta \), using \( \hat{\beta}_w \) and \( \hat{V}_j \), by substituting \( \hat{V}_j \) in place of \( \hat{V}_k \) in interval (3.16). Also the confidence intervals on elements of \( \beta \) can be constructed analogous to (3.17). The details are omitted for simplicity.
4.4 A Jackknife variance estimator for $\hat{\beta}_{sw}$

Similar to the jackknife procedure in Section 4.3, we can construct a jackknife estimator of $V_{sk}$, the asymptotic variance–covariance matrix of $\hat{\beta}_{sw}$ (see Section 3.4). After deleting all observations in the $t$-th group at the design point $x_t$, we calculate the approximate MINQU estimator $\hat{\sigma}_v^2$ of $\sigma_v^2$. Then we can construct weights $\hat{w}_{i(-t),a}$ for $w_i$ ($i \neq t$). Using these weights, we can compute a WLS estimator $\hat{\beta}_{sw(-t)}$ of $\beta$ as follows:

$$
\hat{\beta}_{sw(-t)} = (X'_{(-t)} \hat{W}_{a(-t)^{-1}} X_{(-t)})^{-1} X'_{(-t)} \hat{W}_{a(-t)^{-1}} y_{(-t)}
$$

where $X_{(-t)}$ and $y_{(-t)}$ are defined in Section 4.2, and

$$
\hat{W}_{a(-t)} = \operatorname{diag}_{i \neq t}(\hat{w}_{i(-t),a}),
$$

with

$$
\hat{w}_{i(-t),a} = \hat{\sigma}_v^2 + n_t^{-1} s_i^2.
$$

Then the jackknife estimator of $V_{sk}$ is defined as

$$
\hat{V}_{sk} = k^{-1}(k - 1) \sum_t (\hat{\beta}_{sw(-t)} - \hat{\beta}_{sw})(\hat{\beta}_{sw(-t)} - \hat{\beta}_{sw})'.
$$

We now give the following theorem on the consistency of $\hat{V}_{sk}$. It can be proved along the lines of the proof of Theorem 4.5. It is therefore omitted for simplicity.

**Theorem 4.6** Under the assumptions in Theorem 4.5, we have

$$
k(\hat{V}_{sk} - V_{sk}) \rightarrow_p 0 \quad \square
$$

We can construct confidence intervals on the regression coefficients using the estimator $\hat{V}_{sk}$ of the covariance matrix. These confidence intervals performed well with regard to coverage probability, in a simulation study, but the variance estimator $\hat{V}_{sk}$ is very unstable, and biased, when $\sigma_v^2$ is relatively small or moderate, and the $n_t$'s are small (say less than 5). Hence, this jackknife variance estimator is a viable alternative only when the number of replicates is moderate or large.
4.5 Simulation Results

In this section, using the Monte Carlo method we compare our four estimators of $\sigma^2_v$, viz., the untruncated and truncated AMINQU estimators $\hat{\sigma}^2_v$, $\hat{\sigma}^*_v$ (given in Section 2.4), and the corresponding jackknife estimators $\hat{\sigma}^2_{(J)}$ and $\hat{\sigma}^2_{(J)}$ (given in Section 4.2). We have shown that all these variance estimators are asymptotically normally distributed with common asymptotic variance $V_a$ when $\sigma^2_v > 0$. For estimation of $V_a$, we proposed two substitution estimators $\hat{V}_a$ and $\hat{V}_a$ in Section 2.5.3 by using $\hat{\sigma}^2_v$ and $\hat{\sigma}^*_v$, respectively. We also obtained two jackknife estimators $\hat{V}_J$ and $\hat{V}_{Jh}$ of $V_a$ by delete-group jackknifing $\hat{\sigma}^2_v$ and $\hat{\sigma}^*_v$, respectively, in Section 4.2. In this section, we investigate the relative biases of all four estimators of $\sigma^2_v$, as well as the coverage probabilities (CP) and lengths of confidence intervals (LCI) associated with these estimators of $\sigma^2_v$ and corresponding variance estimators.

For the estimation of regression parameters $\beta$, we focus on the behaviour of estimators of asymptotic covariance matrix since all WLS estimators of $\beta$ are unbiased. The variance estimators we consider here include the substitution estimator $\hat{V}_k$ (given in Section 3.3), the delete-group jackknife variance estimator $\hat{V}_J$ (given in Section 4.3), the substitution estimator $\hat{V}_{sk}$ by alternative procedure (given in Section 3.4), and the delete-group jackknife variance estimator $\hat{V}_{J^*}$ of Shao and Rao (1993). Shao and Rao's variance estimator is considered here since it is also consistent under our model (1.1).

4.5.1 Description of the design

We have used the following simple common mean model with nested error structure and unequal error variances (same as that in Section 1.1.4):

$$y_{ij} = \mu + v_i + e_{ij}, \quad i = 1, \ldots, k = 30, \quad j = 1, 2, n_i \equiv 3.$$ 

Extensive simulation studies in the literature (see, e.g., Rao, 1973; Shao and Rao, 1993) have shown that the results from the balanced model are not very different from the results from the unbalanced model, when the numbers of replicates at
the different design points do not differ significantly. Hence, the balanced model is chosen here for simplicity. Without loss of generality, we take \( \mu = 0 \) in generating \( y_{ij} \)-values. The values of the variance \( \sigma_i^2 \) of \( v_i \) are chosen to be 0.1, 0.25, 0.5, and 2.0. And the error variances \( \sigma_i^2 \) are chosen such that one third of the groups (10 groups) have error variance \( \sigma_i^2 \), one third have error variance \( \sigma_i^2 \), and the remaining one third have error variance \( \sigma_i^2 \). The selected patterns for \( (\sigma_i^2, \sigma_i^2, \sigma_i^2) \) are: (i) \((1/2, 1, 2)\), (ii) \((1/4, 1, 2)\) and (iii) \((1/4, 1, 4)\), ordered according to increasing heteroscedasticity.

We consider three types of distributions for \( v_i \) and \( e_{ij} \):

1. \( v_i \simd N(0, \sigma_i^2) \) and \( e_{ij} \simd N(0, \sigma_i^2) \) for each \( i \).

2. \( v_i \simd \text{Double Exp}(\sigma_i/\sqrt{2}) \), and \( e_{ij} \simd \text{Double Exp}(\sigma_i/\sqrt{2}) \) for each \( i \), where Double Exp(\( \lambda \)), for \( \lambda > 0 \), is the double exponential distribution with probability density function

\[
f(x \mid \lambda) = (2\lambda)^{-1} \exp(-|x|/\lambda), \quad -\infty < x < +\infty.
\]

3. \( v_i \simd \text{uniform distribution on } (-\sigma_i\sqrt{3}, \sigma_i\sqrt{3}) \), \( e_{ij} \simd \text{uniform distribution on } (-\sigma_i\sqrt{3}, \sigma_i\sqrt{3}) \) for each \( i \).

The prior values chosen for the AMINQUE procedures are \( \sigma_i^2 = 1 \) and \( \sigma_i^2 = 1 \) for all \( i \).

For each set of distributions, and each combination of selected values of \( \sigma_i^2 \) and \( (\sigma_i^2, \sigma_i^2, \sigma_i^2) \), we generated 1000 independent sample runs of observations \( \{y_{ij}\} \). The quantities calculated for each run and over all 1000 runs are as follows.

1. The untruncated and truncated AMINQUE estimators \( \hat{\sigma}_v^2 \) and \( \hat{\sigma}_v^2 \), and the corresponding variance estimators \( \hat{V}_a \) and \( \hat{V}_a \); the coverage probability (CP) and length of confidence intervals (LCI) at level 90\% associated with \( \hat{\sigma}_v^2 \) and \( \hat{V}_a \), and the CP and LCI at level 90\% associated with \( \hat{\sigma}_v^2 \) and \( \hat{V}_a \).

2. The untruncated and truncated delete-group jackknife estimators \( \hat{\sigma}_v^{2(J)} \) and \( \hat{\sigma}_v^{2(J)} \), and the corresponding jackknife variance estimators \( \hat{V}_{J} \) and \( \hat{V}_{J} \); the CP and LCI at level 90\% associated with \( \hat{\sigma}_v^{2(J)} \) and \( \hat{V}_{J} \), and the CP and LCI at level 90\% associated with \( \hat{\sigma}_v^{2(J)} \) and \( \hat{V}_{J} \).

3. The substitution variance estimator \( \hat{V}_k \) and the jackknife variance estimator \( \hat{V}_J \) associated with the WLS estimator \( \hat{\beta}_w (\beta = \mu \text{ here}) \); the CP and LCI at level 90\%
CHAPTER 4. JACKKNIFE ESTIMATORS

using \( \hat{\gamma}_k \) and \( \hat{\gamma}_j \) with \( \hat{\beta}_w \); the relative biases (RB) and the coefficients of variation (CV) of these two variance estimators relative to the empirical variance (EV) of \( \hat{\beta}_w \).

(4) For the alternative estimator \( \hat{\beta}_{s,w} \) of \( \beta \), the CP, LCI, RB and CV associated with substitution variance estimator \( \hat{\gamma}_{s,k} \).

(5) The CP, LCI, RB and CV associated with the WLS estimator \( \hat{\beta}_{w}^{*} \) and the corresponding delete-group jackknife variance estimator proposed in Shao and Rao (1993).

(6) The percentage efficiency (E) of our WLS estimator \( \hat{\beta}_w \), relative to \( \hat{\beta}_{s,w} \) and Shao and Rao’s WLS estimator \( \beta_{w}^{*} \).

Here the coverage probability (CP) and length of confidence interval (LCI) at level 90% associated with an estimator \( \hat{\theta} \) and its variance estimator \( \hat{\gamma}_{\theta} \) (for a parameter \( \theta \)), the empirical variance (EV(\( \hat{\theta} \))) of parameter estimator \( \hat{\theta} \), the relative bias (RB) and the coefficient of variation (CV) of the variance estimator \( \hat{\gamma}_{\theta} \), and the relative bias (RB\( _v \)) of \( \hat{\theta} \), are defined as

\[
\text{CP} = 100 \times \sum_{r=1}^{1000} I(\hat{\theta}^{(r)} - t_{0.95,29}(\hat{\gamma}_{\theta}^{(r)})^{1/2} \leq \theta_0 \leq \hat{\theta}^{(r)} + t_{0.95,29}(\hat{\gamma}_{\theta}^{(r)})^{1/2})/1000,
\]

\[
\text{LCI} = 2 \sum_{r=1}^{1000} t_{0.95,29}(\hat{\gamma}_{\theta}^{(r)})^{1/2}/1000,
\]

\[
\text{EV}(\hat{\theta}) = \sum_{r=1}^{1000} (\hat{\theta}^{(r)})^2/1000 - \left( \sum_{r=1}^{1000} \hat{\theta}^{(r)}/1000 \right)^2,
\]

\[
\text{RB} = 100 \times \left( \frac{1}{1000} \sum_{r=1}^{1000} \hat{\gamma}_{\theta}^{(r)}/\text{EV}(\hat{\theta}) - 1 \right),
\]

\[
\text{CV} = 100 \times \sqrt{\text{MSE}(\hat{\gamma}_{\theta})/\text{EV}(\hat{\theta})},
\]

\[
\text{RB}_v = 100 \times \left( \frac{1}{1000} \sum_{r=1}^{1000} \hat{\theta}^{(r)}/\theta_0 - 1 \right),
\]

where

\[
\text{MSE}(\hat{\gamma}_{\theta}) = \sum_{r=1}^{1000} [\hat{\gamma}_{\theta}^{(r)} - \text{EV}(\hat{\theta})]^2/1000,
\]

\( \theta_0 \) is the true value of the parameter \( \theta \), \( I(a \leq \theta_0 \leq b) = 1 \) if \( a \leq \theta_0 \leq b \), and = 0 otherwise, \( \hat{\theta}^{(r)} \) and \( \hat{\gamma}_{\theta}^{(r)} \) are the calculated values of \( \hat{\theta} \) and \( \hat{\gamma}_{\theta} \) from the data in the \( r \)-th run \( (r = 1, \ldots, 1000) \), and \( t_{0.95,29} \) is the 95% percentile of a student t-variante with 29 degrees of freedom.
CHAPTER 4. JACKKNIFE ESTIMATORS

The percentage efficiency (E) of an estimator $\hat{\beta}^{(2)}$ relative to another estimator $\hat{\beta}^{(1)}$ of $\beta$ is defined as

$$E(\hat{\beta}^{(2)}, \hat{\beta}^{(1)}) = 100 \times EV(\hat{\beta}^{(1)})/EV(\hat{\beta}^{(2)}).$$

We found that the results when the errors follow a uniform distribution are quite similar to the results when the errors follow a normal distribution. Hence only results for errors following normal and double exponential distributions are presented in Tables 4.1–4.8.

4.5.2 Empirical results on estimating $\sigma^2_v$

Table 4.1 reports the values of the relative biases ($RB_v$) of four estimators of $\sigma^2_v$ under normal and double exponential distributions respectively. We have the following results: (1) Both AMINQUE estimators $\hat{\sigma}^2_v$ and $\hat{\sigma}^2_v$ underestimate $\sigma^2_v$ slightly when $\sigma^2_v$ is moderate or large relative to $\sigma^2_i$'s. (2) The relative biases decrease as $\sigma^2_v$ increases, while they increase as the heteroscedasticity increases. (3) Both jackknife estimators of $\sigma^2_v$ reduces bias. The truncated jackknife estimator $\hat{\sigma}^2_{vJ}$ is almost unbiased for all cases, while the untruncated jackknife estimator $\hat{\sigma}^2_{vJ}$ is almost unbiased only for moderate or large $\sigma^2_v$ relative to $\sigma^2_i$'s.

Tables 4.2 and 4.3 report the values of coverage probabilities (CP) and lengths of confidence intervals (LCI) for nominal level 90% associated with four sets of estimators of $\sigma^2_v$ and corresponding variance estimators. We have the following results: (1) The confidence intervals using truncated variance estimator $\hat{V}_a$ with truncated AMINQUE estimator $\hat{\sigma}^2_v$ or using corresponding truncated jackknife estimator $\hat{V}_{/h}$ with $\hat{\sigma}^2_{/h}$ fail to give good coverage when $\sigma^2_v$ is small relative to $\sigma^2_i$'s. From the lengths of confidence intervals, we can see that the first one severely overcovers $\sigma^2_v$ due to truncation, and the second one seriously undercovers $\sigma^2_v$. (2) The untruncated estimator $\hat{\sigma}^2_v$ and the corresponding variance estimator $\hat{V}_a$ by substitution lead to undercoverage which increases with $\sigma^2_v$. This undercoverage is more severe in the case of double exponential distribution with long tails than in the case of normal distribution. (3) The untruncated jackknife estimator $\hat{\sigma}^2_{v(J)}$ and the variance esti-
mator \( \hat{V}_J \) give reasonably good coverage in the case of normal distribution when \( \sigma^2_v \) is small (\( \leq 0.5 \)), but lead to undercoverage as \( \sigma^2_v \) increases. The undercoverage is again more severe in the case of double exponential distribution.

Although our methods for inference on \( \sigma^2_v \) are asymptotically correct, the above empirical results suggest the need for alternative methods using suitable transformations on \( \sigma^2_v \) to reduce the undercoverage in finite samples, especially for long-tailed distributions and larger \( \sigma^2_v \).

4.5.3 Empirical results on estimating \( \beta \)

Table 4.4 reports the percentage efficiency of our WLS estimator \( \hat{\beta}_w \) relative to the alternative WLS estimator \( \hat{\beta}_{sw} \), and Shao and Rao's (1993) WLS estimator \( \hat{\beta}_{w*} \). We have the following results from Tables 4.4: (1) Our estimator \( \hat{\beta}_w \) is substantially more efficient than the alternative estimator \( \hat{\beta}_{sw} \) when \( \sigma^2_v \) is small or moderate relative to \( \sigma^2 \). (2) In the case of normal distribution, our estimator \( \hat{\beta}_w \), compared to Shao and Rao’s estimator \( \hat{\beta}_{w*} \), gains efficiency when \( \sigma^2_v \) is relatively moderate or large, and loses efficiency when \( \sigma^2_v \) is relatively small. This result seems to be reversed in the case of double exponential distribution. But either gain or loss in efficiency is very small (\( \leq 10\% \) for most cases).

Table 4.5 reports coverage probabilities (CP) of confidence intervals from the above procedures of estimating \( \beta \) for nominal level 90\%. Table 4.6 reports lengths of these confidence intervals (LCI). We have the following results from these two tables: (1) Compared to Shao and Rao’s (1993) WLS estimator \( \hat{\beta}_{w*} \) and associated delete-group jackknife variance estimator \( \hat{V}_{f*} \), our method based on WLS estimator \( \hat{\beta}_w \) and associated substitution variance estimator \( \hat{V}_h \) gives more accurate and higher coverage probabilities with shorter or comparable lengths of confidence intervals (LCI). Shao and Rao’s method leads to slight undercoverage (\(< 5\% \)) for most cases. (2) Our jackknife variance estimator \( \hat{V}_J \) leads to slightly higher CP but wider LCI compared to \( \hat{V}_h \). (3) The alternative substitution variance estimator, \( \hat{V}_{sh} \), performs well with respect to CP, but leads to wider LCI compared to \( \hat{V}_h \) or \( \hat{V}_J \) when \( \sigma^2_v \) is small or moderate relative to \( \sigma^2 \)'s and errors follow normal distribution or for all \( \sigma^2 \).
CHAPTER 4. JACKKNIFE ESTIMATORS

in the case of double exponential distribution.

Table 4.7 reports the values of relative biases (RB). We have the following results from Table 4.7: (1) Under normal distribution, relative biases for all variance estimators except Shao and Rao’s \( \hat{V}^{*}_{j} \) decrease as \( \sigma_v^2 \) increases. The relative biases of our variance estimators \( \hat{V}_k, \hat{V}_j \) and \( \hat{V}_{sk} \) are small (\(|RB| < 10\%\)) for moderate or large \( \sigma_v^2 \) relative to \( \sigma_v^2 \)'s, while they are over 20\% for some cases when \( \sigma_v^2 \) is relatively small. (2) For all cases, Shao and Rao’s (1993) variance estimator has a positive bias (ranging from 5\% to 15\% under normal distribution, and from 15\% to 20\% under exponential distribution).

Table 4.8 reports the values of coefficients of variation (CV) of all the variance estimators. We have the following results from this table on the stability of these variance estimators as measured by CV: (1) Our jackknife variance estimator \( \hat{V}_j \) is always less stable than our substitution variance estimator \( \hat{V}_k \), especially for relatively small or moderate \( \sigma_v^2 \). (2) The alternative substitution variance estimator \( \hat{V}_{sk} \) is less stable than \( \hat{V}_k \), especially for small or moderate \( \sigma_v^2 \). (3) Shao and Rao’s variance estimator \( \hat{V}^{*}_{j} \) is less stable than \( \hat{V}_k \), especially as \( \sigma_v^2 \) increases.

Overall, our substitution variance estimator \( \hat{V}_k \) performed better than the other variance estimators \( \hat{V}_j \) and \( \hat{V}_{sk} \) as well as Shao and Rao’s (1993) variance estimator \( \hat{V}^{*}_{j} \). The delete-group jackknife variance estimator \( \hat{V}_j \) is less stable than \( \hat{V}_k \), but it is robust to mis-specification of the covariance matrix of the within group observations, unlike \( \hat{V}_k \).
Table 4.1: Relative Biases (%) for Estimators of $\sigma_v^2$

<table>
<thead>
<tr>
<th>$\sigma_v^2$</th>
<th>$\sigma_i^2$</th>
<th>Normal</th>
<th>Double Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\sigma}_v^2$</td>
<td>$\hat{\sigma}_i^2$</td>
<td>$\hat{\sigma}_v^2$</td>
</tr>
<tr>
<td>0.1</td>
<td>(i)</td>
<td>10.0</td>
<td>-15.3</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>10.6</td>
<td>-16.3</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>43.3</td>
<td>-24.1</td>
</tr>
<tr>
<td>0.25</td>
<td>(i)</td>
<td>-6.0</td>
<td>-8.4</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>-6.0</td>
<td>-8.8</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>0.4</td>
<td>-12.0</td>
</tr>
<tr>
<td>0.5</td>
<td>(i)</td>
<td>-6.2</td>
<td>-6.4</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>-6.2</td>
<td>-6.6</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>-6.6</td>
<td>-8.0</td>
</tr>
<tr>
<td>2</td>
<td>(i)</td>
<td>-4.9</td>
<td>-4.9</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>-4.9</td>
<td>-4.9</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>-5.2</td>
<td>-5.2</td>
</tr>
</tbody>
</table>

(1) Patterns for $\sigma_i^2$: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{\sigma}_v^2 =$ truncated AMINQU estimator.
$\hat{\sigma}_i^2 =$ untruncated AMINQU estimator.
$\hat{\sigma}_v(J) =$ jackknife estimator corresponding $\hat{\sigma}_v^2$.
$\hat{\sigma}_i(J) =$ jackknife estimator corresponding $\hat{\sigma}_i^2$. 
Table 4.2: Coverage Probability (CP %) and Length of Confidence Interval (LCI) for Estimators of \( \sigma^2 \) at Level 1 - \( \alpha = 90\% \), when Errors Follow Normal Distribution

<table>
<thead>
<tr>
<th>( \sigma^2 )</th>
<th>( \sigma_i^2 )</th>
<th>(( \hat{\sigma}^2_v ), ( \hat{V}_a ))</th>
<th>(( \hat{\sigma}^2_v ), ( \hat{V}_a ))</th>
<th>(( \hat{\sigma}^2_{v(J)} ), ( \hat{V}_{Jh} ))</th>
<th>(( \hat{\sigma}^2_{v(J)} ), ( \hat{V}_{Jt} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>(i)</td>
<td>98.0 (0.242)</td>
<td>87.0 (0.241)</td>
<td>62.2 (0.184)</td>
<td>89.3 (0.251)</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>98.0 (0.249)</td>
<td>87.3 (0.248)</td>
<td>61.3 (0.187)</td>
<td>89.3 (0.258)</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>97.9 (0.409)</td>
<td>87.9 (0.407)</td>
<td>54.9 (0.267)</td>
<td>88.9 (0.423)</td>
</tr>
<tr>
<td>0.25</td>
<td>(i)</td>
<td>88.4 (0.289)</td>
<td>86.4 (0.289)</td>
<td>82.0 (0.278)</td>
<td>87.9 (0.302)</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>88.9 (0.297)</td>
<td>86.4 (0.297)</td>
<td>81.5 (0.283)</td>
<td>88.3 (0.311)</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>96.2 (0.450)</td>
<td>86.1 (0.449)</td>
<td>69.3 (0.378)</td>
<td>88.0 (0.468)</td>
</tr>
<tr>
<td>0.5</td>
<td>(i)</td>
<td>85.1 (0.376)</td>
<td>85.1 (0.376)</td>
<td>86.9 (0.390)</td>
<td>87.3 (0.394)</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>85.4 (0.385)</td>
<td>85.4 (0.385)</td>
<td>87.1 (0.398)</td>
<td>87.4 (0.403)</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>86.7 (0.526)</td>
<td>85.3 (0.526)</td>
<td>83.0 (0.519)</td>
<td>87.6 (0.550)</td>
</tr>
<tr>
<td>2</td>
<td>(i)</td>
<td>82.9 (0.942)</td>
<td>82.9 (0.942)</td>
<td>84.9 (0.990)</td>
<td>84.9 (0.990)</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>83.0 (0.953)</td>
<td>83.0 (0.953)</td>
<td>85.2 (1.002)</td>
<td>85.2 (1.002)</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>83.6 (1.062)</td>
<td>83.6 (1.062)</td>
<td>86.1 (1.115)</td>
<td>86.1 (1.115)</td>
</tr>
</tbody>
</table>

(1) Note: The pair \( (a, b) \) implies that the quantities in the columns are obtained by using \( a \) as estimator of \( \sigma^2 \), and \( b \) as estimator of asymptotic variance.

