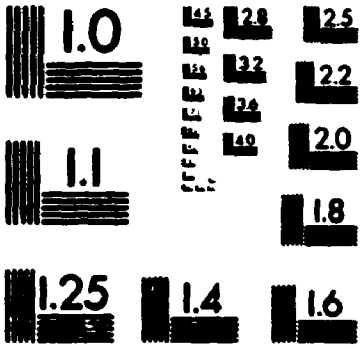
1



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

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

ISBN 0-315-89723-6

Name YU, Mingyu

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

STATISTICS — SUBJECT TERM

0463 — SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences

Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

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

Chairman, Department of Mathematics and Statistics

Thesis Supervisor

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 $\hat{\sigma}_v^2$ 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 β 56

4 Jackknife Estimators **60**
4.1 Review of Jackknife Methodology 61
4.2 Jackknife Estimators and Variance Estimators for σ_v^2 63
4.3 A Jackknife Variance Estimator for $\hat{\beta}_w$ 73
4.4 A Jackknife variance estimator for $\hat{\beta}_{sw}$ 79
4.5 Simulation Results 80
4.5.1 Description of the design 80

4.5.2 Empirical results on estimating σ_v^2 83
4.5.3 Empirical results on estimating β 84

5 Two–Fold Nested Error Regression Models **94**
5.1 Description of the Model 95
5.2 Estimation of σ_v^2 and σ_u^2 97
5.2.1 MINQU, ASR and WLS estimators 98
5.2.2 Approximate MINQUE approach 101
5.2.3 Asymptotic properties of AMINQU estimators 104
5.3 Estimation of Regression Coefficients 108

II Small Area Estimation by Combining Time Series and Cross–Sectional Data **111**

6 EBLUP Estimators **112**
6.1 BLUP Estimator 114
6.2 Two–Stage Estimator: ρ Known 115
6.2.1 Estimation of σ^2 115
6.2.2 Estimation of σ_v^2 116
6.2.3 A two–stage estimator 117
6.3 Estimator of MSE 118
6.3.1 Second order approximation to MSE 118
6.3.2 Second order approximation to estimator of MSE 120
6.4 Two–Stage Estimators: ρ Unknown 120
6.4.1 Estimation of ρ 121
6.4.2 Two–stage or EBLUP estimator 122
6.4.3 Approximation to MSE and Estimation of MSE 122
6.4.4 Two other alternatives 125
6.5 Simulation Study 126

6.5.1 Autocorrelation known 126
6.5.2 Autocorrelation unknown 127

7 Hierarchical Bayes Analysis **134**
7.1 HB Estimator: ρ Known 134
7.1.1 Numerical integration 135
7.1.2 Gibbs sampling 137
7.2 HB Estimation: ρ 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 σ_v^2 86

4.2 Coverage Probability (CP %) and Length of Confidence Interval (LCI) for Estimators of σ_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 σ_v^2 at level $1-\alpha = 90\%$, when Errors follow Double Exponential Distribution 88

4.4 Percentage Efficiency of $\hat{\beta}_w$ Relative to $\hat{\beta}_{sw}$ and $\hat{\beta}_w^{rs}$ 89

4.5 Coverage Probability (CP %) of Confidence Intervals on β 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 mse$[\hat{\theta}_{1T}(\rho_0)]$: $\rho_0 = 0.1, 0.3$, $\rho = 0.2$, $T = 5$ 131
6.6 Relative bias (%) of 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 mse$[\hat{\theta}_{1T}(\hat{\rho}_N)]$: $\rho = 0.2$ 133
6.8 Relative bias (%) of 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:

$$y_{ij} = x_i'\beta + \varepsilon_{ij} \quad \text{with} \quad \varepsilon_{ij} = v_i + e_{ij}, \tag{1.1}$$

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

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, v_i's are independent identically distributed variables with mean 0 and σ_v^2 (denoted as $v_i \overset{iid}{\sim} (0, \sigma_v^2)$), also for given i, $e_{ij} \overset{iid}{\sim} (0, \sigma_i^2)$, and v_i's and e_{ij}'s are mutually independent. Note that σ_i^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 σ_i^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 σ_α^2, σ_i^2 and β 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 σ_v^2 and σ_i^2 are restricted or when $x_i'\beta$ takes special forms, model (1.1) reduces to the following simpler models.