(2) Patterns for \( \sigma_i^2 \): (i) \( (1/2, 1, 2) \); (ii) \( (1/4, 1, 2) \); (iii) \( (1/4, 1, 4) \).

(3) \( \hat{\sigma}^2_v \), \( \hat{\sigma}^2_v \), \( \hat{\sigma}^2_{v(J)} \) and \( \hat{\sigma}^2_{v(J)} \) are the same those in Table 4.1.

(4) \( \hat{V}_a \) = substitution estimator of \( V_a \) using \( \hat{\sigma}^2 \).
\( \hat{V}_a \) = substitution estimator of \( V_a \) using \( \hat{\sigma}^2 \).
\( \hat{V}_{Jh} \) = truncated jackknife variance estimator.
\( \hat{V}_{Jt} \) = untruncated jackknife variance estimator.
Table 4.3: Coverage Probability (CP %) and Length of Confidence Interval (LCI) for estimators of $\sigma^2_v$ at level $1 - \alpha = 90\%$, when Errors follow Double Exponential Distribution

<table>
<thead>
<tr>
<th>$\sigma^2_v$</th>
<th>$\sigma^2_t$</th>
<th>$(\hat{\sigma}^2_v, \hat{V}_a)$</th>
<th>$(\hat{\sigma}^2_v, \hat{V}_a)$</th>
<th>$(\hat{\sigma}^2_v(J), \hat{V}_{Jh})$</th>
<th>$(\hat{\sigma}^2_v(J), \hat{V}_{Jl})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1/2, 1, 2)</td>
<td>0.1</td>
<td>96.1 (0.231)</td>
<td>86.2 (0.230)</td>
<td>62.7 (0.176)</td>
<td>88.4 (0.240)</td>
</tr>
<tr>
<td>(1/4, 1, 2)</td>
<td></td>
<td>96.2 (0.237)</td>
<td>85.8 (0.236)</td>
<td>62.3 (0.179)</td>
<td>88.0 (0.246)</td>
</tr>
<tr>
<td>(1/4, 1, 4)</td>
<td></td>
<td>97.6 (0.380)</td>
<td>88.2 (0.378)</td>
<td>55.3 (0.251)</td>
<td>90.8 (0.393)</td>
</tr>
<tr>
<td>(1/2, 1, 2)</td>
<td>0.25</td>
<td>84.7 (0.294)</td>
<td>82.1 (0.294)</td>
<td>77.6 (0.281)</td>
<td>84.5 (0.308)</td>
</tr>
<tr>
<td>(1/4, 1, 2)</td>
<td></td>
<td>85.3 (0.301)</td>
<td>82.5 (0.301)</td>
<td>77.2 (0.285)</td>
<td>85.5 (0.314)</td>
</tr>
<tr>
<td>(1/4, 1, 4)</td>
<td></td>
<td>93.1 (0.434)</td>
<td>86.1 (0.433)</td>
<td>69.4 (0.364)</td>
<td>88.1 (0.451)</td>
</tr>
<tr>
<td>(1/2, 1, 2)</td>
<td>0.5</td>
<td>79.6 (0.416)</td>
<td>79.4 (0.416)</td>
<td>80.7 (0.429)</td>
<td>81.7 (0.436)</td>
</tr>
<tr>
<td>(1/4, 1, 2)</td>
<td></td>
<td>80.8 (0.423)</td>
<td>80.6 (0.423)</td>
<td>81.7 (0.434)</td>
<td>82.8 (0.443)</td>
</tr>
<tr>
<td>(1/4, 1, 4)</td>
<td></td>
<td>84.5 (0.542)</td>
<td>82.2 (0.542)</td>
<td>79.5 (0.528)</td>
<td>84.0 (0.566)</td>
</tr>
<tr>
<td>(1/2, 1, 2)</td>
<td>2</td>
<td>77.5 (1.226)</td>
<td>77.5 (1.226)</td>
<td>79.9 (1.289)</td>
<td>79.9 (1.289)</td>
</tr>
<tr>
<td>(1/4, 1, 2)</td>
<td></td>
<td>77.9 (1.232)</td>
<td>77.9 (1.232)</td>
<td>80.3 (1.295)</td>
<td>80.3 (1.295)</td>
</tr>
<tr>
<td>(1/4, 1, 4)</td>
<td></td>
<td>76.6 (1.316)</td>
<td>76.6 (1.316)</td>
<td>80.1 (1.382)</td>
<td>80.1 (1.382)</td>
</tr>
</tbody>
</table>

(1) Note: The pair (a, b) implies that the quantities in the columns are obtained by using a as estimator of $\sigma^2_v$, and b as estimator of asymptotic variance.

(2) Patterns for $\sigma^2_t$: (i) (1/2, 1, 2); (ii) (1/4, 1, 2); (iii) (1/4, 1, 4).

(3) $\hat{\sigma}^2_v$, $\hat{\sigma}^2_v$, $\hat{\sigma}^2_v(J)$ and $\hat{\sigma}^2_v(J)$ are the same those in Table 4.1.

(4) $\hat{V}_a = \text{substitution estimator of } V_a$ using $\hat{\sigma}^2_v$.

$\hat{V}_a = \text{substitution estimator of } V_a$ using $\hat{\sigma}^2_v$.

$\hat{V}_{Jh} = \text{truncated jackknife variance estimator}$.

$\hat{V}_{Jl} = \text{untruncated jackknife variance estimator}$. 
Table 4.4: Percentage Efficiency of $\hat{\beta}_w$ Relative to $\hat{\beta}_{sw}$ and $\hat{\beta}_{rs}^w$

<table>
<thead>
<tr>
<th>$\sigma^2_v$</th>
<th>$\sigma^2_\epsilon$</th>
<th>Normal</th>
<th>Double Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E(\hat{\beta}<em>w, \hat{\beta}</em>{sw})$</td>
<td>$E(\hat{\beta}<em>w, \hat{\beta}</em>{rs}^w)$</td>
<td>$E(\hat{\beta}<em>w, \hat{\beta}</em>{sw})$</td>
</tr>
<tr>
<td>0.1</td>
<td>(i) 191</td>
<td>89</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>(ii) 192</td>
<td>84</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>(iii) 213</td>
<td>97</td>
<td>188</td>
</tr>
<tr>
<td>0.25</td>
<td>(i) 121</td>
<td>102</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>(ii) 124</td>
<td>99</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>(iii) 181</td>
<td>103</td>
<td>167</td>
</tr>
<tr>
<td>0.5</td>
<td>(i) 105</td>
<td>108</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>(ii) 105</td>
<td>105</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>(iii) 114</td>
<td>107</td>
<td>121</td>
</tr>
<tr>
<td>2</td>
<td>(i) 100</td>
<td>108</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>(ii) 100</td>
<td>109</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>(iii) 101</td>
<td>108</td>
<td>110</td>
</tr>
</tbody>
</table>

(1) Patterns for $\sigma^2_\epsilon$: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{\beta}_w = \text{AMINQU WLS estimator}$.

$\hat{\beta}_{sw} = \text{alternative WLS estimator using sample variance } s^2_\epsilon \text{ as estimator of } \sigma^2_\epsilon$.

$\hat{\beta}_{rs}^w = \text{Shao and Rao's (1993) WLS estimator}$.
Table 4.5: Coverage Probability (CP %) of Confidence Intervals on $\beta$ for Four Variance Estimators at level $1 - \alpha = 90\%$

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$\sigma^2$</th>
<th>$\hat{V}^*$</th>
<th>$\hat{V}_k$</th>
<th>$\hat{V}_j$</th>
<th>$\hat{V}_{sk}$</th>
<th>$\hat{V}^*$</th>
<th>$\hat{V}_k$</th>
<th>$\hat{V}_j$</th>
<th>$\hat{V}_{sk}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>(i)</td>
<td>86.8</td>
<td>90.8</td>
<td>91.4</td>
<td>90.9</td>
<td>87.9</td>
<td>90.7</td>
<td>90.4</td>
<td>89.6</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>87.0</td>
<td>90.5</td>
<td>90.3</td>
<td>90.1</td>
<td>87.7</td>
<td>89.4</td>
<td>89.6</td>
<td>89.9</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>87.8</td>
<td>91.0</td>
<td>91.8</td>
<td>91.0</td>
<td>88.3</td>
<td>90.6</td>
<td>90.9</td>
<td>89.4</td>
</tr>
<tr>
<td>0.25</td>
<td>(i)</td>
<td>86.1</td>
<td>89.0</td>
<td>91.0</td>
<td>90.1</td>
<td>87.8</td>
<td>91.1</td>
<td>92.1</td>
<td>91.4</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>86.6</td>
<td>89.4</td>
<td>91.1</td>
<td>90.2</td>
<td>87.4</td>
<td>90.8</td>
<td>92.2</td>
<td>91.0</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>87.5</td>
<td>89.6</td>
<td>91.7</td>
<td>90.6</td>
<td>87.4</td>
<td>91.5</td>
<td>92.1</td>
<td>91.2</td>
</tr>
<tr>
<td>0.5</td>
<td>(i)</td>
<td>86.9</td>
<td>88.6</td>
<td>88.9</td>
<td>89.1</td>
<td>88.2</td>
<td>91.2</td>
<td>91.3</td>
<td>91.8</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>86.7</td>
<td>88.7</td>
<td>89.6</td>
<td>89.2</td>
<td>87.9</td>
<td>91.8</td>
<td>92.0</td>
<td>91.7</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>86.3</td>
<td>88.7</td>
<td>89.8</td>
<td>89.9</td>
<td>87.7</td>
<td>91.9</td>
<td>92.7</td>
<td>92.0</td>
</tr>
<tr>
<td>2</td>
<td>(i)</td>
<td>84.1</td>
<td>87.9</td>
<td>88.1</td>
<td>87.5</td>
<td>89.3</td>
<td>91.5</td>
<td>92.1</td>
<td>91.6</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>84.7</td>
<td>87.5</td>
<td>88.0</td>
<td>87.5</td>
<td>88.5</td>
<td>91.6</td>
<td>92.1</td>
<td>91.7</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>84.0</td>
<td>88.1</td>
<td>88.3</td>
<td>88.1</td>
<td>88.5</td>
<td>91.9</td>
<td>92.4</td>
<td>92.1</td>
</tr>
</tbody>
</table>

(1) Patterns for $\sigma^2$: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{V}^*$ = the variance estimators by delete-group jackknifing WLS in Shao and Rao (1993).

$\hat{V}_k$ = the substitution variance estimator of WLS estimator $\hat{\beta}_w$.

$\hat{V}_j$ = the delete-group jackknife variance estimator of WLS estimator $\hat{\beta}_w$.

$\hat{V}_{sk}$ = the substitution variance estimator of WLS estimator $\hat{\beta}_w$. 


Table 4.6: Average Lengths of Confidence Intervals (LCI) for Four Variance Estimators at Level $1 - \alpha = 90\%$

<table>
<thead>
<tr>
<th>$\sigma^2_v$</th>
<th>$\sigma_i^2$</th>
<th>Normal</th>
<th>Double Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{V}^{*}$</td>
<td>$\hat{V}_k$</td>
<td>$\hat{V}_J$</td>
</tr>
<tr>
<td>0.1</td>
<td>(i) 0.414</td>
<td>0.437</td>
<td>0.469</td>
</tr>
<tr>
<td></td>
<td>(ii) 0.446</td>
<td>0.487</td>
<td>0.520</td>
</tr>
<tr>
<td></td>
<td>(iii) 0.462</td>
<td>0.472</td>
<td>0.511</td>
</tr>
<tr>
<td>0.25</td>
<td>(i) 0.493</td>
<td>0.486</td>
<td>0.512</td>
</tr>
<tr>
<td></td>
<td>(ii) 0.517</td>
<td>0.519</td>
<td>0.549</td>
</tr>
<tr>
<td></td>
<td>(iii) 0.540</td>
<td>0.540</td>
<td>0.577</td>
</tr>
<tr>
<td>0.5</td>
<td>(i) 0.594</td>
<td>0.570</td>
<td>0.583</td>
</tr>
<tr>
<td></td>
<td>(ii) 0.613</td>
<td>0.590</td>
<td>0.605</td>
</tr>
<tr>
<td></td>
<td>(iii) 0.637</td>
<td>0.618</td>
<td>0.641</td>
</tr>
<tr>
<td>2</td>
<td>(i) 0.987</td>
<td>0.935</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>(ii) 0.999</td>
<td>0.943</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>(iii) 1.022</td>
<td>0.973</td>
<td>0.984</td>
</tr>
</tbody>
</table>

(1) Patterns for $\sigma_i^2$: (i) (1/2, 1, 2); (ii) (1/4, 1, 2); (iii) (1/4, 1, 4).

(2) $\hat{V}^{*}$: the variance estimators by delete–group jackknifing WLS in Shao and Rao (1993).

$\hat{V}_k$ = the substitution variance estimator of WLS estimator $\hat{\beta}_w$.

$\hat{V}_J$ = the delete–group jackknife variance estimator of WLS estimator $\hat{\beta}_w$.

$\hat{V}_{sk}$ = the substitution variance estimator of WLS estimator $\hat{\beta}_{sw}$.
Table 4.7: Relative Biases (RB %) OF Four Variance Estimators

<table>
<thead>
<tr>
<th>$\sigma_i^2$</th>
<th>Normal</th>
<th>Double Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_i^2$</td>
<td>$\hat{V}_j^*$ $\hat{V}_k$ $\hat{V}<em>j$ $\hat{V}</em>{sk}$</td>
<td>$\hat{V}_j^*$ $\hat{V}_k$ $\hat{V}<em>j$ $\hat{V}</em>{sk}$</td>
</tr>
<tr>
<td>(i) 0.1</td>
<td>15.12  10.63  32.33  6.60</td>
<td>15.26  10.77  26.51  3.23</td>
</tr>
<tr>
<td>(ii)</td>
<td>2.06   9.62   29.14  5.52</td>
<td>12.17  6.76   25.51  0.53</td>
</tr>
<tr>
<td>(iii)</td>
<td>18.26  15.77  39.70  9.29</td>
<td>18.27  20.47  37.58  7.43</td>
</tr>
<tr>
<td>(i) 0.25</td>
<td>10.67  2.26   15.39  3.34</td>
<td>16.56  9.26   18.93  9.88</td>
</tr>
<tr>
<td>(ii)</td>
<td>8.10   1.97   16.20  2.86</td>
<td>14.93  7.88   18.12  8.47</td>
</tr>
<tr>
<td>(iii)</td>
<td>12.75  8.96   28.82  4.91</td>
<td>18.53  10.94  24.80  2.43</td>
</tr>
<tr>
<td>(i) 0.5</td>
<td>8.32   -1.88  3.87   -3.19</td>
<td>18.65  7.06   10.69  4.97</td>
</tr>
<tr>
<td>(ii)</td>
<td>7.90   -2.21  4.33   -3.17</td>
<td>16.71  6.06   10.48  4.30</td>
</tr>
<tr>
<td>(iii)</td>
<td>8.99   1.59   11.27  4.87</td>
<td>19.58  9.60   14.99  9.61</td>
</tr>
<tr>
<td>(i) 2</td>
<td>5.36   -6.30  -4.03  -7.63</td>
<td>19.81  6.98   8.70   4.21</td>
</tr>
<tr>
<td>(ii)</td>
<td>4.71   -6.46  -4.13  -7.74</td>
<td>19.57  6.78   8.54   4.13</td>
</tr>
<tr>
<td>(iii)</td>
<td>6.49   -5.01  -2.82  -6.41</td>
<td>20.48  7.69   9.42   4.89</td>
</tr>
</tbody>
</table>

(1) Patterns for $\sigma_i^2$: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{V}_j^*$ = the variance estimators by delete-group jackknifing WLS in Shao and Rao (1993).

$\hat{V}_k$ = the substitution variance estimator of WLS estimator $\hat{\beta}_w$.

$\hat{V}_j$ = the delete-group jackknife variance estimator of WLS estimator $\hat{\beta}_w$.

$\hat{V}_{sk}$ = the substitution variance estimator of WLS estimator $\hat{\beta}_{sw}$. 
### Table 4.8: Coefficient of Variation (CV %) of Four Variance Estimators

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Double Exponential</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
<td>$\hat{V}_k$</td>
<td>$\hat{V}_j$</td>
<td>$\hat{V}_{sk}$</td>
<td>$\hat{V}_k$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>74.70</td>
<td>59.83</td>
<td>104.22</td>
<td>215.24</td>
<td>80.82</td>
<td>59.40</td>
</tr>
<tr>
<td>(ii)</td>
<td>68.59</td>
<td>60.24</td>
<td>98.10</td>
<td>212.80</td>
<td>73.24</td>
<td>56.18</td>
</tr>
<tr>
<td>(iii)</td>
<td>80.48</td>
<td>65.84</td>
<td>106.83</td>
<td>216.78</td>
<td>88.84</td>
<td>78.31</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>70.16</td>
<td>35.57</td>
<td>64.25</td>
<td>126.78</td>
<td>82.95</td>
<td>44.90</td>
</tr>
<tr>
<td>(ii)</td>
<td>65.93</td>
<td>37.47</td>
<td>65.85</td>
<td>131.96</td>
<td>76.64</td>
<td>43.54</td>
</tr>
<tr>
<td>(iii)</td>
<td>74.18</td>
<td>51.02</td>
<td>98.70</td>
<td>227.63</td>
<td>88.53</td>
<td>52.74</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>67.93</td>
<td>26.58</td>
<td>45.96</td>
<td>39.26</td>
<td>85.72</td>
<td>38.11</td>
</tr>
<tr>
<td>(ii)</td>
<td>63.59</td>
<td>26.73</td>
<td>56.95</td>
<td>37.84</td>
<td>81.79</td>
<td>36.80</td>
</tr>
<tr>
<td>(iii)</td>
<td>70.94</td>
<td>32.77</td>
<td>55.37</td>
<td>132.16</td>
<td>87.52</td>
<td>43.73</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>67.66</td>
<td>25.79</td>
<td>26.12</td>
<td>25.42</td>
<td>90.57</td>
<td>39.88</td>
</tr>
<tr>
<td>(ii)</td>
<td>64.07</td>
<td>25.78</td>
<td>26.09</td>
<td>25.40</td>
<td>88.88</td>
<td>39.34</td>
</tr>
<tr>
<td>(iii)</td>
<td>70.68</td>
<td>26.01</td>
<td>26.50</td>
<td>25.63</td>
<td>93.60</td>
<td>40.21</td>
</tr>
</tbody>
</table>

1. Patterns for $\sigma^2$: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

\( \hat{V}_k \) = the substitution variance estimator of WLS estimator $\hat{\beta}_u$.

\( \hat{V}_j \) = the delete-group jackknife variance estimator of WLS estimator $\hat{\beta}_u$.

\( \hat{V}_{sk} \) = the substitution variance estimator of WLS estimator $\hat{\beta}_u$. 
Chapter 5

Two-Fold Nested Error Regression Models

In this chapter, we extend the approximate minimum norm quadratic unbiased (AMINQU) estimation in Chapter 2 to the two-fold nested-error regression model. The model is briefly introduced in Section 5.1. The formula for the WLS estimator of the vector of regression parameters is derived assuming all variance components are known.

In Section 5.2, we present several estimators of variance components using minimum norm quadratic unbiased (MINQU) estimation, average of squared residuals (ASR) and weighted least squares (WLS) methods. The AMINQU estimators of variance components are then derived. Due to complexity of the model, explicit expressions of AMINQU estimators are obtained only for one special set of prior values. These AMINQU estimators of variance components (except error variances) are shown to be asymptotically consistent and normally distributed. Consistent estimators of these asymptotic variances are obtained by the substitution method.