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

Case 2. $\sigma_v^2 = 0$, σ_i^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

1988; Shao, 1989a, 1989b; Shao, 1991; Shao and Rao, 1993).

Case 3. $\sigma_v^2 \neq 0, \sigma_i^2 = \sigma_e^2$. 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

$$y_i = x_i'\beta + e_i, \quad i = 1, 2, \ldots, k \tag{1.2}$$

where β is a $p \times 1$ vector of unknown parameters, $e_i \overset{iid}{\sim} N(0, \sigma^2)$. This is a special case of model (1.1) with $\sigma_v^2 = 0$, $\sigma_i^2 \equiv \sigma^2$ 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 β is estimated through the ordinary least squares (OLS) procedure which gives

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

The error variance σ^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 $\sigma_v^2 = 0$ and σ_i^2 are not all equal, model (1.1) becomes

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

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 β from this model:

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

$$\sigma_i^2 = h(x_i, \beta, \theta), \tag{1.4}$$

see e.g., Carroll (1982), Davidian and Carroll (1987), Carroll and Ruppert (1988). The inference on β involves starting with the initial values of β and θ, and then estimating σ_i^2 using the above function. Then β is updated by the weighted least squares (WLS) estimator of β from model (1.3) using the estimated σ_i^2. The method also specifies an updating method for θ. With the updated β and θ, σ_i^2 is updated from (1.4), and the iteration is continued until estimators of σ_i^2 and β converge. The advantages of this approach are: (1) The estimator of σ_i^2 is consistent, and, (2) The estimator of β is almost as efficient as the WLS estimator with known σ_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 σ_i^2 and the design points x_i since there are occasions when we don't have a priori information about the structure of σ_i^2. In this case a nonparametric approach is needed.

(C) Bayesian approach. A newer approach to the heteroscedastic linear model is the Bayesian approach. This approach assumes a priori distribution on σ_i^2 and the regression coefficients β. Then the inference is based on the posterior distributions of these parameters σ_i^2 and β given the observations $\{y_{ij}\}$ (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

$$\mathcal{Y} = \mathcal{X}\beta + e, \tag{1.5}$$

where

$$\begin{aligned}
\mathcal{Y}_{n\times 1} &= (y_{11}, \ldots, y_{1n_1}, \ldots, y_{k1}, \ldots, y_{1n_k})', \\
\mathcal{X}_{n\times p} &= (x_1 1'_{n_1} : x_2 1'_{n_2} : \cdots : x_k 1'_{n_k})', \\
e_{n\times 1} &= (e_{11}, \ldots, e_{1n_1}, \ldots, e_{k1}, \ldots, e_{1n_k})',
\end{aligned}$$

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

$$V = Var(e) = \text{block diag}(\sigma_1^2 I_{n_1}, \ldots, \sigma_k^2 I_{n_k}). \tag{1.6}$$

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

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

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

$$\hat{\beta}_w = (\mathcal{X}'\hat{V}^{-1}\mathcal{X})^{-1}\mathcal{X}'\hat{V}^{-1}\mathcal{Y}. \tag{1.7}$$

where $\hat{V}$ is obtained by substituting $\hat{\sigma}_i^2$ into σ_i^2 in (1.6).

When $n_i \geq 2$ for all i, an intuitive estimator of σ_i^2 is the sample variance in the i-th group

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

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 σ_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 β 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 σ_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 $p'\sigma^2 = \sum_i^k p_i\sigma_i^2$, 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 β_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

$$A\mathcal{X} = 0 \tag{1.8}$$

and so

$$\mathcal{Y}'A\mathcal{Y} = e'Ae.$$

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

$$tr(AF_i) = p_i, \quad i = 1, \ldots, k, \tag{1.9}$$

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.

When β is known, e is also known. Then a reasonable estimator of $p'\sigma^2$ 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 $e'Ae - e'\Delta e$ 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

$$\mathcal{Y}'A\mathcal{Y} = \sum_{i}^{k} \lambda_i \hat{e}' F_i \hat{e}, \tag{1.10}$$

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

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

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

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

Rao (1973) showed that the MINQUE of σ_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 β 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 σ_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 β from (1.7). Under the assumption of boundedness of σ_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 $k \to \infty$

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 $\mathcal{X}$, 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 $\sigma_v^2 \neq 0$ and $\sigma_i^2 = \sigma_e^2$ for all i.