Since the AMINQU estimators of variance components under the two-fold nested-error model do not guarantee positive estimators of weights for WLS estimation of regression parameters, we adopt the approach of Section 3.4 in Section 5.3, i.e., use within group sample variances as estimators of random error variances. The resulting WLS estimator of regression parameters is shown to be asymptotically normal. A consistent estimator of the asymptotic covariance matrix is also obtained by the
substitution method.

5.1 Description of the Model

We consider the two-fold nested error regression model with unequal error variances:

\[ y_{ijk} = x_i' \beta + \varepsilon_{ijk} \quad \text{with} \quad \varepsilon_{ijk} = v_i + u_{ij} + e_{ijk}, \quad (5.1) \]

\[ i = 1, \ldots, m, \quad j = 1, \ldots, m_i, \quad k = 1, \ldots, n_{ij}, \]

\[ \sum_{j=1}^{m_i} n_{ij} = n_i, \quad \sum_{i=1}^{m} n_i = n, \]

where \( y_{ijk} \) is the \( k \)-th observation in the \( j \)-th subgroup of the \( i \)-th group, \( x_i' = (x_{i1}, x_{i2}, \ldots, x_{ip}) \) is a vector of known constants, \( \beta' = (\beta_1, \ldots, \beta_p) \) is a vector of unknown regression parameters, \( v_i \)'s are independent identically distributed random variables with mean 0 and variance \( \sigma_v^2 \) (or \( v_i \overset{iid}{\sim} (0, \sigma_v^2) \)), \( u_{ij} \overset{iid}{\sim} (0, \sigma_u^2) \) for all \( i \) and \( j \). Also, for given \( i \) and \( j \), \( e_{ijk} \overset{iid}{\sim} (0, \sigma_e^2) \) for all \( k \), and \( v_i \)'s, \( u_{ij} \) and \( e_{ijk} \)'s are assumed to be mutually independent. Note that \( k \) is used as an index in this chapter, unlike the rest of the thesis. Let

\[ y_{ij} = (y_{ij1}, \ldots, y_{ijn_j})', \]

\[ \varepsilon_{ij} = (\varepsilon_{ij1}, \ldots, \varepsilon_{ijn_j})', \]

\[ \mathbf{Y} = (y'_{11} : \cdots : y'_{1m_1} : \cdots : y'_{m_1} : \cdots : y'_{mn_m})', \]

\[ \mathbf{X} = (x_{11}' : x_2' : \cdots : x_{m1}' : \cdots : x_{mn_m}')', \]

\[ \mathbf{E} = (e_{11}' : \cdots : e_{1n_1}' : \cdots : e_{m1}' : \cdots : e_{mn_m}'). \]

where \( 1_m \) is a \( m \times 1 \) vector with all elements equal 1. A matrix form of model (5.1) is then given by

\[ \mathbf{Y} = \mathbf{X} \beta + \mathbf{E}, \quad (5.2) \]

with dispersion matrix

\[ V = \text{Var}(\mathbf{E}) = \text{block diag}_i(V_i), \]
where

\[ V_i = \sigma_i^2 I_{n_i} + \sigma_u^2 \text{ block diag}_i (J_{n_i}) + \sigma_v^2 J_{n_i} = \text{ block diag}_i (\sigma_i^2 I_{n_i} + \sigma_u^2 J_{n_i}) + \sigma_v^2 J_{n_i} \]

with \( J_m = I_m - 1_m \).

Our principle interest is the estimation of \( \sigma_i^2 \), \( \sigma_u^2 \) and the regression parameters \( \beta \). For simplicity, \( \mathcal{X} \) is assumed to be of full column rank (and so is the matrix \( \mathcal{X} \) defined below). If \( \sigma_i^2 \), \( \sigma_u^2 \) and \( \sigma_v^2 \)'s are known, the following weighted least squares (WLS) estimator is usually used for estimating regression coefficient \( \beta \):

\[
\hat{\beta}_w = (\mathcal{X}' V^{-1} \mathcal{X})^{-1} \mathcal{X} V^{-1} y, \tag{5.3}
\]

where \( V^{-1} = \text{ block diag}_i (V_i^{-1}) \). We evaluate \( V_i^{-1} \) using the following lemma repeatedly.

**Lemma 5.1** Assume that matrices \( A, B \) and \( C \) are of proper orders. If both \( A \) and \( A + BCB' \) are non-singular, then

\[
(A + BCB')^{-1} = A^{-1} - A^{-1} BC(I + B'A^{-1}BC)^{-1} B'A^{-1}.
\]

We obtain, after considerable simplifications, that

\[
V_i^{-1} = C_i - w_i d_i d_i' / \left( \sum_{j=1}^{m_i} w_{ij} \right), \tag{5.4}
\]

where

\[
C_i^{-1} = \text{ block diag}_i \left( \sigma_i^{-2} (I_{n_i} - (\sigma_u^2 / \sigma_i^2) J_{n_i}) \right),
\]

\[
d_i = \text{ col}_{1 \leq j \leq m_i} \left( \sigma_i^{-1} w_{ij} 1_{n_i} \right),
\]

\[
w_{ij}^{-1} = \sigma_u^2 + \sigma_v^2 / n_i,
\]

\[
w_i = \frac{\sum_{j} w_{ij}}{1 + \sigma_v^2 \sum_{j} w_{ij}}.
\]

Using (5.4) in (5.3), \( \hat{\beta}_w \) is written as

\[
\hat{\beta}_w = \left( \sum_i w_i x_i x_i' \right)^{-1} \sum_i w_i x_i \bar{y}_i. \tag{5.5}
\]
where
\[
\bar{y}_{ij} = \frac{\sum_{k=1}^{n_{ij}} y_{ijk}}{n_i}
\]
\[
\bar{y}_{i..} = \frac{\sum_{j=1}^{m_i} w_{ij} \bar{y}_{ij}}{\sum_{j=1}^{m_i} w_{ij}}.
\]

Since \( \sigma_v^2, \sigma_u^2, \) and \( \sigma_i^2 \) are not observable in practice, the problem is how to estimate these variances. We'll consider some estimation procedures in the next section.

P. S. R. S. Rao and Heckler (1992) considered a model similar to (5.1) with \( x_i' \beta = \mu \) (but they do not require the common variance \( \sigma_v^2 \) and common \( \sigma_i^2 \) within \( i \)-th group.) In their paper, they derived estimators of variance components using ANOVA, USS, MINQUE, ASR and their so-called weighted average of means (WAM) procedures and the maximum likelihood estimation procedure. They also considered these two methods in C. R. Rao (1984) for obtaining non-negative quadratic estimators. They computed the empirical variance (or MSE) of variance estimators using Monte Carlo simulation.

### 5.2 Estimation of \( \sigma_v^2 \) and \( \sigma_u^2 \)

In this section some estimation procedures for estimating variance components are presented. The MINQUE estimation procedure, ASR procedure and weighted least squares (WLS) are derived in Section 5.2.1. ANOVA estimators can be seen as a special case of WLS estimators for the variance components. The estimator from the AMINQUE procedure is derived in Section 5.2.2. Explicit representation of estimators of \( \sigma_v^2, \sigma_u^2 \) and \( \sigma_i^2 \) are given for a special set of prior values for these variances. The estimators of \( \sigma_v^2 \) and \( \sigma_u^2 \) are shown to be consistent and asymptotically normally distributed in Section 5.2.3. Estimators of these asymptotic variances are obtained by the substitution method in Section 5.2.4.
5.2.1 MINQU, ASR and WLS estimators

The linear model (5.2) can be written as

$$\mathbf{Y} = \mathbf{X}\beta + U_v\xi_v + U_u\xi_u + U_1\xi_1 + \cdots + U_m\xi_m$$

where

$$U_i' = (0_{n_i} \times \sum_{i'=1}^{n_{i-1}} n_{i'} : I_{n_i} : 0_{n_i} \times \sum_{i'=m+1}^{m} n_{i'})_{n_i \times n},$$

$$U_v = \text{block diag}_v(1_{n_v}), \quad U_u = \text{block diag}_u(\text{block diag}_v(1_{n_u})).$$

$\xi_i$ is an $n_i$-random vector with mean 0 and dispersion matrix $\sigma_i^2 I_{n_i}$, $\xi_v$ is an $m$-random vector with mean 0 and dispersion matrix $\sigma_v^2 I_m$, and $\xi_u$ is an $\sum_{i=1}^{m} m_i = m$-random vector with mean 0 and dispersion matrix $\sigma_u^2 I_{m_u}$.

Let $\sigma_{v0}^2$, $\sigma_{u0}^2$ and $\sigma_i^2$ be prior values of $\sigma_v^2$, $\sigma_u^2$ and $\sigma_i^2$ respectively. And denote

$$w_{i0} = (\sigma_{v0}^2 + \sigma_{u0}^2/n_{ij})^{-1}, \quad w_{i0} = \sum_{j=1}^{m_i} w_{ij0},$$

$$w_0 = w_{i0}/(1 + \sigma_{v0}^2 w_{i0}),$$

$$\tilde{y}_{i0} = \sum_{j=1}^{m_i} w_{ij0}\tilde{y}_{ij}/w_{i0}, \quad \tilde{y}_{i0} = (\mathbf{x}' V_0^{-1} \mathbf{x})^{-1} \mathbf{x}' V_0^{-1} \mathbf{y},$$

$$V_0 = \sigma_{10}^2 F_1 + \cdots + \sigma_{m}^2 F_m + \sigma_{v0}^2 F_v + \sigma_{u0}^2 F_u,$$

(5.7)

where $\tilde{y}_{ij}$ is defined in (5.6), $F_v = U_v U_v'$, $F_u = U_u U'_u$, and $F_i = U_i U'_i$.

(1) MINQU estimators

The MINQU estimation procedure with prior values (C. R. Rao, 1971 and 1972) was elaborated in Section 2.2. In our case here, the MINQU estimators of $\sigma_v^2$, $\sigma_u^2$ and $\sigma_i^2$ are the solutions to the following consistent linear system (referred to as MINQUE equations, C. R. Rao and Kleffe, 1988):

$$H_{UL}\hat{\theta} = q_{UL}$$

(5.9)

where $H_{UL} = (h_{ij})_{(m+2)\times(m+2)}$, $\hat{\theta} = (\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2, \hat{\sigma}_v^2, \hat{\sigma}_u^2)'$, $q_{UL} = (Q_1, \ldots, Q_m, Q_u, Q_v)'$.
CHAPTER 5. TWO-FOLD NESTED ERROR REGRESSION MODELS

with

\[
    h_{ij} = \begin{cases} 
        \text{tr}(MV_0M)^+F_i(MV_0M)^+F_j & \text{for } 1 \leq i, j \leq m \\
        \text{tr}(MV_0M)^+F_i(MV_0M)^+F_u & \text{for } 1 \leq i \leq m, j = m + 1 \\
        \text{tr}(MV_0M)^+F_i(MV_0M)^+F_v & \text{for } 1 \leq i \leq m, j = m + 2 \\
        \text{tr}(MV_0M)^+F_u(MV_0M)^+F_u & \text{for } i = j = m + 1 \\
        \text{tr}(MV_0M)^+F_u(MV_0M)^+F_v & \text{for } i = m + 1, j = m + 2 \\
        \text{tr}(MV_0M)^+F_v(MV_0M)^+F_v & \text{for } i = j = m + 2 
    \end{cases}
\]

and

\[
    Q_v = \mathcal{Y}(MV_0M)^+F_v(MV_0M)^+\mathcal{Y}, \\
    Q_u = \mathcal{Y}(MV_0M)^+F_u(MV_0M)^+\mathcal{Y}, \\
    Q_i = \mathcal{Y}(MV_0M)^+F_i(MV_0M)^+\mathcal{Y}, \\
    M = I_n - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}',
\]

where \(A^+\) denotes the Moore–Penrose inverse of \(A\). Using the following identity (same as (2.5))

\[
    (MV_0M)^+ = V_0^{-1} - V_0^{-1}X_1(X_1'V_0^{-1}X_1)^{-1}X_1'V_0^{-1},
\]

we obtain that

\[
    Q_i = \sigma_i^2 \sum_j \sum_k (y_{ijk} - \bar{y}_{ij})^2 \tag{5.11} \\
        + \sum_j n_{ij}^{-1}w_{i0}^2[\bar{y}_{ij} - x_i'\hat{\beta}_0 - \sigma_{i0}^2w_{i0}(\bar{y}_{i\cdot} - x_i'\hat{\beta}_0)]^2,
\]

\[
    Q_u = \sum_i \sum_j w_{i0}^2[\bar{y}_{ij} - x_i'\hat{\beta}_0 - \sigma_{i0}^2w_{i0}(\bar{y}_{i\cdot} - x_i'\hat{\beta}_0)]^2 \tag{5.12},
\]

\[
    Q_v = \sum_i [w_{i0}(\bar{y}_{i\cdot} - x_i'\hat{\beta}_0)]^2 \tag{5.13}.
\]

For general design points, \(x_i\)'s, explicit expressions for the elements of the matrix \(H_{ii}^\prime\) are not available. The estimators of these error variances are unbiased, but they may take negative values. Further asymptotic results are not available.
(2) **ASR estimators**

From the general result in P. S. R. S. Rao (1977) and P. S. R. S. Rao and Chaubey (1978), the ASR estimators for the variance components are

\[
\hat{\sigma}_v^2 = (\sigma_{v0}^2/m)Q_v,
\]

\[
\hat{\sigma}_u^2 = (\sigma_{u0}^2/\sum_{i=1}^m m_i)Q_u,
\]

\[
\hat{\sigma}_i^2 = (\sigma_{i0}^2/n_i)Q_i.
\]

These estimator are also called MINQUE without unbiasedness or MINQE. They are clearly non-negative, but the bias may be large.

(3) **WLS estimators**

The weighted least squares (WLS) procedure, is a generalization of the classical ANOVA procedure. The weighted average of means (WAM) procedure in P. S. R. S. Rao and Heckler (1992) is also a special case of our WLS procedure. Consider the following weighted sum of squares of residuals from the weighted least square using prior values \(\sigma_{v0}^2\), \(\sigma_{u0}^2\) and \(\sigma_{i0}^2\):

\[
R_0^2 = (Y - X\hat{\beta}_0)'V_0^{-1}(Y - X\hat{\beta}_0).
\]

Denote

\[
S_i = (n_i - m_i)s_i^2 = \sum_{j=1}^{m_i} \sum_{k=1}^{n_{i,j}} (y_{ij,k} - \bar{y}_{i,j})^2,
\]

\[
S_u = \sum_{i=1}^m \sum_{j=1}^{m_i} w_{i,j}(\bar{y}_{ij} - \bar{y}_{i-o})^2,
\]

\[
S_v = \sum_{i=1}^m w_{i0}(\bar{y}_{i-o} - x_i'\hat{\beta}_0)^2.
\]

Then \(R_0^2\) can be rewritten as

\[
R_0^2 = \sum_{i=1}^m \sigma_{i0}^{-2}S_i + S_u + S_v.
\]

By equating \(S_i\), \(S_u\) and \(S_v\) to their expectations, unbiased estimators of \(\sigma_i^2\), \(\sigma_u^2\) and \(\sigma_v^2\) can be obtained from the following expressions:

\[
\hat{\sigma}_i^2 = \hat{s}_i^2 = S_i/(n_i - m_i),
\]

\[
\hat{\sigma}_u^2 = \hat{s}_u^2 = S_u/n,
\]

\[
\hat{\sigma}_v^2 = \hat{s}_v^2 = S_v/m_0.
\]
\[ \sum_{i=1}^{m} \sum_{j=1}^{m_i} B_{ij} \hat{\sigma}_u^2 = Q_i - \sum_{i=1}^{m} \hat{\sigma}_i^2 \sum_{j=1}^{m_i} B_{ij} / n_{ij}, \]  
\[ \sum_{i=1}^{m} A_i \hat{\sigma}_v^2 = Q_v - \hat{\sigma}_v^2 \sum_{i=1}^{m} A_i w_{i0}^{-1} \sum_{j=1}^{m_i} w_{ij0}^2 - \sum_{i=1}^{m} A_i \hat{\sigma}_v^2 w_{i0}^{-1} \sum_{j=1}^{m_i} w_{ij0}^2 / n_{ij}, \]

where

\[ B_{ij} = w_{ij0} - w_{i0}^{-1} w_{ij0}^2, \]
\[ A_i = w_{i0} \left( 1 - w_{i0} x_i \left( \sum_{i=1}^{m_i} w_{i0} x_i x_i^{-1} \right) \right). \]

P. S. R. S. Rao and Heckler (1992) discussed the advantages and disadvantages of these estimators.

ANOVA Estimators The classical ANOVA estimators are a special case of WLS estimators with \( \sigma^2_{\hat{\theta}_0} = \sigma^2_{\hat{\theta}_0} = 0 \) and \( \sigma^2_{\hat{\theta}} = 1 \). With these prior values, \( \hat{y}_{ij}, \hat{y}_{i0} \) and \( \hat{\theta}_0 \) are obtained, and then \( S_i, S_u \) and \( S_v \) are computed from (5.14)–(5.16). Then the ANOVA estimators of \( \sigma^2_{\hat{\theta}}, \sigma^2_{\hat{\theta}} \) and \( \sigma^2_{\hat{\theta}} \) can be obtained from (5.17)–(5.19).

5.2.2 Approximate MINQUE approach

Note that in the MINQUE equation (5.9), the matrix \( H_{UI} \) is quite complicated. There is generally no explicit expression for the solution to this equation. The idea of AMINQUE in Chapter 2 is to simplify this equation so that we can obtain explicit estimators of variance components.

Under certain conditions on the model (which will be spelled out in subsection 5.2.3) and for suitably chosen \( \sigma^2_{\hat{\theta}_0}, \sigma^2_{\hat{\theta}_0} \) and \( \sigma^2_{\hat{\theta}}, \) we have

\[ (MV_0 M)^+ = V_0^{-1} + O(m^{-1}) \]

from (5.10). Hence \( H_{UI} \) in (5.9) is approximately equal to \( H^*_{UI} \) in the sense

\[ H_{UI} = H^*_{UI} + O(m^{-1}), \]

with

\[ H^*_{UI} = (h_{ij})_{(m+2) \times (m+2)} \]
and

\[
\begin{align*}
    h_{ji}^* &= h_{ij}^* = \\
    &\begin{cases}
        \text{tr}(V_0^{-1} F_i V_0^{-1} F_j) & \text{for } 1 \leq i, j \leq m, \\
        \text{tr}(V_0^{-1} F_i V_0^{-1} F_u) & \text{for } 1 \leq i \leq m, j = m + 1, \\
        \text{tr}(V_0^{-1} F_i V_0^{-1} F_v) & \text{for } 1 \leq i \leq m, j = m + 2, \\
        \text{tr}(V_0^{-1} F_u V_0^{-1} F_u) & \text{for } i = j = m + 1, \\
        \text{tr}(V_0^{-1} F_u V_0^{-1} F_u) & \text{for } i = m + 1, j = m + 2, \\
        \text{tr}(V_0^{-1} F_v V_0^{-1} F_v) & \text{for } i = j = m + 2.
    \end{cases}
\end{align*}
\]

It can be shown that

\[
\begin{align*}
    \text{tr}(V_0^{-1} F_i V_0^{-1} F_j) &= 0 & \text{for } i \neq j, \\
    \text{tr}(V_0^{-1} F_i V_0^{-1} F_i) &= \sum_{j=1}^{m} n_{ij}^{-1} w_{i0}^2 \left[ 1 + 2(n_{ij} - 1)\sigma_{i0}^2/\sigma_{i0}^2 + n_{ij}(n_{ij} - 1)(\sigma_{i0}^2/\sigma_{i0}^2)^2 \right] \\
    &\quad - 2\sigma_{i0}^2 \sum_{j=1}^{m} n_{ij}^{-2} w_{i0}^2/(1 + \sigma_{i0}^2 w_{i0}) + \sigma_{i0}^4 \left( \sum_{j=1}^{m} n_{ij}^{-1} w_{i0}^2/(1 + \sigma_{i0}^2 w_{i0}) \right)^2, \\
    \text{tr}(V_0^{-1} F_i V_0^{-1} F_u) &= \sum_{j=1}^{m} n_{ij}^{-1} w_{ij0}^2 - 2\sigma_{i0}^2 \sum_{j=1}^{m} n_{ij}^{-1} w_{ij0}^2/(1 + \sigma_{i0}^2 w_{i0}) \\
    &\quad + \sigma_{i0}^4 \sum_{j=1}^{m} w_{ij0}^2 \sum_{j=1}^{m} n_{ij}^{-1} w_{ij0}^2/(1 + \sigma_{i0}^2 w_{i0})^2, \\
    \text{tr}(V_0^{-1} F_i V_0^{-1} F_v) &= \sum_{j=1}^{m} n_{ij}^{-1} w_{ij0}^2 - 2\sigma_{i0}^2 \sum_{j=1}^{m} n_{ij}^{-1} w_{ij0}^2 + \sigma_{i0}^4 w_{i0}^2 \sum_{j=1}^{m} n_{ij}^{-1} w_{ij0}^2, \\
    \text{tr}(V_0^{-1} F_u V_0^{-1} F_u) &= \sum_{i,j} \left[ \sum_{j=1}^{m} w_{ij0}^2 - \sigma_{i0}^2 w_{i0}^2 \sum_{j=1}^{m} w_{ij0}^2 \right]/(1 + \sigma_{i0}^2 w_{i0}), \\
    \text{tr}(V_0^{-1} F_u V_0^{-1} F_v) &= \sum_{i,j} \left[ \sum_{j=1}^{m} w_{ij0}^2 - 2\sigma_{i0}^2 \sum_{j=1}^{m} w_{ij0}^2/(1 + \sigma_{i0}^2 w_{i0}) \right] \\
    &\quad + \sigma_{i0}^4 \left( \sum_{j=1}^{m} w_{ij0}^2/(1 + \sigma_{i0}^2 w_{i0}) \right)^2, \\
    \text{tr}(V_0^{-1} F_v V_0^{-1} F_v) &= \sum_{i,j}^2 w_{i0}^2.
\end{align*}
\]