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

$$y_{ij} - w_i\bar{y}_i = \sum_{l=1}^{p}(1 - w_i)x_{il}\beta_l + \varepsilon_{ij}^*$$

where

$$w_i = 1 - [\sigma_e^2/(\sigma_e^2 + n_i\sigma_v^2)]^{1/2},$$

and

$$\varepsilon_{ij}^* = \varepsilon_{ij} - w_i\bar{\varepsilon}_i, \quad \text{with } \bar{\varepsilon}_i = \textstyle\sum_{i=1}^{n_i} \varepsilon_{ij}/n_i.$$

Since σ_v^2 and σ_e^2 are unknown, their estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ are obtained through the

well-known "fitting-of-constants" method proposed by Henderson (1953) as follows:

$$\hat{\sigma}_e^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2/(n - k),$$
$$\hat{\sigma}_v^2 = \max(0, \tilde{\sigma}_v^2),$$

where

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

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 $\hat{\sigma}_e^2$ and $\tilde{\sigma}_v^2$ are unbiased estimators of σ_e^2 and σ_v^2 respectively, while the truncated estimator $\hat{\sigma}_v^2$ is biased but consistent.

With these estimators of σ_v^2 and σ_e^2, an OLS estimator, $\tilde{\beta}$, of β is obtained from the following model:

$$y_{ij} - \hat{w}_i\bar{y}_i = \sum_{k=1}^{p}(1 - \hat{w}_i)x_{ik}\beta_k + \varepsilon_{ij}^*$$

with

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

This estimator, $\tilde{\beta}$, is identical to the two-step WLS estimator obtained from (1.7) using the above estimators of σ_v^2 and σ_e^2. Fuller and Battese (1973) showed that $\tilde{\beta}$ is unbiased for β under the assumptions of symmetrically distributed errors ε_{ij} with fourth moments and the existence of the expectation of $(\hat{\sigma}_e^2)^{-1}$. Further they showed that $\tilde{\beta}$ has the same asymptotic distribution as the WLS estimator using the true values of σ_v^2 and σ_e^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 σ_v^2 and σ_i^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 σ_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 μ. They also considered the grand mean, the mean of grouped means, and the MLE for estimating μ. 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 $x_i'\beta = \alpha + \beta x_i$. For estimation of variance components σ_v^2 and σ_i^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 α and β 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 σ_v^2 and σ_i^2, calculate the WLS of (α, β). With this estimate of (α, β), compute the first–stage estimator of σ_v^2 and σ_i^2, and using the first–stage estimators of σ_v^2 and σ_i^2, obtain first–stage WLS estimator of (α, β). (2) From the first–stage WLS estimator of (α, β), obtain the second–stage estimator of σ_v^2 and σ_i^2. Then the second–stage WLS estimator of (α, β) 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 (1989) 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 θ_{iT} ($i = 1, \ldots, k$). We assume that y_{iT} is design-unbiased for θ_{iT}, i.e.,

$$y_{iT} = \theta_{iT} + e_{iT}, \tag{1.12}$$

where e_{iT}'s are sampling errors with $E(e_{iT}) = 0$, given θ_{iT}. We assume that a vector of fixed concomitant variables $x_{iT} = (x_{iT1}, \ldots, x_{iTp})'$ related to θ_{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 θ_{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 θ_{iT}:

$$\theta_{iT} = x'_{iT}\beta_T, \tag{1.13}$$

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

$$\tilde{\theta}_{iT}(reg) = x'_{iT}\,\tilde{\beta}_T, \tag{1.14}$$

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

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

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:

$$\theta_{iT} = x'_{iT}\beta_T + v_{iT}, \tag{1.15}$$

where the error terms v_{iT}'s are independent random variables with mean 0 and unknown variance σ^2_{vT}. For sampling errors, we assume that the e_{iT}'s are independent

normal variables with $E(e_{iT}) = 0$ and $V(e_{iT}) = \sigma^2_{iT}$, where σ^2_{iT} is known. The combined model is given by

$$y_{iT} = x'_{iT}\beta_T + v_{iT} + e_{iT}. \tag{1.16}$$

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

$$\tilde{\theta}^H_{iT} = \frac{\sigma^2_{vT}}{\sigma^2_{vT} + \sigma^2_{iT}}\, y_{iT} + \frac{\sigma^2_{iT}}{\sigma^2_{vT} + \sigma^2_{iT}}\, x'_{iT}\hat{\beta}_T(\sigma^2_{vT}). \tag{1.17}$$

where $\hat{\beta}_T(\sigma^2_{vT})$ is the weighted least squares estimator of β_T under the combined model (1.16). In the second step, a suitable estimator $\hat{\sigma}^2_{vT}$ is substituted for σ^2_{vT} 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 $\hat{\theta}_{iT}(reg) = x'_{iT}\hat{\beta}_T(\hat{\sigma}^2_{vT})$:

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

where $y_T = (y_{1T}, \ldots, y_{kT})'$, and $w_{iT} = \hat{\sigma}^2_{vT}/(\hat{\sigma}^2_{vT} + \sigma^2_{iT})$. This result holds without assuming normal distributions for v_{iT} and e_{iT}.

Note that the ratio $\hat{\sigma}^2_{vT}/\sigma^2_{iT}$ 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 σ^2_{vT}:

$$\begin{aligned} \hat{\sigma}^2_{vT} &= \max(\tilde{\sigma}^2_{vT}, 0), \\ \tilde{\sigma}^2_{vT} &= (k-p)^{-1}\left[\sum_{i=1}^{k} \hat{a}^2_{iT} - \sum_{i=1}^{k} \sigma^2_{iT}(1 - x'_{iT}(X'_T X_T)^{-1} x_{iT})\right], \end{aligned} \tag{1.19}$$

where $\hat{a}_{iT} = y_{iT} - x'_{iT}(X'_T X_T)^{-1} X'_T y_T$ 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).

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

$$MSE(\hat{\theta}^H_{iT}) \approx g_{1i}(\sigma^2_{vT}) + g_{2i}(\sigma^2_{vT}) + g_{3i}(\sigma^2_{vT}),$$

where

$$g_{1i}(\sigma^2_{vT}) = \sigma^2_{vT}\sigma^2_{iT}(\sigma^2_{vT} + \sigma^2_{iT})^{-1},$$
$$g_{2i}(\sigma^2_{vT}) = \sigma^4_{iT}(\sigma^2_{vT} + \sigma^2_{iT})^{-2}x'_{iT}(X'_T V^{-1} X_T)^{-1}x_{iT},$$

and

$$g_{3i}(\sigma^2_{vT}) = \sigma^4_{iT}(\sigma^2_{vT} + \sigma^2_{iT})^{-3}[2k^{-1}[\sigma^4_{vT} + 2\sigma^2_{vT}\sum_{i=1}^{k}\sigma^2_{iT}/k + \sum_{i=1}^{k}\sigma^4_{iT}/k],$$

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

$$\mathrm{mse}(\hat{\theta}^H_{iT}) = g_{1i}(\hat{\sigma}^2_{iT}) + g_{2i}(\hat{\sigma}^2_{iT}) + 2g_{3i}(\hat{\sigma}^2_{iT}), \tag{1.20}$$

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 $\hat{\theta}^H_{iT}$ and $\mathrm{mse}(\hat{\theta}^H_{iT})$ is robust to nonnormality of the random effects.

Empirical Bayes Estimation

In the empirical Bayes (EB) approach , y_{iT} is assumed to be normal distributed with mean θ_{iT} and variance σ^2_{iT} conditional on θ_{iT}, and θ_{iT} is assumed to be normal distributed with mean $x'_{iT}\beta_T$ and variance σ^2_{vT}. Hence, the posterior distribution of θ_{iT} given y_{iT}, β_T and σ^2_{vT} is normal with mean θ^B_{iT} and variance $g_{1T}(\sigma^2_v)$, where

$$\theta^B_{iT} = E(\theta_{iT} \mid y_{iT}, \beta_T, \sigma^2_{vT}) = \frac{\sigma^2_{vT}}{\sigma^2_{vT} + \sigma^2_{iT}}\, y_{iT} + \frac{\sigma^2_{iT}}{\sigma^2_{vT} + \sigma^2_{iT}}\, x^t_i\beta_T. \tag{1.21}$$