The above expressions are still quite complicated for general prior values of \(\sigma_{i0}^2\), \(\sigma_{i}^2\) and \(\sigma_{i}^2\), so an explicit expression for \(H_{II}'\) is not available. To avoid this problem, we use \(\sigma_{i0}^2 = \sigma_{i}^2 = 0\) and \(\sigma_{i}^2 = 1\), like the prior values corresponding to the ANOVA procedure. In this case, \(w_{ij0} = n_{ij}\), and \(w_{i0} = n_{i}\), and \(Q_i, Q_u\) and \(Q_v\) in (5.11)-(5.13) can be rewritten as

\[
    Q_i = \sum_{j=1}^{m} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2 + \sum_{j=1}^{m} n_{ij} (\bar{y}_{ij} - \hat{\beta}_0)^2, \quad (5.20)
\]
CHAPTER 5. TWO-FOLD NESTED ERROR REGRESSION MODELS

\[ Q_u = \sum_{i=1}^{m} \sum_{j=1}^{n_i} n_{ij}^2 (\tilde{y}_{ij} - x'_i \hat{\beta}_0)^2, \]

\[ Q_v = \sum_{i=1}^{m} n_i^2 (\tilde{y}_{i.0} - x'_i \hat{\beta}_0)^2, \]

where \( \tilde{y}_{ij} \) is the same as that in (5.6), and \( \hat{\beta}_0 \) is the ordinary least square estimator of \( \beta \) from model (5.1). Further after computing the entries in \( H_{UU} \), we find that

\[
H_{UU} = \begin{pmatrix}
\left( \text{diag}(n_i) \right)_{(m \times m)} & \mu & \mu \\
\mu' & \sum_i \sum_j n_{ij}^2 & \sum_i \sum_j n_{ij}^2 \\
\mu' & \sum_i \sum_j n_{ij}^2 & \sum_i (\sum_j n_{ij})^2 \\
\end{pmatrix}
\]

with

\[
\mu = (n_1, \ldots, n_m)'.
\]

Applying Lemma 2.1 twice, we can find the inverse of \( H_{UU} \) as

\[
(H_{UU})^{-1} = \begin{pmatrix}
\left( \text{diag}(n_i^{-1}) + N_1^{-1}1_m 1_m' \right)_{(m \times m)} & -N_1^{-1}1_m & 0 \\
-N_1^{-1}1_m' & N_1^{-1} + N_2^{-1} & N_2^{-1} \\
0 & N_2^{-1} & N_2^{-1} \\
\end{pmatrix}
\]

with

\[
N_1 = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (n_{ij}^2 - n_{ij}), \quad N_2 = \sum_{i=1}^{m} n_i^2 - \sum_{i=1}^{m} \sum_{j=1}^{n_i} n_{ij}^2.
\]

The AMINQUE estimators of \( \sigma^2_i, \sigma_a^2 \) and \( \sigma_v^2 \) can be obtained from the linear system

\[
H_{UU} \tilde{\theta} = q_{U}, \quad \text{or} \quad \tilde{\theta} = (H_{UU})^{-1} q_{U},
\]

where

\[
\tilde{\theta} = (\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_m^2, \tilde{\sigma}_a^2, \tilde{\sigma}_v^2)',
\]

which gives

\[
\tilde{\sigma}_i^2 = n_i^{-1} Q_i - N_1^{-1} (Q_u - \sum_i Q_i),
\]

\[
\tilde{\sigma}_a^2 = N_1^{-1} (Q_u - \sum_i Q_i) - N_2^{-1} (Q_v - Q_u),
\]

\[
\tilde{\sigma}_v^2 = N_2^{-1} (Q_v - Q_u).
\]

(5.23)

(5.24)
CHAPTER 5. TWO-FOLD NESTED ERROR REGRESSION MODELS

Note that both $\hat{\sigma}^2_v$ and $\hat{\sigma}^2_u$ may assume negative values. Non-negative AMINQUE estimators of $\sigma^2_v$ and $\sigma^2_u$ are obtained by truncating them at zero, viz.

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

5.2.3 Asymptotic properties of AMINQUE estimators

In this section, the asymptotic normality of estimators $\hat{\sigma}^2_v$ and $\hat{\sigma}^2_u$ in (5.23) and (5.24) are established. First some regularity conditions are imposed on model (5.1):

(A1) There are positive constants $\sigma_0^2, \sigma_\infty^2, c_\infty$, and positive integers $n_\infty, m_\infty$ such that $\sigma_0^2 \leq \sigma_i^2 \leq \sigma_\infty^2, 2 \leq n_{ij} \leq n_\infty, 2 \leq m_i \leq m_\infty$ and $\|x_i\| \leq c_\infty$ for all $i$.

(B1) There is a positive constant $c_0$ so that

$$c_0 \leq m^{-1} (\text{the minimum eigenvalue of } X'X).$$

where $X = (x_1 : x_2 : \cdots : x_m)'$.

(C1) All variance components $v_i, u_{ij}$ and $e_{ijk}$ are symmetrically distributed about 0, and there is a constant $b$ such that

$$E|v_i|^{4+2\delta} \leq b, \quad E|u_{ij}|^{4+2\delta} \leq b \quad \text{and} \quad E|e_{ijk}|^{2+\delta} \leq b \quad \text{for all } i, j, k$$

for some positive constant $\delta$.

Assumption (B1) together with the boundedness of $\|x_i\|$ implies that

$$\lambda_0 I_p \leq I^{-1}X'X \leq \lambda_\infty I_p \quad \text{for all } m$$

for some positive constants $\lambda_0, \lambda_\infty$.

The symmetry of the distributions in condition (C1) implies that $E v_i = 0, E v_i^3 = 0, E u_i = 0, E u_i^3 = 0, E e_{ij} = 0$, and

$$E \left[ \frac{\bar{\bar{e}}_{ij} - \bar{\bar{e}}_{ij}}{b_1 + b_2 \sum_j \sum_k (e_{ijk} - \bar{\bar{e}}_{ij})^2 + b_3 \bar{\bar{e}}_{ij}^2} \right]^r = 0$$

with

$$\bar{\bar{e}}_{ij} = v_i + u_{ij} + \bar{\bar{e}}_{ij}, \quad \bar{\bar{e}}_{ij} = \sum_{k=1}^{n_{ij}} e_{ijk}/n_{ij},$$

$$r > 1.$$
for any positive constants $b_1$, $b_2$, $b_3$, and positive integer $r$.

By the independence between $v_i$, $u_{ij}$, and $e_{ijk}$, and from (A1)

$$\text{var}(\varepsilon_{ij}) = \sigma_v^2 + \sigma_u^2 + \sigma_e^2 / n_{ij} \leq C$$

for some positive constant $C$, and from (5.25), for the $\delta$ given in condition (C1)

$$E|\varepsilon_{ij}|^{2 + \delta} \leq C_1$$

(5.28)

for some positive constant $C_1$ by the $C_r$-inequality.

Note that the asymptotic properties in this chapter are discussed as $m \to \infty$. A subscript $m$, which should be attached to many quantities below, is omitted except when confusion can arise.

Now we turn to the estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$. To prove the asymptotic normality of these estimators, we prove the result on the untruncated estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$ first. By substituting (5.20)-(5.22) into (5.23) and (5.24), $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$ can be rewritten as

$$\hat{\sigma}_v^2 = N_2^{-1} \sum_i Z_{2i} - 2R_4 + R_5,$$

(5.29)

$$\hat{\sigma}_u^2 = N_1^{-1} \sum_i Z_{3i} - \hat{\sigma}_v^2 - 2R_6 + R_7,$$

(5.30)

where

$$Z_{2i} = \sum_j \sum_{j \neq j_i} n_{ij} n_{ij} \left[ v_i^2 + 2v_i (u_{ij} + \bar{e}_{ij}) + (u_{ij} + \bar{e}_{ij})(u_{ij'} + \bar{e}_{ij'}) \right],$$

$$Z_{3i} = \sum_j (n_{ij}^2 - n_{ij}) \left[ v_i^2 + u_{ij}^2 + 2v_i u_{ij} + 2(u_i + u_{ij})\bar{e}_{ij} \right. + \sum \sum_{k \neq k_i} e_{ijk} e_{ijk'} / (n_{ij}^2 - n_{ij}) \right],$$

$$R_4 = N_2^{-1} \sum_i \sum_j n_{ij} \left( n_{ij} - n_{ij} \right) \bar{e}_{ij} (\hat{\beta}_0 - \beta)' x_i,$$

$$R_5 = N_2^{-1} \sum_i \left( n_{i}^2 - \sum_j n_{ij}^2 \right) (\hat{\beta}_0 - \beta)' x_i,$$

$$R_6 = N_1^{-1} \sum_i \sum_j (n_{ij}^2 - n_{ij}) \bar{e}_{ij} (\hat{\beta}_0 - \beta)' x_i,$$

$$R_7 = N_1^{-1} \sum_i \sum_j (n_{ij}^2 - n_{ij}) (\hat{\beta}_0 - \beta)' x_i.$$  

Using the independence between $v$, $u_{ij}$ and $e_{ijk}$, we get

$$E(Z_{2i}) = \sigma_v^2 \left( n_{i}^2 - \sum_j n_{ij}^2 \right) = N_2 \sigma_v^2,$$

(5.31)

$$E(Z_{3i}) = (\sigma_v^2 + \sigma_u^2) \left( \sum_j n_{ij}^2 - n_{i} \right) = N_1 (\sigma_v^2 + \sigma_u^2),$$

(5.32)
and

\[
\text{var}(Z_{2i}) = \sum \sum_{j \neq j'} n_{ij}^2 n_{i'j'} \left[ M_{u4} - \sigma_u^4 \right. \\
+ 4 \sigma_u^2 (\sigma_u^2 + n_{ij}^{-1} \sigma_i^2) + (\sigma_u^2 + n_{ij}^{-1} \sigma_i^2)(\sigma_u^2 + n_{i'j'}^{-1} \sigma_{i'}^2) \left. \right], \\
\text{var}(Z_{3i}) = \sum_j (n_{ij}^2 - n_{ij}) \left[ M_{u4} - \sigma_u^4 + M_{u4} - \sigma_u^4 + 4 \sigma_u^2 \sigma_u^2 \right. \\
+ 4 (\sigma_u^2 + \sigma_u^2) \sigma_i^2/n_{ij} + \sigma_i^2/(n_{ij}^2 - n_{ij}) \left. \right], \\
\text{cov}(Z_{2i}, Z_{3i}) = (M_{u4} - \sigma_u^4) \sum_j (n_{ij}^2 - n_{ij}) \left( n_{ij}^2 - \sum_j n_{ij}^2 \right) \\
+ 4 \sigma_u^2 \sigma_u^2 \sum_j n_{ij} (n_{ij}^2 - n_{ij})(n_{ij} - n_{ij}),
\]

with

\[
M_{u4} = Ev_i^4 \quad \text{and} \quad M_{u4} = Ev_{ij}^4.
\]

Denote

\[
V_1 = N_i^{-2} \sum_1 \text{var}(Z_{2i}) \\
V_2 = N_i^{-2} N_2^{-2} \sum_1 \left( N_i^2 \text{var}(Z_{2i}) - 2 N_1 N_2 \text{cov}(Z_{2i}, Z_{3i}) + N_2^2 \text{var}(Z_{3i}) \right)
\]

(5.33) \hspace{1cm} (5.34)

Now we have the following theorem on the asymptotic normality of \( \hat{\sigma}_u^2 \) and \( \hat{\sigma}_i^2 \).

**Theorem 5.1** Suppose that assumptions (A1), (B1) and (5.25) hold. Then as \( m \to +\infty \)

\[
V_1^{-1/2}(\hat{\sigma}_u^2 - \sigma_u^2) \to_d N(0, 1), \\
V_2^{-1/2}(\hat{\sigma}_i^2 - \sigma_i^2) \to_d N(0, 1).
\]

(5.35) \hspace{1cm} (5.36)

The proof of the above theorem is similar to that of Theorem 2.1. So it is omitted for simplicity. From the above theorem, \( V_1 \) and \( V_2 \) are the asymptotic variances of \( \hat{\sigma}_u^2 \) and \( \hat{\sigma}_i^2 \) respectively. Since both of these go to zero as \( m \to \infty \) under conditions in Theorem 5.1, we have the following corollary.

**Corollary 5.1** Suppose the conditions in Theorem 5.1 hold. Then

\[
\hat{\sigma}_u^2 \to_p \sigma_u^2, \quad \hat{\sigma}_i^2 \to_p \sigma_i^2.
\]
Using arguments similar to those for Corollary 2.2, we have the following results for AMINQUE after truncation.

**Corollary 5.2** Suppose \( \sigma_\nu^2 > 0 \) and \( \sigma_u^2 > 0 \). Then under the conditions in Theorem 5.1,

\[
\hat{\sigma}_\nu^2 = \max(0, \hat{\sigma}_\nu^2) \rightarrow_p \sigma_\nu^2,
\]

\[
\hat{\sigma}_u^2 = \max(0, \hat{\sigma}_u^2) \rightarrow_p \sigma_u^2,
\]

\[
V_1^{-1/2}(\hat{\sigma}_\nu^2 - \sigma_\nu^2) \rightarrow_d N(0, 1),
\]

\[
V_2^{-1/2}(\hat{\sigma}_u^2 - \sigma_u^2) \rightarrow_d N(0, 1).
\]

Therefore the asymptotic variances of \( \hat{\sigma}_\nu^2 \) and \( \hat{\sigma}_u^2 \) are \( V_1 \) and \( V_2 \) correspondingly.

Similar to results in section 2.5.3, we propose estimators for \( V_1 \), and \( V_2 \) by the substitution method:

\[
\tilde{V}_1 = N_\nu^{-2} \sum_i \left[ \sum_j n_{ij} \left( ((\bar{y}_{ij} - x_i^t \hat{\beta}_0) - x_i^t \hat{\beta}_0 - \hat{\sigma}_\nu^2)^2 \right) \right],
\]

(5.37)

\[
\tilde{V}_2 = N_1^{-2} N_\nu^{-2} \sum_i \left[ N_1 \sum_j (n_{ij}^2 - n_{ij}) \left( \bar{y}_{ij} - x_i^t \hat{\beta}_0 \right)^2 - N_2 (\hat{\sigma}_\nu^2 + \hat{\sigma}_u^2) \sum_j (n_{ij}^2 - n_{ij}) 
\]

- \( N_2 \sum_j \sum_k (\bar{y}_{ijk} - \bar{y}_{ij})^2 - N_2 (\hat{\sigma}_\nu^2 + \hat{\sigma}_u^2) \sum_j (n_{ij}^2 - n_{ij}) 
\]

- \( N_1 \sum_j \sum_k (\bar{y}_{ijk} - \bar{y}_{ij}) \left( (\bar{y}_{ij} - x_i^t \hat{\beta}_0) - x_i^t \hat{\beta}_0 \right) + N_1 \hat{\sigma}_\nu^2 \left( n_{ij}^2 - \sum_j n_{ij}^2 \right) \right)^2.
\]

These estimators are consistent as shown in the following theorem.

**Theorem 5.2** Suppose that assumptions (A1), (B1), and (5.25) hold. Then

\[
m(\tilde{V}_1 - V_1) \rightarrow_p 0,
\]

\[
m(\tilde{V}_2 - V_2) \rightarrow_p 0.
\]

The proof is similar to that of Theorem 2.2, so it is omitted for simplicity.

Note that we can also get another set of estimators of \( V_1 \) and \( V_2 \) by using \( \hat{\sigma}_\nu^2 \) and \( \hat{\sigma}_u^2 \) instead of using \( \hat{\sigma}_\nu^2 \) and \( \hat{\sigma}_u^2 \) in (5.37) and (5.38), especially when the true value of \( \sigma_\nu^2 \) and \( \sigma_u^2 \) are significantly away from 0 relative to values of \( \sigma_\nu^2 \). The resulting variance estimators are also consistent, but may be inefficient.
5.3 Estimation of Regression Coefficients

Unlike in the case of the one-fold nested-error model, we have difficulty in obtaining a positive estimator of \( \sigma^2_u + n_i^{-1} \sigma^2_i \) using AMINQU estimators of \( \sigma^2_u \) and \( \sigma^2_i \). Therefore we want to estimate the regression vector \( \beta \) using the approach in Section 3.4.

We proved the consistency and asymptotic normality of \( \hat{\delta}_v^2 \) and \( \hat{\delta}_u^2 \) in section 5.2.3. The within group sample variance \( s^2_i = \left( \sum_{j=1}^{m_i} \frac{1}{n_{ij} - 1} \right)^{-1} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2 \) is an unbiased estimator of \( \sigma^2_i \). From (5.5), a weighted least square estimator \( \hat{\beta}_w \) of \( \beta \) can be obtained from

\[
\hat{\beta}_w = (X'WX)^{-1} \sum_{i=1}^{m} \hat{w}_i x_i \hat{y}_i.
\]  

(5.39)

where

\[
W = \text{diag}(\hat{w}_i),
\]

\[
\hat{w}_{ij} = (\hat{\delta}_u^2 + n_i^{-1} s^2_i)^{-1},
\]

\[
\hat{w}_i = \sum_{j=1}^{m_i} \hat{w}_{ij} / \left( 1 + \hat{\delta}_u^2 \sum_{j=1}^{m_i} \hat{w}_{ij} \right),
\]

\[
\hat{y}_{i..} = \sum_{j=1}^{m_i} \hat{w}_{ij} \bar{y}_{ij} / \left( \sum_{j=1}^{m_i} \hat{w}_{ij} \right),
\]

where \( \bar{y}_{ij} \) is the same as that in (5.6).

We now establish the asymptotic normality of \( \hat{\beta}_w \) along the lines of Section 3.1–3.2. Let

\[
\hat{w}_{ij}^{(1)} = (\sigma^2_u + n_i^{-1} s^2_i)^{-1}, \quad \hat{w}_i^{(1)} = \left( 1 + \sigma^2_u \sum_{j=1}^{m_i} \hat{w}_{ij}^{(1)} \right)^{-1} \sum_{j=1}^{m_i} \hat{w}_{ij}^{(1)}
\]

\[
\tau_{\delta_i} = \sum_j E \hat{w}_{ij}^{(1)},
\]

\[
\tau_{\tau_i} = E[\hat{w}_{ij}^{(1)} \hat{y}_{ij}^{(1)}]^2, \quad \text{where} \quad \hat{y}_{i..}^{(1)} = \sum_j \hat{w}_{ij}^{(1)} \bar{y}_{ij} / \sum_j \hat{w}_{ij}^{(1)},
\]

\[
D_\delta = \text{diag}(\tau_{\delta_i}),
\]

\[
D_\tau = \text{diag}(\tau_{\tau_i}).
\]

It is easy to check that when \( \sigma^2_u > 0 \), both \( \tau_{\delta_i} \) and \( \tau_{\tau_i} \) are finite.

We have the following theorem on the asymptotic normality of \( \hat{\beta}_w \).
CHAPTER 5. TWO-FOLD NESTED ERROR REGRESSION MODELS

Theorem 5.3 Suppose that assumptions (A1), (B1) and (C1) hold. Then
\[ V^{-1/2}_\beta (\hat{\beta}_w - \beta) \rightarrow_p N(0, I_p) \quad \text{as} \ m \rightarrow +\infty \]
where \( V^{-1/2}_\beta = (V^{1/2}_\beta)^{-1} \) and \( V^{1/2}_\beta \) is a square root of
\[ V_\beta = (X'D_\theta X)^{-1} X'D_\gamma X(X'D_\theta X)^{-1}. \]
(note that \( X = (x_1 : x_2 : \cdots : x_m)' \)).

Proof. Similar to the arguments in Sections 3.1 and 3.2, we can show that
\[
\sum_i \hat{w}_i x_i x'_i = \sum_i \tau_{\theta i} x_i x'_i + o_p(m) = (X' D_\theta X) + o_p(m), \tag{5.40}
\]
\[
\sum_i x_i \hat{w}_i \sum_j \hat{w}_{ij} \epsilon_{ij} / \sum_i \hat{w}_{ij} = \sum_i \hat{w}_i^{(1)} \epsilon_i^{(1)} + o_p(m^{1/2})
= X' col_i(\hat{w}_i^{(1)} \epsilon_i^{(1)}) + o_p(m^{1/2}). \tag{5.41}
\]

From condition (C1),
\[ Ecol_i(\hat{w}_i^{(1)} \epsilon_i^{(1)}) = 0, \]
and
\[ \text{var}[X' col_i(\hat{w}_i^{(1)} \epsilon_i^{(1)})] = X' D_\gamma X. \]

Along the lines of the proof of (3.7), we can show that as \( m \rightarrow +\infty \)
\[(X' D_\gamma X)^{-1/2} X' col_i(\hat{w}_i^{(1)} \epsilon_i^{(1)}) \rightarrow_d N(0, I_p). \tag{5.42}\]

Hence,
\[(X' D_\gamma X)^{-1/2} \sum_i x_i \hat{w}_i \sum_j \hat{w}_{ij} \epsilon_{ij} / \sum_j \hat{w}_{ij} \rightarrow_d N(0, I_p) \quad \text{as} \ m \rightarrow +\infty. \]

Then from (5.40) we can prove that
\[(X' D_\gamma X)^{-1/2}(X' D_\theta X)(\hat{\beta}_w - \beta) \rightarrow_p N(0, I_p) \quad \text{as} \ m \rightarrow +\infty, \]

analogous to the proof of (3.12). The theorem is proved by an argument similar to those used in the last paragraph of Theorem 3.1. The details are omitted for simplicity. \( \square. \)
An estimator of $V_{\beta}$ can be obtained by the substitution method as follows:

$$\hat{V}_{\beta} = (X'WXX)^{-1}X'\hat{D}_{\tau}X(X'WXX)^{-1}.$$ 

where

$$D_{\tau} = \text{diag}(\hat{\psi}_i^2(\hat{y}_{i.} - x_i'\hat{\beta}_0)^2).$$

This variance estimator is consistent as shown in the next theorem.

**Theorem 5.4** Suppose that assumptions (A1), (B1) and (C1) hold. Then as $m \to +\infty$,

$$m(\hat{V}_{\beta} - V_{\beta}) \to_p 0.$$ 

The proof is similar to that of Theorem 2.2 and hence is omitted for simplicity.
Part II

Small Area Estimation by Combining Time Series and Cross-Sectional Data
Chapter 6

EBLUP Estimators

In Part II, which consists of this chapter and the following Chapter 7, we study the small area model (1.27) by combining time series and cross-sectional data. We use the empirical best linear unbiased prediction (EBLUP) approach in this chapter, and use the hierarchical Bayes approach in Chapter 7.

Let \( y_{it} \) be the direct survey estimator of \( \theta_{it} \), the \( i \)--th small area mean for the period \( t \) \((i = 1, \ldots, k; t = 1, \ldots, T) \). We assume that

\[
\begin{align*}
y_{it} &= \theta_{it} + e_{it} = x_{it}' \beta + u_{it} + e_{it}, \\
u_{it} &= \rho u_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1, \quad \text{for } t > 1, \quad \text{and} \quad u_{ii} = \epsilon_{ii}/\sqrt{1-\rho^2},
\end{align*}
\]

(6.1)

where \( x_{it} = (x_{i1t}, x_{i2t}, \ldots, x_{itp})' \), are vectors of fixed concomitant variables related to \( \theta_{it} \), \( i = 1, \ldots, k, \ t = 1, \ldots, T \); \( \beta = (\beta_1, \beta_2, \ldots, \beta_p)' \) is a vector of regression parameters; \( v_i \) is a random effect corresponding to area \( i \), following normal distribution with mean 0 and variance \( \sigma_v^2 \) (i.e., \( v_i \overset{iid}{\sim} N(0, \sigma_v^2) \)); for given \( i \), the \( u_{it}'s \) follow an AR(1) process with common autocorrelation \( \rho \) over time period \( t = 1, \ldots, T \); \( \epsilon_{it} \) are random errors associated with \( u_{it} \) and \( \epsilon_{it} \overset{iid}{\sim} N(0, \sigma^2) \); \( e_i = (e_{i1}, e_{i2}, \ldots, e_{iT})' \) is the vector of sampling errors for area \( i \), following a multivariate normal distribution with mean 0, and known covariance matrix \( \Sigma_i \) \( (e_i \overset{iid}{\sim} N(0, \Sigma_i)) \); \( \{v_i\}, \{\epsilon_{it}\} \) and \{\( e_i \)\} are mutually independent. We are interested in estimating the small area means \( \theta_{iT} \) for the current period.

Our model (6.1) can be expressed as a special case of a general mixed linear model as follows. Arranging the data \{\( y_{it} \)\} as \( y = (y_{i1}, \ldots, y_{iT}; \ldots; y_{k1}, \ldots, y_{kT})' = \)
(y', \ldots, y')', the proposed model (6.1) may be written in matrix form as

\[ y = X\beta + Zv + u + e \]  

(6.2)

where

\[ X = (X_1', \ldots, X_k')', \quad X_i' = (x_{i1}, \ldots, x_{iT}), \]
\[ Z = I_k \otimes 1_T, \]
\[ v = (v_1, \ldots, v_k)', \]
\[ u = (u_1', \ldots, u_k')', \quad u_i' = (u_{i1}, \ldots, u_{iT}) \]
\[ e = (e_1', \ldots, e_k')', \quad e_i' = (e_{i1}, \ldots, e_{iT}) \]

with \( I_k \) denoting the identity matrix of order \( k \), \( 1_T \) denoting a \( T \)-vector of 1's, and \( \otimes \) denoting the direct Kronecker product. Further,

\[ E(v) = 0, \quad Cov(v) = \sigma_v^2 I_k, \]
\[ E(u) = 0, \quad Cov(u) = \sigma^2 I_k \otimes \Gamma = \sigma^2 R, \]
\[ E(e) = 0, \quad Cov(e) = \Sigma = \text{block diag}(\Sigma_1, \ldots, \Sigma_k), \]  

(6.3)

and \( v, u \) and \( e \) are mutually independent, where \( \Gamma \) is a \( T \times T \) matrix with elements \( \rho^{i-i}/(1 - \rho^2) \). It follows from (6.2) and (6.3) that

\[ Cov(y) = V = \Sigma + \sigma^2 R + \sigma^2_0 ZZ' \]
\[ = \text{block diag}(\Sigma_i + \sigma^2 \Gamma + \sigma^2_0 J_T) = \text{block diag}(V_i) \quad \text{say}, \]

with \( J_T = 1_T 1'_T \).

In this chapter, we follow the standard EBLUP approach by first deriving the BLUP estimator of \( \theta_{1T} \) in Section 6.1. In Section 6.2, assuming known autocorrelation \( \rho \), we give "method of moments" type estimators of variance components \( \sigma_v^2 \) and \( \sigma^2 \), and two-stage estimators of small area means are then obtained. Extending the approach of Prasad and Rao (1986, 1990) for the Fay–Herriot model, an estimator of MSE of the two-stage estimator is obtained in Section 6.3, correct to terms of order \( O(k^{-1}) \) for small or moderate number of time points, \( T \), and a
relatively large number of small areas, $k$. The case of unknown autocorrelation $\rho$ is considered in Section 6.4. We provide an estimator of $\rho$. But the estimator turned out to be unsatisfactory unless the number of small areas is very large. Hence two other alternative methods are suggested. The two-step estimators associated with these three methods of dealing with unknown $\rho$ are also obtained. Methods for estimating the MSE of the corresponding two-step estimators are considered. The results of a simulation study on the efficiency of two-stage estimators and relative bias of estimators of MSE are reported in Section 6.5.

6.1 BLUP Estimator

The current mean of the $t$-th small area $\theta_{it} = x_{it}'\beta + \upsilon_i + u_{it}$ is a special case of the linear combination $\tau = l'\beta + l_1'\upsilon + l_2'u$ with $l = x_{iT}$, $l_1$ is the $k$-vector with 1 in the $i$-th position and 0 elsewhere, and $l_2$ is the $(kT)$-vector with 1 in the $(iT)$-th position and 0 elsewhere. Noting that model (6.2) is a special case of the general mixed linear model, the BLUP estimator of $\theta_{it}$ can be obtained from Henderson's (1950) general results for arbitrary linear combinations $\tau$.

Assuming first that $\sigma^2_v$, $\sigma^2_\upsilon$ and $\rho$ are known, the BLUP estimator of $\tau$ is given by

$$\tilde{\tau} = l'\tilde{\beta} + (\sigma^2_v l_1'Z' + \sigma^2_\upsilon l_2'R)V^{-1}(y - X\hat{\beta}),$$

where $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ is the weighted least squares estimator of $\beta$. Using the special structures of $l_1$, $l_2$, $Z$, $R$ and $V$, it is easily seen that (6.4) reduces to

$$\tilde{\theta}_{iT} = \varphi(\sigma^2, \sigma^2_v, \rho, y)$$

$$= x_{iT}'\tilde{\beta} + (\sigma^2_v l_1'T + \sigma^2_\upsilon T)'(\Sigma_i + \sigma^2\Gamma + \sigma^2_\upsilon J_T)^{-1}(y_i - X_i\tilde{\beta}),$$

where $\gamma_T$ is the $T$-th row of $\Gamma$. The BLUP estimator may also be written as a weighted sum of the direct estimator $y_{iT}$, the synthetic estimator $x_{iT}'\tilde{\beta}$ and the prediction errors at other time points in the same area, $y_{it} - x_{it}'\tilde{\beta}$, $t = 1, \ldots, T - 1$:

$$\tilde{\theta}_{iT} = w_{iT}y_{iT} + (1 - w_{iT})x_{iT}'\tilde{\beta} + \sum_{i=1}^{T-1} w_{it}(y_{it} - x_{it}'\tilde{\beta}),$$
where

\[(w_{i1}^*, \ldots, w_{iT}^*) = (\sigma_0^2 I_T + \sigma_\rho^2 \gamma_T)^1 V_i^{-1}.\]

### 6.2 Two-Stage Estimator: \(\rho\) Known

In practice, the parameters, \(\sigma^2\), \(\sigma_\rho^2\) and \(\rho\) are usually unknown. We first assume that \(\rho\) is known and replace \(\sigma^2\) and \(\sigma_\rho^2\) in (6.5) by their consistent estimators \(\hat{\sigma}^2(\rho)\), \(\hat{\sigma}_\rho^2(\rho)\) to obtain a two-stage estimator \(\hat{\theta}_{1T}(\rho)\). The case of unknown \(\rho\) is studied in Section 6.4.

#### 6.2.1 Estimation of \(\sigma^2\)

Pantula and Pollock (1985) estimated \(\sigma^2\), \(\sigma_\rho^2\) in the nested error regression model (1.26) with autocorrelated errors \(w_{it}\) by extending the method of fitting constants for the special case of independent errors (Fuller and Battese, 1973). We now extend their method to the more general model (6.1) with both autocorrelated errors \(u_{it}\) and sampling errors \(e_{it}\), assuming \(\rho\) is known.

We first obtain an unbiased estimator of \(\sigma^2\). For this purpose, we transform model (6.1) to eliminate the random effect \(v_i\). First transform \(y_i\) to \(z_i = Py_i\) such that the covariance matrix of \(Pu_i\) is \(\sigma^2 I_T\), i.e., \(\Gamma = P^{-1}(P^{-1})^{-1}\). The \(T \times T\) matrix \(P\) has the following form: first diagonal element \((1 - \rho^2)^{1/2}\), remaining diagonal elements 1, \((t+1, t)\)-th element \(-\rho\) for \(t = 1, \ldots, T - 1\), and remaining elements 0 (see Judge et al., 1985, p.285). The transformed model is given by

\[z_i = PX_i \beta + f v_i + P(u_i + e_i), \quad i = 1, \ldots, k.\]  

(6.6)

where \(f = (f_1, \ldots, f_T)'\) with \(f_t = (1 - \rho^2)^{1/2}\) and \(f_t = 1 - \rho\) for \(2 \leq t \leq T\). Next we transform \(z_i\) to \(z_i^{(1)} = (I_T - D)z_i\), where \(D = (ff^{'}) / c\) with \(c = f^{'f} = (1 - \rho)[T - (T - 2)\rho]\). This leads to the following reduced model:

\[z_i^{(1)} = H_i^{(1)} \beta + e_i^*, \quad i = 1, \ldots, k,\]  

(6.7)

where \(H_i^{(1)} = (I_T - D)PX_i\) and \(e_i^* = (I_T - D)P(u_i + e_i)\).
CHAPTER 6. EBLUP ESTIMATORS

Since
\[ \text{Cov}(e_i^*) = (I_T - D)(\sigma^2 I_T + P\Sigma_i P')(I_T - D)' \] (6.8)
does not involve \( \sigma_i^2 \), we can estimate \( \sigma^2 \) through reduced model (6.7) using the residual sum of squares. Let \( z^{(1)} = [(z_1^{(1)})', \ldots, (z_k^{(1)})']', H^{(1)} = [(H_1^{(1)})', \ldots, (H_k^{(1)})']', \) and \( \hat{e}' \hat{e} \) be the residual sum of squares obtained by regressing \( z^{(1)} \) on \( H^{(1)} \) using ordinary least squares. An unbiased estimator \( \hat{\sigma}^2 \) is then given by
\[
\hat{\sigma}^2(\rho) = \hat{e}' \hat{e} - tr\{(\text{block diag}(I_T - D) - H^{(1)}(H^{(1)'H^{(1)}})^{-1}H^{(1)'}) \cdot \\
\text{block diag}(P\Sigma_i P')\} [k(T - 1) - \text{rank}(H^{(1)})]^{-1},
\] (6.9)
where \( A^{-} \) is a generalized inverse of \( A \). The unbiasedness of \( \hat{\sigma}^2(\rho) \) follows by noting that \( \hat{e}' \hat{e} = e'(I_T - H^{(1)}(H^{(1)'H^{(1)}})^{-1}H^{(1)'})e \), then using (6.8) and the following lemma:

**Lemma 6.1** If \( y \) is a random \( p \)-dimension multivariate with mean 0 and covariance matrix \( \Omega \), then \( E(y'Gy) = tr(G\Omega) \) for any symmetric matrix \( G \).

Since \( \hat{\sigma}^2(\rho) \) can take negative value, we truncate it at zero and use
\[
\hat{\sigma}^2 = \max\{0, \hat{\sigma}^2(\rho)\}. \] (6.10)

The truncated estimator \( \hat{\sigma}^2(\rho) \) is no longer unbiased, but it is asymptotically consistent with bounded (small or moderate) number of time points, \( T \), as \( k \to \infty \).

6.2.2 Estimation of \( \sigma_i^2 \)

Turning to the estimation of \( \sigma_i^2 \), we transform (6.6) by changing \( z_i \) to \( c^{-1/2}f'z_i \) such that \( u_i^* = c^{-1/2}f'P \bar{u}_i \) has mean 0 and variance \( \sigma^2 \). The transformed model is given by
\[
c^{-1/2}f'z_i = c^{-1/2}f'PX_i\delta + c^{1/2}u_i + u_i^* + c^{-1/2}f'P\bar{u}_i
\]
with error variance \( \sigma_i^2 + \sigma^2 + c^{-1}f'P\Sigma_i P'f \). Let \( \hat{u}' \hat{u} \) be the residual sum of squares obtained by regressing \( c^{-1/2}f'z_i \) on \( c^{-1/2}f'PX_i \) using ordinary least squares. An
unbiased estimator $\hat{\sigma}_0^2$ is then given by

$$\hat{\sigma}_0^2(\rho) = c^{-1}[k - \text{rank}(F)]^{-1}[\hat{u}'\hat{u} - \text{tr}\{(I_k - F(F'F)^{-1}F')\text{diag}(c^{-1}f'P\Sigma_fP'f)\}] - c^{-1}\hat{\sigma}^2(\rho), \quad (6.11)$$

where $F = (X_1'P'f, \ldots, X_k'P'f)'$. The unbiasedness of $\hat{\sigma}_0^2(\rho)$ follows by noting that

$$E(\hat{u}'\hat{u}) = (c\sigma_0^2 + \sigma^2)(k - \text{rank}(F)) + \text{tr}[(I_k - F(F'F)^{-1}F')\text{diag}(c^{-1}f'P\Sigma_fP'f)].$$

Also, noting that $\hat{\sigma}_0^2(\rho)$ can take negative values, we truncate it at zero and use

$$\hat{\sigma}_0^2 = \max\{0, \hat{\sigma}_0^2(\rho)\}. \quad (6.12)$$

The truncated estimator $\hat{\sigma}_0^2(\rho)$ is not unbiased, but it is also asymptotically consistent as $k \to \infty$.

**Remark.** The consistency of $\hat{\sigma}^2$ and $\hat{\sigma}_0^2$, for small or moderate number of time periods, can be established using techniques similar to those in Chapter 2 under suitable regularity conditions.

### 6.2.3 A two-stage estimator

A two-stage estimator of $\theta_{IT}$ is now obtained from (6.5) by substituting $\hat{\sigma}^2(\rho)$ and $\hat{\sigma}_0^2(\rho)$ for $\sigma^2$ and $\sigma_0^2$ respectively:

$$\hat{\theta}_{IT}(\rho) = \varphi[\hat{\sigma}^2(\rho), \hat{\sigma}_0^2(\rho), \rho, y]. \quad (6.13)$$

The two-stage estimator, $\hat{\theta}_{IT}(\rho)$, remains unbiased, noting that $\hat{\sigma}^2(\rho)$ and $\hat{\sigma}_0^2(\rho)$ are even functions of $y$ and translation invariant, i.e., they remain unchanged when $y$ is changed to $-y$ or to $y - Xa$ for all $y$ and $a$; see Kackar and Harville (1984).

It is not necessary to assume normality of the errors in the model (6.1); only symmetric distributions are needed. However, we need normality to derive an estimator of MSE of $\hat{\theta}_{IT}(\rho)$, correct to a second-order approximation as $k \to \infty$. 
6.3 Estimator of MSE

We first obtain a second order approximation to the MSE of the two-stage estimator \( \hat{\theta}_{IT}(\rho) \), in the sense that the neglected terms are of order \( o(k^{-1}) \), for large \( k \). Using the approximation, an estimator of \( MSE[\hat{\theta}_{IT}(\rho)] \), correct to the same order of approximation, is then obtained.

6.3.1 Second order approximation to MSE

Following Kackar and Harville (1984), we have

\[
MSE[\hat{\theta}_{IT}(\rho)] = MSE(\hat{\theta}_{IT}) + E[\hat{\theta}_{IT} - \hat{\theta}_{IT}(\rho)]^2, \tag{6.14}
\]

where \( MSE(\hat{\theta}_{IT}) = E(\hat{\theta}_{IT} - \theta_{IT})^2 \). Further, using Henderson’s (1975) general result, an exact expression for \( MSE(\hat{\theta}_{IT}) \) is given by

\[
MSE(\hat{\theta}_{IT}) = g_{1IT}(\sigma^2, \sigma_v^2, \rho) + g_{2IT}(\sigma^2, \sigma_v^2, \rho), \tag{6.15}
\]

where

\[
g_{1IT}(\sigma^2, \sigma_v^2, \rho) = \sigma_v^2 + \frac{\sigma^2}{1 - \rho^2} - (\sigma_v^2)_{IT}^{-1} - \sigma_v^2(\sigma_v^2)_{IT}^{-1} + \sigma^2(\sigma_v^2)_{IT}^{-1}, \tag{6.16}
\]

and

\[
g_{2IT}(\sigma^2, \sigma_v^2, \rho) = [x_{IT} - X'_i V_i^{-1}(\sigma_v^2)_{IT}^{-1} + \sigma^2(\sigma_v^2)_{IT}^{-1}] (X'V^{-1}X)^{-1}[x_{IT} - X'_i V_i^{-1}(\sigma_v^2)_{IT}^{-1} + \sigma^2(\sigma_v^2)_{IT}^{-1}]. \tag{6.17}
\]

The second term \( g_{2IT}(\sigma^2, \sigma_v^2, \rho) \) in (6.15), due to estimating \( \beta \), is of order \( O(k^{-1}) \), while the first term \( g_{1IT}(\sigma^2, \sigma_v^2, \rho) \) is \( O(1) \).

It remains to evaluate the term \( E[\hat{\theta}_{IT} - \hat{\theta}_{IT}(\rho)]^2 \) in (6.14). Following Kackar and Harville (1984), we propose a Taylor approximation to this term:

\[
E[\hat{\theta}_{IT} - \hat{\theta}_{IT}(\rho)]^2 \approx E[(\partial \hat{\theta}_{IT}/\partial \sigma^2)(\hat{\sigma}^2(\rho) - \sigma^2) + (\partial \hat{\theta}_{IT}/\partial \sigma_v^2)(\hat{\sigma}_v^2(\rho) - \sigma_v^2)]^2.
\]

Following Prasad and Rao (1990), a further approximation is obtained as

\[
E[\hat{\theta}_{IT} - \hat{\theta}_{IT}(\rho)]^2 \approx tr[\Delta'V\Delta\Sigma^2] = g_{3IT}(\sigma^2, \sigma_v^2, \rho), \text{ say} \tag{6.18}
\]
where $\Sigma^*$ is the $2 \times 2$ covariance matrix of the unbiased estimators $\hat{\sigma}^2(\rho)$ and $\hat{\sigma}_v^2(\rho)$. Further $\Delta = (\partial b/\partial \sigma^2, \partial b/\partial \sigma_v^2)$ with $b' = (\sigma^2_{E1T} + \sigma^2_{E})V_i^{-1}$. Calculating the derivatives $\partial b/\partial \sigma^2$ and $\partial b/\partial \sigma_v^2$ we obtain, after simplification, $\Delta'V\Delta = A$, where $A$ is a $2 \times 2$ symmetric matrix with diagonal elements

\[
a_{11} = [\gamma_T - \Gamma V_i^{-1}(\sigma^2_{E1T} + \sigma^2_{E})]'V_i^{-1}[\gamma_T - \Gamma V_i^{-1}(\sigma^2_{E1T} + \sigma^2_{E})],
\]
\[
a_{22} = [1 - J TV_i^{-1}(\sigma^2_{E1T} + \sigma^2_{E})]'V_i^{-1}[1 - J TV_i^{-1}(\sigma^2_{E1T} + \sigma^2_{E})],
\]
and off-diagonal elements

\[
a_{12} = a_{21} = [\gamma_T - \Gamma V_i^{-1}(\sigma^2_{E1T} + \sigma^2_{E})]'V_i^{-1}[1 - J TV_i^{-1}(\sigma^2_{E1T} + \sigma^2_{E})].
\]

The term $g_{3T}(\sigma^2, \sigma_v^2, \rho)$ is of the same order as $g_{3T}(\sigma^2, \sigma_v^2, \rho)$. Combining (6.15) and (6.18), we get a second order approximation to $MSE[\hat{\theta}_{iT}(\rho)]$ as

\[
MSE[\hat{\theta}_{iT}(\rho)] \approx g_{1T}(\sigma^2, \sigma_v^2, \rho) + g_{3T}(\sigma^2, \sigma_v^2, \rho) + g_{3T}(\sigma^2, \sigma_v^2, \rho).
\]  
(6.19)

The neglected terms in the approximation (6.19) are of lower order, $o(k^{-1})$, for large $k$. A rigorous proof of this assertion is quite involved, but follows along the lines of Prasad and Rao (1990).