Under quadratic loss, θ^B_{iT} is the Bayes estimator of θ_{iT}. Since σ^2_{vT} and β_T are unknown, estimators $\hat{\sigma}^2_{vT}$ and $\hat{\beta}_T$ are obtained using ML, REML or method of

moments as in the two-step estimation. The estimated posterior distribution is $N(\hat{\theta}_{iT}^{EB}, g_{1i}(\hat{\sigma}_{vT}^2))$, where $\hat{\theta}_{iT}^{EB}$ is identical to $\hat{\theta}_{iT}^{H}$. A naive EB approach uses $\hat{\theta}_{EB}$ as the estimator of θ_{iT} and measure its uncertainty by the estimated posterior variance

$$Var(\theta_{iT} \mid y_{iT}, \hat{\beta}_T, \hat{\sigma}_{vT}^2) = g_{1i}(\hat{\sigma}_{vT}^2).$$

Although this EB estimator $\hat{\theta}_{iT}^{EB}$ is approximately equal to the true posterior mean $E(\theta_{iT} \mid y_T)$, $g_{1i}(\hat{\sigma}_{vT}^2)$ underestimates the true posterior variance $Var(\theta_{iT} \mid y_T)$ since the variation due to estimating σ_{vT}^2 and β_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 θ_{iT} as those in the EB approach, i.e., $y_{iT} \mid v_{iT} \overset{iid}{\sim} N(0, \sigma_{iT}^2)$, and $\theta_{iT} \mid \beta_T, \sigma_{vT}^2 \overset{iid}{\sim} N(x_{iT}'\beta_T, \sigma_{vT}^2)$. Further we assume prior distributions for β_T and σ_{vT}^2. The inference are then based on the posterior distribution of θ_{iT} given observations y_T.

By assuming that the prior distribution of β_T is a uniform over R^p, we can show that the posterior distribution of θ_{iT} given y_T and σ_{vT}^2 is normal distribution with mean equal to the BLUP $\tilde{\theta}_{iT}^{H}$ and variance equal to $g_{1i}(\sigma_{vT}^2) + g_{2i}(\sigma_{vT}^2)$. Hence when σ_{vT}^2 is assumed to be known, the HB and the BLUP approach lead to identical inferences.

To take the uncertainty about σ_{vT}^2 into account, we need to calculate the posterior distribution of θ_{iT} given the observations y_T, by assuming a suitable prior on σ_v^2. The HB estimator of θ_{iT} is then given by

$$\hat{\theta}_{iT}^{HB} = E(\theta_{iT} \mid y_T) \equiv E_{\sigma_{vT}^2}(\tilde{\theta}_{iT}^{H}). \tag{1.22}$$

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

$$Var(\theta_{iT} \mid y_T) = E_{\sigma_{vT}^2}[g_{1i}(\sigma_{vT}^2) + g_{2i}(\sigma_{vT}^2)] + V_{\sigma_{vT}^2}(\tilde{\theta}_{iT}^{H}), \tag{1.23}$$

where $E_{\sigma_v^2}$ and $V_{\sigma_v^2}$ respectively denote the expectation and variance with respect to the posterior distribution of σ_{vT}^2 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

$$y_{it} = \theta_{it} + e_{it}, \quad i = 1, \ldots, k; \quad t = 1, \ldots, T, \tag{1.24}$$

where the e_{it}'s are sampling errors with $E(e_{it}) = 0$, given the small area means θ_{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 θ_{it}:

$$\text{(I)} \qquad \theta_{it} = x'_{it}\beta + v_i + \varepsilon_{it},$$

where $x_{it} = (x_{it1}, \ldots, x_{itp})'$ 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 ε_{it}'s are independent normal variables with mean 0 and variance σ^2, abbreviated $\varepsilon_{it} \overset{iid}{\sim} (0, \sigma^2)$.

$$\text{(II)} \qquad \theta_{it} = x'_{it}\beta + v_i + \varepsilon_{it},$$

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

$$\text{(III)} \qquad \theta_{it} = x'_{it}\beta + v_i + u_t + \varepsilon_{it},$$