It remains to obtain the covariance matrix $\Sigma^*$. For this purpose we rewrite $\hat{\sigma}^2(\rho)$ and $\hat{\sigma}_v^2(\rho)$ as

\[
\hat{\sigma}^2(\rho) = [(k - 1)T - rank(H^{(1)})]^{-1}a'C_1a + \text{const}.
\]
(6.20)
\[
\hat{\sigma}_v^2(\rho) = c^{-1}[k - rank(F)]^{-1}a'C_2a
- c^{-1}[(k - 1)T - rank(H^{(1)})]^{-1}a'C_1a + \text{const}.
\]
(6.21)

where $a = Zu + u + e \sim N(0, V)$,

\[
C_1 = C'[I_n - CX(X'C'C')X^{-1}X'C']C
\]

with $C = \text{block diag}_{i}((I_T - D)P)$ and

\[
C_2 = C'^*[I_n - C'X(X'C'C')X^{-1}X'C']*C
\]

with $C^* = \text{block diag}_{i}(c^{-1/2}f'P)$. We can now evaluate the elements of $\Sigma^*$ using (6.20) and (6.21) in the following well-known lemma on the covariance of two quadratic forms of normally distributed variables.
Lemma 6.2 If $y \sim N(0, \Omega)$, then $\text{Cov}(y'G_1y, y'G_2y) = 2\text{tr}(G_1\Omega G_2\Omega)$, where $G_1$ and $G_2$ are two symmetric matrices.

6.3.2 Second order approximation to estimator of MSE

Following Prasad and Rao (1990), it can be shown that

$$E[g_{1T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho)] \approx g_{1T}(\sigma^2, \sigma_v^2, \rho) - g_{3T}(\sigma^2, \sigma_v^2, \rho). \quad (6.22)$$

$$E[g_{2T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho)] \approx g_{2T}(\sigma^2, \sigma_v^2, \rho), \quad (6.23)$$

and

$$E[g_{3T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho)] \approx g_{3T}(\sigma^2, \sigma_v^2, \rho). \quad (6.24)$$

The neglected terms in the approximations (6.22)–(6.24) are of lower order, $o(k^{-1})$. A rigorous proof of this assertion is quite involved, but it essentially involves showing that $Eo_p(k^{-1}) = o(k^{-1})$.

It now follows from (6.22)–(6.24) and (6.19) that an estimator of $MSE[\hat{\theta}_{1T}(\rho)]$ with expectation correct to $O(k^{-1})$ is given by

$$\text{mse}[\hat{\theta}_{1T}(\rho)] \approx g_{1T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho) + g_{2T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho) + 2g_{3T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho). \quad (6.25)$$

A naive estimator of MSE, which ignores the uncertainty in the estimators $\hat{\sigma}^2(\rho)$ and $\sigma_v^2(\rho)$ is given by

$$\text{mse}_N[\hat{\theta}_{1T}(\rho)] = g_{1T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho) + g_{2T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho). \quad (6.26)$$

The naive MSE estimator (6.26) is computationally simpler than (6.25) since it does not require the calculation of $g_{3T}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho)$ which involves the estimation of the covariance matrix of the variance components $\hat{\sigma}^2(\rho)$ and $\hat{\sigma}_v^2(\rho)$. However, it can lead to severe underestimation of the true MSE (see Section 6.5).

6.4 Two–Stage Estimators: $\rho$ Unknown

In Sections 6.2 and 6.3 we obtained a two–stage estimator of $\theta_{1T}$ and an estimator of its MSE, assuming known $\rho$. However, in practice $\rho$ is seldom known. In this
section, we introduce an estimator \( \hat{\rho} \) of \( \rho \). Using \( \hat{\rho} \), we obtain the estimators \( \hat{\sigma}^2(\hat{\rho}) \) and \( \hat{\sigma}_0^2(\hat{\rho}) \). Hence, an EBLUP estimator of \( \theta_{iT} \) can be obtained by substituting \( \hat{\sigma}^2(\hat{\rho}) \), \( \hat{\sigma}_0^2(\hat{\rho}) \), and \( \hat{\rho} \) for \( \sigma^2 \), \( \sigma_0^2 \) and \( \rho \), respectively, in (6.5). Similar to the procedure in Section 6.3, we derive a second order approximation to the MSE and an estimator of the MSE for \( \hat{\theta}_{iT} \) to the same order of approximation.

The above estimator of \( \rho \) is not very satisfactory for small or moderate number of small areas, as shown in our Monte Carlo simulation study. Therefore, two alternative methods are also suggested in this section.

### 6.4.1 Estimation of \( \rho \)

We now obtain a consistent moment estimator of \( \rho \) by taking into account the sampling errors. Let \( a_{it} = v_i + u_{it} + e_{it} \). Then we have

\[
E\left[k^{-1}(T-2)^{-1} \sum_{i=1}^{k} \sum_{t=1}^{T-2} a_{it}(a_{it} - a_{i,t+1}) \right] = \sigma_a^2/(1+\rho) + k^{-1}(T-2)^{-1} \sum_{i=1}^{k} \sum_{t=1}^{T-2} (\sigma_{it}^{(i)} - \sigma_{t,i+1}^{(i)}) \tag{6.27}
\]

and

\[
E\left[k^{-1}(T-2)^{-1} \sum_{i=1}^{k} \sum_{t=1}^{T-2} a_{it}(a_{i,t+1} - a_{i,t+2}) \right] = \sigma_a^2/(1+\rho) + k^{-1}(T-2)^{-1} \sum_{i=1}^{k} \sum_{t=1}^{T-2} (\sigma_{t,i+1}^{(i)} - \sigma_{t,i+2}^{(i)}), \tag{6.28}
\]

where \( \sigma_{it}^{(i)} = \text{cov}(e_{it}, e_{i,t+s}) \) and \( \sigma_{it}^{(i)} = \text{var}(e_{it}) \). It follows from (6.27) and (6.28) that a moment estimator of \( \rho \), assuming known errors \( a_{it} \), is given by

\[
\rho^* = \frac{\sum_{i=1}^{k} \sum_{t=1}^{T-2} a_{it}(a_{i,t+1} - a_{i,t+2}) - (\sigma_{t,i+1}^{(i)} - \sigma_{t,i+2}^{(i)})}{\sum_{i=1}^{k} \sum_{t=1}^{T-2} a_{it}(a_{i,t} - a_{i,t+1}) - (\sigma_{t,i}^{(i)} - \sigma_{t,i+1}^{(i)})}. \tag{6.29}
\]

Since \( a_{ij} \) are not observable in practice, we replace the \( a_{ij} \)'s in (6.29) by the ordinary least squares residuals \( \hat{a}_{it} = y_{it} - x_{it}'(X'X)^{-1}X'y \) to get an estimator \( \hat{\rho} \). Under certain regularity conditions \( \hat{\rho} \) is a consistent estimator of \( \rho \) as \( k \to \infty \). Since the absolute value \( |\hat{\rho}| \) may be greater than or equal to 1, we need to truncate \( \hat{\rho} \):

\[
\hat{\rho} = \text{sign}(\hat{\rho}) \min(1-\delta, |\hat{\rho}|). \tag{6.30}
\]
where \( \text{sign}(x) = x/|x| \) when \( x \neq 0 \), or \( = 0 \) otherwise, and \( \delta > 0 \) is some arbitrarily small number. It can be shown that \( \hat{\rho} \) is asymptotically equal to \( \rho^* \).

### 6.4.2 Two-stage or EBLUP estimator

With the estimator \( \hat{\rho} \) from (6.30), estimators of \( \sigma^2 \) and \( \sigma_v^2 \) can be obtained from (6.10) and (6.12) as

\[
\hat{\sigma}^2(\hat{\rho}) = \max\{0, \hat{\sigma}^2(\hat{\rho})\},
\]

\[
\hat{\sigma}_v^2(\hat{\rho}) = \max\{0, \hat{\sigma}_v^2(\hat{\rho})\},
\]

where \( \hat{\sigma}^2 \) and \( \hat{\sigma}_v^2 \) are obtained by substituting \( \hat{\rho} \) for \( \rho \) in equations (6.9) and (6.11) respectively. Hence a two-stage estimator of \( \theta_{IT} \) is obtained from (6.13) as

\[
\hat{\theta}_{IT}(\hat{\rho}) = a[\hat{\sigma}^2(\hat{\rho}), \hat{\sigma}_v^2(\hat{\rho}), \hat{\rho}, y].
\]

This two-stage estimator \( \hat{\theta}_{IT}(\hat{\rho}) \) remains unbiased since \( \hat{\rho} \) and the induced variance estimators \( \hat{\sigma}^2(\hat{\rho}) \) and \( \hat{\sigma}_v^2(\hat{\rho}) \) are even functions of \( y \) and translation invariant.

### 6.4.3 Approximation to MSE and Estimation of MSE

Similar to the procedure in Section 6.3, we first obtain a second-order approximation to \( \text{MSE}(\hat{\theta}_{IT}(\hat{\rho})) \). Following Kackar and Harville (1984), we have

\[
\text{MSE}(\hat{\theta}_{IT}(\hat{\rho})) = \text{MSE}(\tilde{\theta}_{IT}) + E[\hat{\theta}_{IT}(\hat{\rho}) - \tilde{\theta}_{IT}]^2, \tag{6.31}
\]

where \( \text{MSE}(\tilde{\theta}_{IT}) \) is given in (6.15). For the second term in (6.31), following Prasad and Rao (1990), we propose a Taylor series approximation by writing

\[
E[\hat{\theta}_{IT}(\hat{\rho}) - \tilde{\theta}_{IT}]^2
\]

\[
= E[(\hat{\theta}_{IT}(\hat{\rho}) - \varphi(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho, y)) + \varphi(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho, y) - \varphi(\sigma^2, \sigma_v^2, \rho, y)]^2
\]

where \( \varphi(\sigma^2, \sigma_v^2, \rho, y) \) is defined in (6.5). We have

\[
E[\hat{\theta}_{IT}(\hat{\rho}) - \tilde{\theta}_{IT}]^2 \approx E[(d\varphi(\sigma^2, \sigma_v^2, \rho)/d\hat{\rho})(\hat{\rho} - \rho)
\]

\[
+ (\sigma^2, \sigma_v^2, \rho)/\partial \sigma^2)(\hat{\sigma}^2(\rho) - \sigma^2) + (\partial \sigma_v^2, \rho)/\partial \sigma_v^2)(\hat{\sigma}_v^2(\rho) - \sigma_v^2)]^2, \tag{6.32}
\]
where the dependence on $y$ is suppressed for simplicity. Note that the truncated estimators $\hat{\sigma}^2$, $\hat{\sigma}_o^2$ and $\hat{\rho}$ are replaced by their untruncated counterparts $\tilde{\sigma}^2(\rho)$, $\tilde{\sigma}_o^2(\rho)$ and $\tilde{\rho}$ in (6.32). This amounts to ignoring terms of lower order, $o(k^{-1})$, for large $k$ and small or moderate $T$.

Note that in (6.32)

$$\frac{\partial \varphi(\sigma^2, \sigma_o^2, \rho)}{\partial \sigma^2} = (f + \frac{\partial b'_1}{\partial \sigma^2})(Zv + u + e) \approx \frac{\partial b'_1}{\partial \sigma^2}(Zv + u + e) \tag{6.33}$$

since $f = O(k^{-1})$, where

$$b'_1 = (\sigma_o^2 l'_1 Z' + \sigma^2 l'_2 R)V^{-1},$$
$$f = (l' - b'_1 X)(X'V^{-1}X)^{-1}X' \left( \frac{\partial V^{-1}}{\partial \sigma^2} \right) A - \left( \frac{\partial b'_1}{\partial \sigma^2} \right) X(X'V^{-1}X)^{-1}X'V^{-1},$$
$$A = I_n T - X(X'V^{-1}X)^{-1}X'V^{-1},$$

and

$$\frac{\partial b'_1}{\partial \sigma^2} = [l'_2 - (\sigma_o^2 l'_1 Z' + \sigma_o^2 l'_2 R)V^{-1}]RV^{-1}. $$

Similarly,

$$\frac{\partial \varphi(\sigma^2, \sigma_o^2, \rho)}{\partial \sigma_o^2} \approx \frac{\partial b'_1}{\partial \sigma_o^2}(Zv + u + e), \tag{6.34}$$

$$\frac{\partial \varphi(\tilde{\sigma}^2(\rho), \tilde{\sigma}_o^2(\rho), \rho)}{\partial \rho} \approx b'_2(Zv + u + e), \tag{6.35}$$

where

$$\frac{\partial b'_1}{\partial \sigma_o^2} = [l'_1 - (\sigma_o^2 l'_1 Z' + \sigma_o^2 l'_2 R)V^{-1}Z]Z'V^{-1},$$
$$b'_2 = \frac{\partial b'_1}{\partial \sigma_o^2} \cdot \frac{d\tilde{\sigma}^2(\rho)}{d\rho} + \frac{\partial b'_1}{\partial \sigma^2} \cdot \frac{d\tilde{\sigma}_o^2(\rho)}{d\rho} + \sigma^2 l'_2 - (\sigma_o^2 l'_1 Z' + \sigma_o^2 l'_2 R)V^{-1} \frac{dR}{d\rho} V^{-1}. $$

Here $dR/d\rho$ can be evaluated easily from the expression for matrix $R$. The evaluation of derivatives $d\tilde{\sigma}^2(\rho)/d\rho$ and $d\tilde{\sigma}_o^2(\rho)/d\rho$ involves the derivatives of matrices in (6.9) and (6.11) with respect to $\rho$. These expressions are lengthy, and so are omitted.

From the approximations in (6.33), (6.34) and (6.35), the left hand side of (6.32) can be further approximated by

$$E[\hat{\theta}_T(\hat{\rho}) - \hat{\theta}_T]^2 \approx tr(\Delta_1 V \Delta_1 \Sigma^2), \tag{6.36}$$
where $\Delta_1 = (b_2, \partial b_1/\partial \sigma^2, \partial b_1/\partial \sigma_v^2)$ and $\Sigma^a$ is the $3 \times 3$ covariance matrix of $\hat{\rho} - \rho$, $\hat{\sigma}^2(\rho) - \sigma^2$, and $\hat{\sigma}_v^2(\rho) - \sigma_v^2$.

Combining (6.31) and (6.36), we get a second order approximation to $MSE(\hat{\theta}_{iT}(\hat{\rho}))$ as

$$MSE(\hat{\theta}_{iT}(\hat{\rho})) \approx g_{1iT}(\sigma^2, \sigma_v^2, \rho) + g_{2iT}(\sigma^2, \sigma_v^2, \rho) + g_{3iT}(\sigma^2, \sigma_v^2, \rho),$$

(6.37)

where $g_{1iT}$, $g_{2iT}$ are defined in (6.16) and (6.17), and

$$g_{3iT} = \text{tr}(\Delta_1' V \Delta_1 \Sigma^a).$$

(6.38)

The neglected terms in (6.37) are of order $o(k^{-1})$.

For the evaluation of covariance matrix $\Sigma^a$, its elements $Var[\hat{\sigma}^2(\rho)]$, $Var[\hat{\sigma}_v^2(\rho)]$ and $Cov[\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho)]$ can be obtained from (6.20) and (6.21) using Lemma 6.2. For other terms involving $\hat{\rho}$, we use the following approximation. Under certain regularity conditions, as $k$ is quite large and $T$ is small or moderate, $\hat{\rho} - \rho \approx \rho^* - \rho$ to get

$$\hat{\rho} - \rho \approx [k(T - 2)\sigma^2]^{-1}(1 + \rho)\alpha'[(\hat{G}_2 - \hat{G}_3) - \rho(\hat{G}_1 - \hat{G}_2)]\alpha + \text{constant},$$

(6.39)

where $\alpha = Zv + u + e$, $\hat{G}_i = G_i \otimes I_4$, $G_1$ is a $T \times T$ matrix with 1 for elements $(t, t)$, $t = 1, \ldots, T - 2$, 0 elsewhere, $G_2$ is a $T \times T$ matrix with 1/2 for elements $(t, t - 1)$ and $(t - 1, t)$, $t = 2, \ldots, T - 1$, 0 elsewhere, and $G_3$ is a $T \times T$ matrix with 1/2 for elements $(t, t - 2)$ and $(t - 2, t)$, $t = 3, \ldots, T$, 0 elsewhere. Using (6.20), (6.21) and (6.39) in Lemma 6.2, we can approximate $Var(\hat{\rho})$, $Cov(\hat{\rho}, \hat{\sigma}^2(\rho))$, and $Cov(\hat{\rho}, \hat{\sigma}_v^2(\rho))$.

Now we turn to the estimation of MSE. Following Prasad and Rao (1990), we have

$$E[g_{1iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_{1iT}(\sigma^2, \sigma_v^2, \rho) + g_{3iT}(\sigma^2, \sigma_v^2, \rho),$$

(6.40)

$$E[g_{2iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_{2iT}(\sigma^2, \sigma_v^2, \rho),$$

(6.41)

$$E[g_{3iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_{3iT}(\sigma^2, \sigma_v^2, \rho).$$

(6.42)

Hence, it follows from (6.37), (6.40), (6.41) and (6.42) that a second order approximation to the estimator of $MSE(\hat{\theta}_{iT})$ is given by

$$\text{mse}(\hat{\theta}_{iT}) \approx g_{1iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho}) + g_{2iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho}) + 2g_{3iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho}).$$

(6.43)
Although our estimator of MSE, (6.43), is correct to terms of order $O(k^{-1})$, our simulation results indicated difficulties with this method since $\rho$ often took values outside the admissible range, $(-1, 1)$, especially for small $T$ or small $\sigma^2$ relative to sampling variation.

### 6.4.4 Two other alternatives

Although $\hat{\rho}$ is theoretically consistent, the probability of obtaining $\hat{\rho}$ with $|\hat{\rho}| > 1$ is too large for small or moderate number of time points, $T$, as shown in our simulation study. This causes a problem. The value of $|\hat{\rho}|$ (after truncating $\hat{\rho}$) is too close to 1. Since $1 - \rho^2$ appears in the denominator in the formula for MSE($\hat{\theta}_{iT}$), the estimated mean squared error can become unrealistically large. To overcome this problem, we present the following two methods.

**Method 1** In this method, we guess a value $\rho_0$ for $\rho$ from past experience. A two-stage estimator $\hat{\theta}_{iT}(\rho_0)$ based on the prior guess $\rho_0$ is used. And its MSE is estimated by substituting $\rho_0$ for $\rho$ in (6.25). Denote this estimator of MSE as $\text{mse}[\hat{\theta}_{iT}(\rho_0)]$.

**Method 2** We ignore the sampling errors $e_{it}$ and obtain a moment estimator of $\rho$, along the lines of Pantula and Pollock (1985). This naive estimator of $\rho$ is given by

$$\hat{\rho}_N = \frac{\sum_{t=1}^{T} \sum_{i=1}^{T-2} \hat{a}_{it}(\hat{a}_{i,t+1} - \hat{a}_{i,t+2})}{\sum_{t=1}^{T} \sum_{i=1}^{T-2} \hat{a}_{it}(\hat{a}_{i,t} - \hat{a}_{i,t+1})}, \quad T > 2,$$

(6.44)

where $\hat{\epsilon}_{it} = y_{it} - x_{it}'X(X'X)^{-1}X'y$ is the $(it)$-th ordinary least squares residual. The estimator $\hat{\rho}_N$ is inconsistent and typically underestimates $\rho$ (see Section 6.5.2). Nevertheless, the resulting two-stage estimator $\hat{\theta}_{iT}(\hat{\rho}_N)$ remains unbiased. The MSE of $\hat{\theta}_{iT}(\hat{\rho}_N)$ is estimated by substituting $\hat{\rho}_N$ for $\rho$ in (6.25). Denote this estimator of MSE as $\text{mse}[\hat{\theta}_{iT}(\hat{\rho}_N)]$. 
CHAPTER 6. EBLUP ESTIMATORS

6.5 Simulation Study

To study the efficiency of two-stage estimators and relative biases of estimators of MSE, we generated samples \( \{y_{it}\} \) from the following simple model:

\[
y_{it} = v_i + u_{it} + e_{it} \\
u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1
\]

(6.45)

with \( \rho = 0.2 \) and \( 0.4 \), \( e_{it} \overset{iid}{\sim} N(0, 1) \), \( v_i \overset{iid}{\sim} N(0, \sigma_v^2) \), \( \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2) \). (Choudhry and Rao (1989) obtained \( \hat{\rho} = 0.36 \) under model (1.26) using data from the Canadian Labour Force Survey.) We used \( k = 40 \) small areas and both small \( T (= 5) \) and moderate \( T (= 10) \), and generated 5000 independent samples for each selected pair \( (\sigma_v^2, \sigma^2) \). Note that \( \sigma_v^2 \) and \( \sigma^2 \) represent between small area variation relative to sampling variation and between time variation relative to sampling variation respectively since \( \text{var}(e_{it}) = 1 \).

From each simulated sample, the two-stage estimators and the Fay–Herriot estimator of \( \theta_{1T} = v_1 + u_{1T} \), and the estimators of MSE were computed. Note that it is sufficient to consider \( \theta_{1T} \) since MSE values are the same for all \( \theta_{1T}, i = 1, \ldots, 40 \). Simulated values of MSE of any estimator \( \hat{\theta}_{1T} \) (say) and the relative bias of an estimator of MSE, say mse, were computed as follows:

\[
\text{MSE}(\hat{\theta}_{1T}) = \frac{1}{5000} \sum_{s=1}^{5000} (\hat{\theta}_{1T}^s - \theta_{1T})^2,
\]

\[
\text{RB[mse]} = \frac{1}{\text{MSE}(\hat{\theta}_{1T})} \left[ \frac{1}{5000} \sum_{s=1}^{5000} (\text{mse}_s) - \text{MSE}(\hat{\theta}_{1T}) \right],
\]

where \( \hat{\theta}_{1T}^s \), and \( \theta_{1T}^s \) and mse are the values of \( \hat{\theta}_{1T} \), \( \theta_{1T} \) and mse respectively of the \( s \)-th simulation. Simulated percentage values of gain in efficiency of a two-stage estimator over the Fay–Herriot (FH) estimator were computed as follows

\[
\text{GE} = \left( \frac{\text{MSE(FH estimator)}}{\text{MSE(two-stage estimator)}} \right) - 1 \right) \times 100.
\]

6.5.1 Autocorrelation known

Tables 6.1 and 6.2 report the values of GE for the two-stage estimator \( \hat{\theta}_{1T}(\rho) \) with \( \rho = 0.2 \) and \( 0.4 \) respectively. We have the following results from Tables 6.1 and
6.2: (1) Substantial gains in efficiency (GE) are achieved when the between time variation relative to sampling variation is small ($\sigma^2 = 0.25, 0.5$), especially when the between small area variation is substantial ($\sigma_v^2 = 1, 2$): For example, GE = 105% when $T = 10$, $\sigma^2 = 0.25$ and $\sigma_v^2 = 1.0$; (2) GE values for $T = 10$ are significantly large than those for $T = 5$, especially for small $\sigma^2$, i.e., the use of data for more time points improves efficiency of the two-stage estimator. For example for $\rho = 0.4$, $\sigma^2 = 0.25$, $\sigma_v^2 = 1.0$, GE = 105% with $T = 10$ vs 74% with $T = 5$; (3) For a fixed $\sigma_v^2$, GE decreases as $\sigma^2$ increases, whereas it increases with $\sigma_v^2$ for a fixed $\sigma^2$.

Tables 6.3 and 6.4 report the values of relative bias (RB) of the naive estimator of MSE, $\text{mse}_N[\hat{\theta}_{1T}(\rho)]$, given by (6.26), and the improved estimator of MSE, $\text{mse}[\hat{\theta}_{1T}(\rho)]$, given by (6.25), for $\rho = 0.2$ and 0.4 respectively. We have the following results from Tables 6.3 and 6.4: (1) The improved estimator of MSE performs well, leading to slight overestimation (RB $\leq 3\%$ for $T = 5$ and RB $\leq 5\%$ for $T = 10$). (2) The naive estimators of MSE leads to significant underestimation for small $\sigma^2$ and small $T$ ($= 5$), but for other values it is quite satisfactory. For example, RB = $-9.1\%$ when $\sigma^2 = 0.25$, $\sigma^2 = 0.25$, $\rho = 0.4$ and $T = 5$.

6.5.2 Autocorrelation unknown

Method 1. We computed the GE-values for the two-stage estimator $\hat{\theta}_{1T}(\rho_0)$ using prior guess $\rho_0 = 0.1, 0.3$ for true $\rho = 0.2$ and $\rho_0 = 0.2, 0.3, 0.5, 0.6$ for true $\rho = 0.4$. Our result indicate that the GE-values are virtually unaffected by the choice of $\rho_0$, i.e., the two-stage estimator $\hat{\theta}_{1T}(\rho_0)$ retains its efficiency even when the prior guess $\rho_0$ deviates significantly from $\rho$.

We have also computed the RB-values for $\text{mse}[\hat{\theta}_{1T}(\rho_0)]$, the estimator of MSE obtained by substituting $\rho_0$ for $\rho$ in (6.25). Our results in Table 6.5 and 6.6 for $T = 5$ suggest that $\text{mse}[\hat{\theta}_{1T}(\rho_0)]$ performs well when $\rho_0 < \rho$, but it can lead to significant overestimation when $\rho_0$ is significantly larger than $\rho$ and $\sigma^2$ is small. For example RB $\geq 10\%$ when $\rho_0 = 0.6$, $\rho = 0.4$ and $\sigma^2 = 0.25$. Our method 1, therefore, appears to be satisfactory provided the prior guess $\rho_0$ is not significantly larger than the true $\rho$. 
Method 2. We computed the GE-values for the two-stage estimator \( \hat{\theta}_{1T}(\hat{\rho}_N) \) using the naive estimator \( \hat{\rho}_N \) given by (6.44). Again, the GE values are close to those under the true \( \rho \); i.e., the two-stage estimator retains efficiency even when a naive estimator \( \hat{\rho}_N \) is used. It may be noted that \( \hat{\rho}_N \) leads to significant underestimation of \( \rho \) which increases with \( \rho \).

Tables 6.7 and 6.8 report the RB-values for \( \text{mse}[\hat{\theta}_{1T}(\hat{\rho}_N)] \), the estimator of MSE obtained by substituting \( \hat{\rho}_N \) for \( \rho \) in (6.25). Our results in Tables 6.7 and 6.8 suggest that \( \text{mse}[\hat{\theta}_{1T}(\hat{\rho}_N)] \) performs well for small \( \rho \), but it can lead to underestimation which increases with \( \rho \). (For \( T = 5 \) and \( \rho = 0.7 \) we obtained \( \text{RB} = -10\% \)). Thus, our method 2 may be satisfactory when \( \rho \) is expected to be small.

Method 3. Method 3 uses the estimator \( \hat{\rho} \) given by (6.30). It takes into account the sampling errors, but often leads to values outside the admissible range, \((-1, 1)\), especially for small \( T \) or small \( \sigma^2 \) relative to sampling variation. Similar difficulties are encountered under measurement error models (see Fuller, 1987). It would be useful to develop suitable modifications of \( \hat{\rho} \), using methods similar to those in Fuller (1987), that lead to more efficient estimators taking values in the admissible range \((-1, 1)\).
Table 6.1: Gain in efficiency (%) of the two-stage estimator $\hat{\theta}_{1T}(\rho)$ over the Fay–Herriot estimator: $\rho = 0.2$

<table>
<thead>
<tr>
<th>$\sigma_v^2$</th>
<th>$\sigma^2$ = 0.25</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>32</td>
<td>19</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>0.5</td>
<td>51</td>
<td>29</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>1.0</td>
<td>77</td>
<td>45</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>2.0</td>
<td>107</td>
<td>64</td>
<td>32</td>
<td>15</td>
</tr>
</tbody>
</table>

| $T = 10$ | | | | |
| 0.25 | 49 | 26 | 13 | 6 |
| 0.5 | 76 | 40 | 18 | 8 |
| 1.0 | 114 | 60 | 28 | 11 |
| 2.0 | 157 | 86 | 41 | 27 |

Table 6.2: Gain in efficiency (%) of the two-stage estimator $\hat{\theta}_{1T}(\rho)$ over the Fay–Herriot estimator: $\rho = 0.4$

<table>
<thead>
<tr>
<th>$\sigma_v^2$</th>
<th>$\sigma^2$ = 0.25</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>33</td>
<td>21</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>0.5</td>
<td>50</td>
<td>30</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>1.0</td>
<td>74</td>
<td>44</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>2.0</td>
<td>101</td>
<td>62</td>
<td>33</td>
<td>16</td>
</tr>
</tbody>
</table>

| $T = 10$ | | | | |
| 0.25 | 47 | 27 | 15 | 9 |
| 0.5 | 71 | 39 | 20 | 10 |
| 1.0 | 105 | 57 | 28 | 13 |
| 2.0 | 144 | 80 | 40 | 19 |
Table 6.3: Relative bias (%) of $\text{mse}[\hat{\theta}_{1T}(\rho)]$ and $\text{mse}_{N}[\hat{\theta}_{1T}(\rho)]$ (in brackets): 
$\rho = 0.2$

<table>
<thead>
<tr>
<th>$\sigma_v^2$</th>
<th>$\sigma^2$</th>
<th>0.25</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>2.4(-7.2)</td>
<td>1.7(-4.1)</td>
<td>1.8(-1.5)</td>
<td>1.6(-0.1)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>2.5(-5.6)</td>
<td>1.8(-3.3)</td>
<td>1.7(-1.2)</td>
<td>1.4(-0.2)</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>2.7(-4.3)</td>
<td>1.9(-2.5)</td>
<td>1.7(-0.8)</td>
<td>1.3(-0.1)</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>2.7(-3.6)</td>
<td>1.9(-2.1)</td>
<td>1.7(-0.6)</td>
<td>1.2(-0.0)</td>
<td></td>
</tr>
<tr>
<td>$T = 10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>4.5(-0.6)</td>
<td>4.1(1.2)</td>
<td>3.5(2.0)</td>
<td>2.8(2.0)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>4.4(-0.1)</td>
<td>4.0(1.5)</td>
<td>3.5(2.1)</td>
<td>2.8(2.0)</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>4.2(0.2)</td>
<td>3.9(1.6)</td>
<td>3.4(2.1)</td>
<td>2.7(2.0)</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>4.1(0.3)</td>
<td>3.8(1.7)</td>
<td>3.3(2.1)</td>
<td>2.6(2.0)</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.4: Relative bias (%) of $\text{mse}[\hat{\theta}_{1T}(\rho)]$ and $\text{mse}_{N}[\hat{\theta}_{1T}(\rho)]$ (in brackets): 
$\rho = 0.4$

<table>
<thead>
<tr>
<th>$\sigma_v^2$</th>
<th>$\sigma^2$</th>
<th>0.25</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>2.9(-9.1)</td>
<td>1.3(-5.5)</td>
<td>1.6(-1.8)</td>
<td>1.7(-0.1)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>2.8(-7.7)</td>
<td>1.2(-4.9)</td>
<td>1.4(-1.8)</td>
<td>1.4(-0.3)</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>2.9(-6.5)</td>
<td>1.3(-4.2)</td>
<td>1.4(-1.5)</td>
<td>1.2(-0.3)</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>2.9(-5.7)</td>
<td>1.3(-3.7)</td>
<td>1.4(-1.2)</td>
<td>1.2(-0.2)</td>
<td></td>
</tr>
<tr>
<td>$T = 10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>4.3(-2.6)</td>
<td>3.9(0.4)</td>
<td>3.5(1.7)</td>
<td>2.9(2.1)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>4.3(-1.9)</td>
<td>3.9(0.7)</td>
<td>3.4(1.9)</td>
<td>2.8(2.0)</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>4.2(-1.6)</td>
<td>3.8(0.9)</td>
<td>3.4(2.0)</td>
<td>2.7(2.0)</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>4.1(-1.5)</td>
<td>3.7(1.0)</td>
<td>3.3(2.0)</td>
<td>2.6(2.0)</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.5: Relative bias (%) of \( \text{mse}(\hat{\theta}_{1T}(\rho_0)) \): \( \rho_0 = 0.1, 0.3, \rho = 0.2; T = 5 \)

<table>
<thead>
<tr>
<th>( \sigma_v^2 )</th>
<th>( \sigma^2 = 0.25 )</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_0 = 0.1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>1.0</td>
<td>0.5</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.5</td>
<td>1.1</td>
<td>0.5</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>1.0</td>
<td>1.1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>2.0</td>
<td>1.1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>( \rho_0 = 0.3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>4.6</td>
<td>3.4</td>
<td>3.0</td>
<td>2.5</td>
</tr>
<tr>
<td>0.5</td>
<td>4.7</td>
<td>3.5</td>
<td>2.9</td>
<td>2.1</td>
</tr>
<tr>
<td>1.0</td>
<td>4.8</td>
<td>3.7</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0</td>
<td>4.9</td>
<td>3.8</td>
<td>3.1</td>
<td>2.1</td>
</tr>
</tbody>
</table>
Table 6.6: Relative bias (%) of $\text{mse}[\hat{\theta}_T(\rho_0)]$: $\rho_0 = 0.2, 0.3, 0.5, 0.6$. $\rho = 0.4$; $T = 5$

<table>
<thead>
<tr>
<th>$\sigma_r^2$</th>
<th>$\sigma^2$ = 0.25</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_0 = 0.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>-1.4</td>
<td>-2.0</td>
<td>-1.1</td>
<td>-0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>-1.5</td>
<td>-2.2</td>
<td>-1.3</td>
<td>-0.5</td>
</tr>
<tr>
<td>1.0</td>
<td>-1.5</td>
<td>-2.3</td>
<td>-1.5</td>
<td>-0.7</td>
</tr>
<tr>
<td>2.0</td>
<td>-1.5</td>
<td>-2.4</td>
<td>-1.7</td>
<td>-1.0</td>
</tr>
</tbody>
</table>

| $\rho_0 = 0.3$ |                  |     |     |     |
| 0.25         | 0.3             | -0.6| 0.2 | 0.6 |
| 0.5          | 0.3             | -0.7| 0.0 | 0.5 |
| 1.0          | 0.4             | -0.7| -0.1| 0.3 |
| 2.0          | 0.3             | -0.7| -0.2| 0.1 |

| $\rho_0 = 0.5$ |                  |     |     |     |
| 0.25         | 6.4             | 3.6 | 3.5 | 2.9 |
| 0.5          | 6.1             | 3.4 | 2.3 | 2.4 |
| 1.0          | 6.2             | 3.6 | 2.9 | 2.0 |
| 2.0          | 6.2             | 3.7 | 3.0 | 2.0 |

| $\rho_0 = 0.6$ |                  |     |     |     |
| 0.25         | 11.2            | 7.0 | 5.8 | 4.0 |
| 0.5          | 10.2            | 6.8 | 4.6 | 3.5 |
| 1.0          | 10.2            | 5.9 | 4.1 | 2.8 |
| 2.0          | 10.3            | 6.1 | 4.3 | 2.6 |
Table 6.7: Relative bias (%) of $\text{mse}[\hat{\theta}_1T(\hat{\rho}_N)]$: $\rho = 0.2$

<table>
<thead>
<tr>
<th>$\sigma_\varphi^2$</th>
<th>$\sigma^2$ = 0.25</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>-1.0</td>
<td>-0.7</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.8</td>
<td>-0.7</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.6</td>
<td>-0.6</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>2.0</td>
<td>-0.4</td>
<td>-0.5</td>
<td>0.0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

| $T = 10$          |                   |     |     |     |
| 0.25              | 1.4              | 1.8 | 2.1 | 2.0 |
| 0.5               | 1.2              | 1.6 | 2.0 | 1.9 |
| 1.0               | 1.0              | 1.4 | 1.8 | 1.8 |
| 2.0               | 0.8              | 1.2 | 1.6 | 1.7 |

Table 6.8: Relative bias (%) of $\text{mse}[\hat{\theta}_1T(\hat{\rho}_N)]$: $\rho = 0.4$

<table>
<thead>
<tr>
<th>$\sigma_\varphi^2$</th>
<th>$\sigma^2$ = 0.25</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>-4.0</td>
<td>-3.8</td>
<td>-1.8</td>
<td>-0.3</td>
</tr>
<tr>
<td>0.5</td>
<td>-4.0</td>
<td>-3.9</td>
<td>-2.1</td>
<td>-0.5</td>
</tr>
<tr>
<td>1.0</td>
<td>-3.9</td>
<td>-4.0</td>
<td>-2.3</td>
<td>-0.7</td>
</tr>
<tr>
<td>2.0</td>
<td>-3.7</td>
<td>-4.0</td>
<td>-2.4</td>
<td>-0.9</td>
</tr>
</tbody>
</table>

| $T = 10$          |                   |     |     |     |
| 0.25              | -2.3             | -1.6| -0.2| 1.1 |
| 0.5               | -2.6             | -1.9| -0.4| 0.9 |
| 1.0               | -3.0             | -2.3| -0.7| 0.8 |
| 2.0               | -3.2             | -2.5| -0.9| 0.6 |
Chapter 7

Hierarchical Bayes Analysis

In this Chapter, we apply the hierarchical Bayes (HB) method to our model (6.1). In Section 7.1, we consider the case of $\rho$ known. Two methods for implementing the HB approach, the numerical integration and Gibbs sampler, are studied. In Section 7.2, the case of $\rho$ unknown is briefly discussed.

7.1 HB Estimator: $\rho$ Known

Consider our model (6.1). Similar to (6.2), the model can be written as

$$y_i = \theta_i + e_i = X_i\beta + u_i1_T + u_i + e_i$$  \hspace{1cm} (7.1)

where $\theta_i = (\theta_{i1}, \ldots, \theta_{iT})'$, $y_i$, $X_i$, $e_i$ are the same as those for (6.2). With $\rho$ known, we use the following four-stage hierarchical Bayes model.

(I) Conditional on $\theta_{ii}$'s, $y_i \overset{i.i.d.}{\sim} N(\theta_i, \Sigma_i)$.

(II) Conditional on $u_i$, $\beta$, and $r_1 = 1/s^2$, $\theta_i \overset{i.i.d.}{\sim} N(X_i\beta + u_i1_T, r_1^{-1}\Gamma)$.

(III) Conditional on $r_2 = 1/s^2$, $u_i \sim N(0, r_2^{-1}I_T)$.

(IV) The marginal distribution of $\beta$, $r_1$, and $r_2$ are assumed independent with

$\beta \sim \text{uniform}(R^n)$,

$r_1 \sim \text{Gamma}(a_1/2, b_1/2),$

$r_2 \sim \text{Gamma}(a_2/2, b_2/2),$

where $\text{Gamma}(a, b)$ is a Gamma random variable with density function

$$g(r) = a^b r^{b-1} e^{-ar}/\Gamma(b), \quad \text{for } r > 0, \ a > 0 \ \text{and} \ b > 0.$$
CHAPTER 7. HIERARCHICAL BAYES ANALYSIS

Note that model (7.1) (with normal errors) is the combination of stages (I), (II) and (III). In stage (IV), the parameter \( \beta \) is assigned a diffuse prior. Also, \( a_i \) and \( b_i \) \((i = 1, 2)\) are some prior constants. Small values of \( a_i \), \( b_i \) are usually used to reflect absence of prior information on the parameters.

In the HB approach, we want to find the posterior distribution of \( \theta_{iT} \) given the observations \( \{y_{it}\} \), say \( g(\theta_{iT} | y) \). The parameter of interest \( \theta_{iT} \) is estimated by \( E(\theta_{iT} | y) \) and its precision is measured by \( V(\theta_{iT} | y) \). Unfortunately, an explicit form of \( g(\theta_{iT} | y) \) is not available due to model complexity. There are at least two approaches to obtaining the posterior expectation and posterior variance of \( \theta_{iT} \) given data \( \{y_{it}\} \): numerical integration and Gibbs sampler.

7.1.1 Numerical integration

From the four steps in our hierarchical Bayes model, we can write the joint probability density function (p.d.f.) of \( y, \theta, v, \beta, r_1 \) and \( r_2 \), as

\[
g(y, \theta, v, \beta, r_1, r_2) \propto \exp\left\{- (y - \theta)' \Sigma^{-1} (y - \theta) / 2 \right\}
\times r_1^{kT/2} \exp\left\{- r_1 (\theta - X \beta - Z v)' R^{-1} (\theta - X \beta - Z v) / 2 \right\}
\times r_2^{-k/2} \exp\left\{- r_2 v' v / 2 \right\} \times r_1^{b_1/2 - 1} e^{-a_1 r_1 / 2} \times r_2^{b_2/2 - 1} e^{-a_2 r_2 / 2},
\]

where \( \theta = (\theta_{11}, \ldots, \theta_{iT}, \ldots, \theta_{k1}, \ldots, \theta_{kT})' \), \( y, X, v, R, Z \) and \( \Sigma \) are the same as those defined in (6.2).