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

$$\begin{aligned} &\text{(IV)} \qquad \theta_{it} = x'_{it}\beta + v_i + u_{it} \\ &\text{and} \qquad u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1, \end{aligned} \tag{1.25}$$

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

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

$$\theta_{it} = \rho\theta_{i,t-1} + (x_{it} - \rho x_{i,t-1})'\beta + (1-\rho)v_i + \varepsilon_{it}. \tag{1.25a}$$

The alternative form (1.25a) relates θ_{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 β (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 θ_{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

$$\begin{aligned} y_{it} &= x_{it}'\beta + v_i + w_{it} \\ w_{it} &= \tilde{\rho} w_{i,t-1} + \tilde{\varepsilon}_{it}, \quad |\tilde{\rho}| < 1, \end{aligned} \tag{1.26}$$

where $\tilde{\varepsilon}_{it} \overset{iid}{\sim} N(0, \tilde{\sigma}^2)$. They obtained a two-stage estimator of small area mean θ_{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.

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.,

$$\begin{aligned} y_{it} &= \theta_{it} + e_{it} = x'_{it}\beta + v_i + u_{it} + e_{it} \\ u_{it} &= \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1. \end{aligned} \tag{1.27}$$

Following Fay and Herriot (1979), we assume that the sampling errors e_{it} are normally distributed with zero mean and block diagonal covariance matrix Σ with known blocks, Σ_i, where Σ_i is a $T \times T$ matrix, i.e., independent of sample errors across small areas. The covariance matrices Σ_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 Σ_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, θ_{iT}, 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

$$\begin{aligned} \mathcal{Y}_{n\times 1} &= (y_{11}, \ldots, y_{1n_1}, \ldots, y_{k1}, \ldots, y_{1n_k})', \\ \mathcal{X}_{n\times p} &= (x_1 1'_{n_1} : x_2 1'_{n_2} : \cdots : x_k 1'_{n_k})', \end{aligned}$$

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

where 1_t is a t-vector with all entries equal to one. A matrix form of model (1.1) is then given by

$$\mathcal{Y} = \mathcal{X}\beta + \mathcal{E}. \tag{1.28}$$

The dispersion matrix of the error $\mathcal{E}$ is

$$V = Var(\mathcal{E}) = \text{block diag}(\sigma_1^2 I_{n_1} + \sigma_v^2 1_{n_1} 1'_{n_1}, \ldots, \sigma_k^2 I_{n_k} + \sigma_v^2 1_{n_k} 1'_{n_k}).$$

Without loss of generality, we assume $\mathcal{X}$ is of full column rank (so is the matrix X defined below). Th primary interest is the estimation of β, and σ_v^2.

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

$$\tilde{\beta}_w = (\mathcal{X}'V^{-1}\mathcal{X})^{-1}\mathcal{X}'V^{-1}\mathcal{Y}. \tag{1.29}$$

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

$$\tilde{\beta}_w = (X'W^{-1}X)^{-1}X'W^{-1}y, \tag{1.30}$$

where

$$X = (x_1 : \cdots : x_k)', \quad W = \text{diag}_i(w_i), \quad w_i = \sigma_v^2 + n_i^{-1}\sigma_i^2,$$
$$y = (\bar{y}_1, \ldots, \bar{y}_k)', \quad \bar{y}_i = \sum_{i=1}^{n_i} y_{ij}/n_i,$$

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

Since σ_v^2 and σ_i^2 are usually unknown, a two-step WLS estimator is used in practice:

$$\hat{\beta}_w = (\mathcal{X}'\hat{V}^{-1}\mathcal{X})^{-1}\mathcal{X}'\hat{V}^{-1}\mathcal{Y} = (X'\hat{W}^{-1}X)^{-1}X'\hat{W}^{-1}y, \tag{1.31}$$

where $\hat{V}$ and $\hat{W}$ are estimators of V and W obtained by substituting suitable estimators into V and W for σ_v^2 and σ_i^2.

For estimating σ_v^2 and σ_i^2, we introduce in Section 2.1 a weighted least squares (WLS) procedure, viz., given a priori values for σ_v^2 and σ_i^2, the WLS estimator of

β using these prior values is calculated. Then σ_v^2 and σ_i^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 σ_v^2 and σ_i^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 σ_v^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 σ_v^2 is then constructed.

Using the AMINQUE estimators of σ_v^2 and σ_i^2, we construct a two–stage WLS estimator of β 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 β 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 σ_v^2 and the within group sample variances as estimator of σ_i^2's. The resulting WLS estimator of β 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-

sider the delete–group jackknife estimators of variance component σ_v^2 in Section 4.2. Delete–group jackknife variance estimators of σ_v^2 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 β, and prove its consistency. The jackknife variance estimator for the alternate WLS estimator of β 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, jacknife estimators of σ_v^2, variance estimators for σ_v^2, 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 β 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, θ_{iT}. We first derive the BLUP estimator of θ_{iT} for

model (1.27) by using general results in Henderson (1950). Then the estimators of variance components σ_v^2 and σ^2 are obtained using moment estimators assuming ρ is known. With ρ known, a two-stage or EBLUP estimator of θ_{iT} is obtained by substituting the estimator of σ_v^2 and σ^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 ρ unknown is discussed in Section 6.4. We present a method for estimating ρ, and construct the EBLUP estimator of θ_{iT} using the estimator of ρ and corresponding estimators of σ_v^2 and σ^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 ρ doesn't perform well in small samples. Therefore we suggest two alternative approaches in Section 6.4.2, and obtain EBLUP estimators of θ_{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 ρ known and ρ unknown. We propose a multistage HB model on the parameter $\theta = (\theta_{11}, \ldots, \theta_{1T}, \ldots, \theta_{k1}, \ldots, \theta_{kT})'$ and other model parameters in each case. Two methods are used to obtain the posterior expectation and posterior covariance matrix of our parameter θ. 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 σ_v^2 is considered for the general nested error regression model (1.1). Estimators of error variances σ_i^2 are also obtained. The estimators of σ_v^2 and σ_i^2 are needed for inference on β, but inference on σ_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 σ_v^2 and σ_i^2 are derived, by dropping lower order terms in MINQU estimation equations with a priori values. The AMINQU estimators of σ_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 σ_v^2 is also suggested.

2.1 WLS Estimators

The weighted least squares (WLS) procedure for estimating the variance components σ_v^2 and σ_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 σ_{v0}^2 and σ_{i0}^2, a WLS estimator of β 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 σ_v^2 and σ_i^2 are obtained by equating these quadratic terms to their expectations. The details are given below.

Given a priori values σ_{v0}^2, σ_{i0}^2 for σ_v^2 and σ_i^2 respectively, we can construct the weights

$$w_{i0} = \sigma_{v0}^2 + \sigma_{i0}^2/n_i.$$

Denote

$$W_0 = \text{diag}_i(w_{i0}),$$

and

$$V_0 = \text{block diag}(\sigma_{10}^2 I_{n_1} + \sigma_{v0}^2 1_{n_1} 1_{n_1}', \ldots, \sigma_{k0}^2 I_{n_k} + \sigma_{k0}^2 1_{n_k} 1_{n_k}').$$

Then we can find $\hat{\beta}_0$ of β 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

$$R_0^2 = (\mathcal{Y} - \mathcal{X}'\hat{\beta}_0)'V_0^{-1}(\mathcal{Y} - \mathcal{X}'\hat{\beta}_0).$$

After simplification, it can be shown that

$$R_0^2 = \sum_{i=1}^{k} \sigma_{i0}^{-2}\left[\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2\right] + \left[\sum_{i=1}^{k} w_{i0}^{-1}(\bar{y}_i - x_i'\hat{\beta}_0)^2\right],$$

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

$$\hat{\sigma}_i^2 = s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 \tag{2.1}$$

$$\hat{\sigma}_v^2 \sum_{i=1}^{k} w_{i0}^{-1} A_i + \sum_{i=1}^{k} n_i^{-1} w_{i0}^{-1} \hat{\sigma}_i^2 A_i = \sum_{i=1}^{k} w_{i0}^{-1}(\bar{y}_i - x_i'\hat{\beta}_0)^2, \tag{2.2}$$

where

$$A_i = 1 - w_{i0}^{-1} x_i' \Big(\sum_{i=1}^{k} w_{i0}^{-1} x_i x_i' \Big)^{-1} x_i.$$