Integrating over \( v \), we can find that the joint p.d.f. of \( y, \theta, \beta, r_1 \) and \( r_2 \) is

\[
g(y, \theta, \beta, r_1, r_2) \propto \exp\left\{- (y - \theta)' \Sigma^{-1} (y - \theta) / 2 \right\}
\times r_1^{kT/2} r_2^{k/2} |r_1 1_T \Gamma^{-1} 1_T + r_2|^{-k/2} \times \exp\left\{- r_1 (\theta - X \beta)' U^{-1} (\theta - X \beta) / 2 \right\}
\times r_1^{b_1/2 - 1} e^{-a_1 r_1 / 2} \times r_2^{b_2/2 - 1} e^{-a_2 r_2 / 2},
\]

where \( U = \text{block diag}_i (\sigma_i^2 1_T 1'_T + \sigma_i^2 \Gamma) \). Further, denote

\[ U_1 = X' U^{-1} X. \]
CHAPTER 7. HIERARCHICAL BAYES ANALYSIS

From the following identity,

\[(\theta - X\beta)'U^{-1}(\theta - X\beta)\]
\[= (\beta - U_1^{-1}X'U^{-1}\theta)'U_1(\beta - U_1^{-1}X'U^{-1}\theta) + \theta'(U_1^{-1} - U_1^{-1}XU_1^{-1}X'U^{-1})\theta,\]

we can obtain the joint p.d.f. of \(y, \theta, r_1\) and \(r_2\) (by integrating over \(\beta\) in (7.3)) as

\[
g(y, \theta, r_1, r_2)
\propto r_1^{(n^* p)/2} r_2^{k/2} |r_1|^{1/2} |r_2|^{k/2} |U_1|^{1/2}
\times \exp\{-(y - \theta)'\Sigma^{-1}(y - \theta)/2\}
\times \exp\{-r_1 \theta'(U^{-1} - U_1^{-1}XU_1^{-1}X'U^{-1})\theta/2\}
\times r_1^{b_1/2} e^{-a_1 r_1/2} \times r_2^{b_2/2 - 1} e^{-a_2 r_2/2}.
\]

It can be seen here that further integration over \(r_1\) or \(r_2\) is difficult since \(r_1\) and \(r_2\) appear in the determinants of \((r_1 1_\Gamma^{-1} 1_T + r_2)\) and \(U_1\). Nonetheless, noting that

\[(y - \theta)'\Sigma^{-1}(y - \theta) + r_1 \theta'(U^{-1} - U_1^{-1}XU_1^{-1}X'U^{-1})\theta
= (\theta - W\Sigma^{-1}y)'W^{-1}(\theta - W\Sigma^{-1}y) + y'(\Sigma^{-1} - \Sigma^{-1}W\Sigma^{-1})y\]

where

\[W^{-1} = \Sigma^{-1} + r_1 U^{-1} - r_1 U_1^{-1}XU_1^{-1}X'U^{-1},\]

we can see that the conditional distribution given \(y, r_1\) and \(r_2\) is of form

\[
g(\theta \mid y, r_1, r_2) \propto \exp\{-(\theta - W\Sigma^{-1}y)'W^{-1}(\theta - W\Sigma^{-1}y)\}.
\]

Hence the conditional expectation and variance of \(\theta\) given \(y, r_1\) and \(r_2\) are

\[
M(y, r_1, r_2) = E(\theta \mid y, r_1, r_2) = W\Sigma^{-1}y,
\]

\[
V(y, r_1, r_2) = V(\theta \mid y, r_1, r_2) = W.
\]

It is interesting to note that \(W\Sigma^{-1}y\) is exactly the BLUP we obtained in Section 6.1 for \(\theta\) due to the following fact

\[W = \Sigma - \Sigma V^{-1}\Sigma + \Sigma V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\Sigma,\]
Further theoretical results are not available. But in a practical situation with particular design matrices, it might be possible to get a better picture of the quantities in (7.4), and from (7.4), we can obtain the joint p.d.f. of $r_1$ and $r_2$ given $y$, say, $g(r_1, r_2 \mid y)$. We can compute the posterior expectation and the posterior covariance matrix of $\theta$ as

$$E(\theta \mid y) = \iint M(y, r_1, r_2)g(r_1, r_2 \mid y)dr_1dr_2$$

and

$$Cov(\theta \mid y) = \iint V(y, r_1, r_2)g(r_1, r_2 \mid y)dr_1dr_2 + \iint [M(y, r_1, r_2) - E(\theta \mid y)][M(y, r_1, r_2) - E(\theta \mid y)]'g(r_1, r_2 \mid y)dr_1dr_2.$$ 

Closed-form expressions for $E(\theta \mid y)$ and $Cov(\theta \mid y)$ cannot be obtained, but numerical integration can be used. In the present case, we need to numerically solve two-dimensional integrals (see Ghosh and Lahari, 1987 for an example of implementing numerical integration).

### 7.1.2 Gibbs sampling

The second approach uses the currently popular Gibbs sampling. The Gibbs sampling was formally introduced in Geman and Geman (1984), and made more popular by Gelfand and Smith (1990). It is a Monte Carlo method for estimating the posterior distribution by drawing samples from the conditional distribution of one parameter given prior values for all other parameters, and updating these prior values step by step. We now review this method.

First, consider a collection of random variables $U_1, U_2, \ldots, U_m$. Denote the marginal, conditional, and joint probability density function as $[U_1]$, $[U_1 \mid U_2]$, and $[U_1, U_2]$ for a pair of variables $U_1$ and $U_2$, and similarly for three or more variables. Assume that the conditional probability density of one variable $U_i$ given all the other variables $U_s$ ($s \neq i$), i.e., $[U_i \mid U_s, s \neq i]$, has explicit expression. We are interested in
the marginal density \([\mathcal{L}_i], i = 1, \ldots, m\). The implementation of the Gibbs sampling scheme is as follows. Given an arbitrary starting set of values \(l_1^{(0)}, l_2^{(0)}, \ldots, l_m^{(0)}\),
draw \(l_1^{(1)}\) from \([l_1 \mid l_2^{(0)}, l_3^{(0)}, \ldots, l_m^{(0)}]\), \(l_2^{(1)}\) from \([l_2 \mid l_1^{(1)}, l_3^{(0)}, \ldots, l_m^{(0)}]\), and so on, up to \(l_m^{(1)}\) from \([l_m \mid l_1^{(1)}, \ldots, l_{m-1}^{(1)}]\). Thus each variable is visited once in this order, and a cycle is completed after \(m\) random number generations. This process is continued with \(l_1^{(1)}, l_2^{(1)}, \ldots, l_m^{(1)}\). After a large number, \(B\), of iterations, we would obtain an \(m\)-tuple \(l_1^{(B)}, l_2^{(B)}, \ldots, l_m^{(B)}\). It is shown in Geman and Geman (1984) that the joint distribution of the above \(m\)-tuple converges at an exponential rate to \([U_1, U_2, \ldots, U_m]\).

Gelfand and Smith (1990) suggest a way of obtaining \(G\) such \(m\)-tuples by retaining the samples \(l_1^{(B+tg)}, l_2^{(B+tg)}, \ldots, l_m^{(B+tg)}\), where \(g = 1, 2, \ldots, G\) and \(t \approx 50\).

A more efficient estimate to the marginal density of \(U_i, [U_i]\), can be constructed by using the following sample-based estimate

\[
\frac{1}{G} \sum_{g=1}^{G} [U_i \mid U_j^{(B+tg)}, j \neq i].
\]

We now apply Gibbs sampling to our hierarchical model (I)-(IV). We obtain the following desired conditional density distributions:

(i) \(\theta \mid v, \beta, r_1, r_2, y \sim N[(\Sigma^{-1} + r_1 R^{-1})^{-1}(\Sigma^{-1}y + r_1 R^{-1}(X\beta + Zv)), (\Sigma^{-1} + r_1 R^{-1})^{-1}]\) (which is independent of \(r_2\)),

(ii) \(v \mid \theta, \beta, r_1, r_2, y \sim N[(r_1 Z' R^{-1}Z + r_2 I_T)^{-1}r_1 Z' R^{-1}(\theta - X\beta), (r_1 Z' R^{-1}Z + r_2^{-1} I_T)^{-1}]\) (which is independent of \(y\)),

(iii) \(\beta \mid \theta, v, r_1, r_2, y \sim N[(X' R^{-1}X)^{-1}X' R^{-1}(\theta - Zv), (X' R^{-1}X)^{-1}]\) (which is independent of \(y\) and \(r_2\)),

(iv) \(r_1 \mid \theta, v, \beta, r_2, y \sim Gamma[\frac{1}{2}a_1 + (\theta - X\beta - Zv)' R^{-1}(\theta - X\beta - Zv)], \frac{1}{2}(kT + b_1)]\) (independent of \(y\) and \(r_2\)),

(v) \(r_2 \mid \theta, v, \beta, r_1, y \sim Gamma[\frac{1}{2}(a_2 + v'v), \frac{1}{2}(k + b_2)]\) (independent of \(y, \theta, r_1\)).

Random variates can be readily generated from the conditional distributions (i) to (v) since only normal and gamma distributions are involved.

Apply the Gibbs sampling procedure, a sample based density estimator of \(\theta\) given
$y$ can be approximated by

$$
\frac{1}{G} \sum_{g=1}^{G} \{ \theta \mid v = \nu^{(B+tg)}, \beta = \beta^{(B+tg)}, r_1 = r_1^{(B+tg)}, r_2 = r_2^{(B+tg)}, y \}.
$$

The posterior expectation and posterior covariance matrix of $\theta$, can be approximated from the above conditional distributions by

$$
E(\theta \mid y) \approx \frac{1}{G} \sum_{g=1}^{G} \Lambda_g,
$$

(7.7)

and

$$
COV(\theta \mid y)
\approx \frac{1}{G} \sum_{g=1}^{G} COV(\theta \mid v = \nu^{(B+tg)}, \beta = \beta^{(B+tg)}, r_1 = r_1^{(B+tg)}, r_2 = r_2^{(B+tg)}, y)
+ \frac{1}{G-1} \left[ \sum_{g=1}^{G} \Lambda_g \Lambda_g' - \frac{1}{G} \sum_{g=1}^{G} \Lambda_g \sum_{i=1}^{G} \Lambda_g' \right],
$$

(7.8)

where

$$
\Lambda_g = E(\theta \mid v = \nu^{(B+tg)}, \beta = \beta^{(B+tg)}, r_1 = r_1^{(B+tg)}, r_2 = r_2^{(B+tg)}, y).
$$

Above quantities can be readily evaluated from the mean and covariance matrix in conditional distribution $[\theta \mid v, \beta, r_1, r_2, y]$ in (i).

Ghosh and Nangia (1993) use a similar model and present a practical application to estimation of median income of four-person families.

### 7.2 HB Estimation: $\rho$ Unknown

The HB estimation for the case of unknown $\rho$ turns out to be much more difficult due to the inclusion of unknown correlation $\rho$. The steps for HB analysis, however, can follow those in Section 7.1. We sketch these steps in this section.

We use a hierarchical model similar to (I)-(IV) in Section 7.1:

(I') is same as (I).

(II') Conditional on $v_i$'s, $\beta$, $r_1$ and $\rho$, $\theta_i \overset{ind}{\sim} N(X_i\beta + v_i1_T, r_1^{-1}\Gamma)$.
(III)' is same as (III).

(IV)' Assume the same prior distributions on $\beta$, $r_1$ and $r_2$ as in (IV).

(V)' Note that difficulty arises regarding the choice of prior distribution for $\rho$. Most papers in the literature avoid this difficulty by assuming a known $\rho$ (the case in Section 7.1), e.g., random walk. A plausible prior distribution for $\rho$ is the uniform distribution over admissible region $(-1,1)$.

Again, similar to the two methods in Section 7.1, there are two approaches of approximating the posterior expectation and posterior covariance of the parameter of interest.

In the first approach, the posterior mean and the posterior covariance matrix of $\theta = (\theta_1', \ldots, \theta_k')'$ given $y$, $r_1$, $r_2$ and $\rho$ and the posterior distribution of $r_1$, $r_2$ and $\rho$ given $y$ are first obtained analytically, using the diffuse prior distributions assumed at stages (IV)' and (V)' of the hierarchical model, similar to the derivation of equations (7.5) and (7.6). Denoting these quantities by $M(y, r_1, r_2, \rho)$, $V(y, r_1, r_2, \rho)$ and $g(r_1, r_2, \rho | y)$, we can compute the posterior mean and the posterior covariance matrix of $\theta$ as

$$E(\theta | y) = \iint M(y, r_1, r_2, \rho)g(r_1, r_2, \rho | y)dr_1dr_2d\rho,$$

and

$$Cov(\theta | y) = \iint V(y, r_1, r_2, \rho)g(r_1, r_2, \rho | y)dr_1dr_2d\rho$$

$$+ \iint [M(y, r_1, r_2, \rho) - E(\theta | y)][M(y, r_1, r_2, \rho)$$

$$- E(\theta | y)]'g(r_1, r_2, \rho | y)dr_1dr_2d\rho.$$}

However, closed-form expression for $E(\theta | y)$ and $Cov(\theta | y)$ cannot be obtained, and numerical integration becomes necessary. In the present case, we need to numerically solve three-dimensional integrals.

The second approach uses Gibbs sampling. The desired posterior mean and posterior covariance matrix of $\theta$ given $y$ can be obtained through an iterative Monte–Carlo procedure by sampling from the full conditional distributions $g(\theta | \beta, v, r_1, r_2, \rho, y)$, $g(v | \theta, \beta, r_1, r_2, \rho, y)$, $g(\beta | \theta, v, r_1, r_2, \rho, y)$, $g(r_1 | \theta, \beta, v, r_2, \rho, y)$,
\( g(r_2 \mid \theta, \beta, v, r_1, \rho, y) \) and \( g(\rho \mid \theta, \beta, v, r_1, r_2, y) \) which determine the joint distribution of \( \theta, \beta, v, r_1, r_2, \rho \) conditional on \( y \). The desired full conditional distributions are obtained from (I')-(IV') as follows:

(i') \( \theta \mid v, \beta, r_1, r_2, \rho, y \sim N((\Sigma^{-1} + r_1 R^{-1})^{-1}(\Sigma^{-1}y + r_1 R^{-1}(X\beta + Zv)), (\Sigma^{-1} + r_1 R^{-1})^{-1}) \) (independent of \( r_2 \)),

(ii') \( v \mid \theta, \beta, r_1, r_2, \rho, y \sim N((r_1 Z' R^{-1} Z + r_2 I_T)^{-1} r_1 Z' R^{-1} (\theta - X\beta), (r_1 Z' R^{-1} Z + r_2^{-1} I_T)^{-1}) \) (independent of \( y \)),

(iii') \( \beta \mid \theta, v, r_1, r_2, \rho, y \sim N((X' R^{-1} X)^{-1} \times X' R^{-1} (\theta - Zv), (X' R^{-1} X)^{-1}) \) (independent of \( y \) and \( r_2 \)),

(iv') \( r_1 \mid \theta, v, \beta, r_2, \rho, y \sim Gamma[\frac{1}{2}(a_1 + (\theta - X\beta - Zv)' R^{-1} (\theta - X\beta - Zv)), \frac{1}{2}(kT + b_1)] \) (independent of \( y \) and \( r_2 \)),

(v') \( r_2 \mid \theta, v, \beta, r_1, \rho, y \sim Gamma[\frac{1}{2}(a_2 + v' v), \frac{1}{2}(k + b_2)] \) (independent of \( y, \theta, r_1 \)),

(vi') \( \rho \mid \theta, v, \beta, r_1, r_2, y = [c(\beta, \theta, v, r_1)]^{-1} |R|^{-1/2} \exp\left(-\frac{1}{2}(\theta - X\beta - Zv)' R^{-1} (\theta - X\beta - Zv)\right) \) (independent of \( y, r_2 \)) where \( g(\rho) \) is the prior probability density imposed on \( \rho \), and

\[
c(\beta, \theta, v, r_1) = \int |R|^{-1/2} \exp\left(-\frac{1}{2}(\theta - X\beta - Zv)' R^{-1} (\theta - X\beta - Zv)\right) g(\rho) d\rho.
\]

Random variates can be readily generated from the conditional distributions (i') to (v') since only normal and gamma distributions are involved. The conditional distribution (vi') does not have a closed form, but can be generated using rejection sampling without evaluating the integral \( c(\beta, \theta, v, r_1) \) (for example, see Zeger and Karim, 1991). After complete Gibbs sampling, the posterior expectation and variance can be approximated in the same fashion as those in (7.7) and (7.8).
Chapter 8

Summary and Proposed Future Research

Our results in Part I and Part II of this thesis are discussed in Sections 8.1 and 8.2, respectively. Some proposals for future research are then made in Section 8.3.

8.1 Summary for Part I

Part I of this thesis deals with one and two-fold nested error linear regression models with unequal error variances. Chapters 2 to 4 are devoted to the former model, while Chapter 5 is devoted to the latter model. For these models, approximate minimum norm quadratic unbiased (AMINQU) estimators are derived by ignoring the lower order terms in the minimum norm quadratic unbiased (MINQU) estimation equations (C. R. Rao and Kleffe, 1988). The AMINQU estimators are biased, but they have smaller mean squared errors than the unbiased analysis of variance (ANOVA) and MINQU estimators as shown in a simulation study. The AMINQU estimator of the variance component $\sigma^2_e$ is shown to be asymptotically consistent and normal. A consistent estimator of the asymptotic variance is obtained, which enables us to construct confidence intervals for the variance component. These confidence intervals perform well but lead to slight undercoverage when $\sigma^2_e$ is small or moderate relative to error variances.

Under the one-fold nested-error model, the AMINQU estimators after trunca-
tion generate positive estimators of weights for weighted least squares (WLS) estimation of the regression parameters $\beta$, unlike ANOVA or MINQU estimators. Using the truncated AMINU estimators, we have obtained a WLS estimator of regression parameters $\beta$ in Chapter 3. The asymptotic normality of this WLS estimator of $\beta$ is established. A consistent estimator of the asymptotic covariance matrix is obtained by the substitution method. The variance estimator is shown empirically to be accurate and stable, especially when $\sigma_0^2$ is moderate or large relative to error variances $\sigma_i^2$. We have also constructed confidence intervals on the elements of $\beta$ using the WLS estimator and the corresponding variance estimator. These confidence intervals lead to accurate coverage probability. An alternative WLS estimator of $\beta$ is also considered, using the AMINU estimator of $\sigma_0^2$ and the within-group sample variances as estimators of error variances $\sigma_i^2$. Its asymptotic properties are established. But in the case of a small number of observations within a group, and small or moderate $\sigma_0^2$ relative to $\sigma_i^2$, the corresponding substitution variance estimator is not stable.

For the one-fold nested error model, the delete-group jackknife method is applied to the AMINU estimators to reduce the bias in Chapter 4. It is also used to estimate the variance of the AMINU estimators as well as the covariance matrix of the WLS estimators of regression parameters. Consistency of the jackknife variance and covariance estimators is established and asymptotically correct jackknife confidence intervals on the parameters are obtained. In a limited simulation study, we have found that the jackknife method reduces the bias in the untruncated and truncated AMINU estimators. We have also found that jackknife variance estimators for the elements of $\beta$ are not as stable as the corresponding substitution variance estimators. However, the jackknife variance estimators may be more robust against possible misspecification of the within group covariance structure. Our WLS estimator and the associated substitution variance estimator lead to more accurate and higher coverage probability than the WLS estimator and associated delete-group jackknife variance estimator proposed in Shao and Rao (1993). Their method leads to slight undercoverage ($< 5\%$).
In Chapter 5, the AMINQU estimation approach is applied to the two-fold nested-error regression model. Due to model complexity, explicit expressions for the AMINQU estimators of variance components could be obtained using only a special set of prior values. Asymptotic properties of these AMINQU estimators are established. Since our AMINQU estimators cannot guarantee positive weights for the WLS estimator of $\beta$, alternative WLS estimators are proposed using within-group sample variances as estimators of error variances. Asymptotic normality of the resulting WLS estimator of $\beta$ is also established, and a consistent estimator of the asymptotic covariance matrix is obtained.

8.2 Summary for Part II

In Part II of this thesis, we have proposed a model for small area estimation, involving autocorrelated random effects and sampling errors, using both time series and cross-sectional data. This model is an extension of a well-known model, due to Fay and Herriot (1979), for cross-sectional data. Chapter 6 is devoted to two-stage estimation of small area means under this model. We first assume that the autocorrelation, $\rho$, is known and obtain a two-stage estimator of a small area mean for the current period. An approximation to its mean squared error (MSE), correct to second order terms, is obtained. An estimator of MSE is provided, also correct to the same order terms. For the case of unknown $\rho$, we have obtained a consistent estimator of $\rho$, but it did not perform well in small or moderate samples. Therefore, two alternatives methods are proposed and the corresponding two-stage estimators and estimators of their MSE are obtained. Method 1 is based on a prior guess of $\rho$, while method 2 uses a naive estimator of $\rho$ that ignores the sampling errors.

Our simulation results in Section 6.5 have shown that the two-stage estimators can lead to substantial gain in efficiency over the Fay-Herriot estimator which uses only the current cross-sectional data, especially when the between time variation relative to sampling variation is small. Further, our method 1 of estimating MSE may be satisfactory, provided the prior guess of $\rho$ is not significantly larger than the
true $\rho$. Our method 2 of estimating MSE may also be satisfactory when $\rho$ is expected to be small. We have also noted that our method, based on a consistent estimator $\hat{\rho}$ that takes sampling errors into account, runs into difficulties since $\hat{\rho}$ often leads to values outside the admissible range $(-1, 1)$, especially for a small number of time points $T$ or when the between time variation relative to sampling variation is small. Suitable modifications of $\hat{\rho}$ that can lead to more efficient estimators taking values in the admissible range are needed.

In Chapter 7, a hierarchical Bayes approach is studied. In the HB approach, a prior distribution on all the model parameters is specified and the posterior distribution of the parameters of interest is then obtained. Inferences are based on the posterior distribution; in particular, a current small area mean is estimated by its posterior mean and its precision is measured by its posterior variance. The HB approach is computer intensive, but the computations can be handled using Gibbs sampling or numerical integration.

### 8.3 Proposed Further Research

This thesis has addressed inferential problems related to two types of models, viz., nested-error regression models and small area models for time series and cross-sectional data. We plan to do further work on these problems as outlined below.

First, it is necessary to develop suitable model diagnostics to validate the nested error regression models assumed in Part I. There has been previous work on model diagnostics for nested error models assuming equal error variances (see Ghosh and Rao, 1993 for a review). Also, as noted in Section 4.5.2, alternative methods of inference on the variance components, using suitable transformations to reduce undercoverage in finite samples, are needed.

Model diagnostics are also needed for our small area models in Part II. Also, in this part, the problem of getting efficient and admissible estimators of the autocorrelation $\rho$ in the presence of sampling errors remains to be studied. Similar difficulties are encountered under measurement error models (see Fuller, 1987). It would be
useful to develop suitable modifications of our consistent estimator $\hat{\rho}$, using methods similar to those in Fuller (1987), that lead to more efficient estimators taking values in the admissible range $(-1, 1)$.

Finally, we have only outlined the key steps in the hierarchical Bayes (HB) approach to small area estimation using both time series and cross-sectional data. Practical implementation of HB inference for our problem, especially when $\rho$ is unknown, remains to be done. Also, we need to study the sensitivity of HB inference to choice of priors, especially when $\rho$ is unknown.
Bibliography


Chen J. H. and Shao J. (1993). Iterative weighted least squares estimators. *Annals*
of Statistics, 21, 1071–1092.


In Small Area Statistics (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh
eds) 91-102, John Wiley & Sons, New York.


area estimators. Unpublished manuscript.


85, 163–171.