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

ANOVA estimation. The analysis of variance (ANOVA) estimator can be obtained from the WLS procedure with $\sigma_{v0}^2 = 0$ and $\sigma_{i0}^2 = 1$ for all i. Hence $\hat{\sigma}_i^2 = s_i^2$ as in (2.1), and $\hat{\sigma}_v^2$ can be solved from

$$\hat{\sigma}_v^2 \sum_{i=1}^{k} n_i A_i + \sum_{i=1}^{k} \hat{\sigma}_i^2 A_i = \sum_{i=1}^{k} n_i (\bar{y}_i - x_i' \hat{\beta}_0)^2, \tag{2.3}$$

where

$$A_i = 1 - n_i x_i' \Big(\sum_{i=1}^{k} n_i^{-1} x_i x_i' \Big)^{-1} x_i.$$

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 $\sum_{i=1}^{k} n_i x_i x_i' = n$, $\hat{\sigma}_v^2$ is given from (2.3) as

$$\hat{\sigma}_v^2 (n - \sum_{i=1}^{k} n_i^2/n) + \sum_{i=1}^{k} \hat{\sigma}_i^2 (1 - n_i/n) = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2,$$

where $\hat{\sigma}_i^2 = s_i^2$.

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 σ_i^2 is given by $\hat{\sigma}_i^2 = s_i^2$. And the estimator of σ_v^2, from (2.2), is given by

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

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 β 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 σ_v^2 and σ_i^2 can be obtained from (2.1) and (2.2) with $\sigma_{v0}^2 = 1$ 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 $U_i' = (0 : I_{n_i} : 0)_{n_i \times n}$, ξ_i is an n_i–vector with mean 0 and dispersion matrix $\sigma_i^2 I_{n_i}$, U_v = block diag$_i(1_{n_i})_{n\times k}$, and ξ_v is a k–vector with mean 0 and dispersion matrix $\sigma_v^2 I_k$. For a set of a priori values σ_{v0}^2, σ_{10}^2, ..., σ_{k0}^2, denote

$$V_0 = \sigma_{10}^2 F_1 + \cdots + \sigma_{k0}^2 F_k + \sigma_{v0}^2 F_v,$$

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

Parameters of interest are of the form $p'\theta = p_1\sigma_1^2 + \cdots + p_k\sigma_k^2 + p_v\sigma_v^2$ for specified $p' = (p_1, \ldots, p_k, p_v)$ where $\theta' = (\sigma_1^2, \ldots, \sigma_k^2, \sigma_v^2)$. 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 $tr(AV_0AV_0)$ subject to the condition of invariance, $A\mathcal{X} = 0$, and the conditions of unbiasedness, $tr(AF_i) = p_i$, $i = 1, \ldots, k$, 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. R. Rao and Kleffe (1988) (see p.96, Chapter 5) extended it to estimating $\theta = (\sigma_1^2, \ldots, \sigma_k^2, \sigma_v^2)'$ jointly. The estimator of θ by MINQUE with a priori values is the solution to the following system of linear equations:

$$H_{UI}\hat{\theta} = q_{UI}, \tag{2.4}$$

where

$$H_{UI} = \begin{pmatrix} (tr(MV_0M)^+F_i(MV_0M)^+F_j)_{k\times k} & \mu \\ \mu' & tr(MV_0M)^+F_v(MV_0M)^+F_v \end{pmatrix},$$

$$q_{UI} = (Q_1, \ldots, Q_k, Q_v)',$$

with

$$M = I_n - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}',$$

$$\mu = \left(tr(MV_0M)^+F_1(MV_0M)^+F_v, \ldots, tr(MV_0M)^+F_k(MV_0M)^+F_v\right)'_{k\times 1},$$

$$\begin{aligned} Q_v &= \mathcal{Y}'(MV_0M)^+F_v(MV_0M)^+\mathcal{Y}, \\ Q_i &= \mathcal{Y}'(MV_0M)^+F_i(MV_0M)^+\mathcal{Y}, \end{aligned}$$

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)

$$(MV_0M)^+ = V_0^{-1} - V_0^{-1}\mathcal{X}(\mathcal{X}'V_0^{-1}\mathcal{X})^{-1}\mathcal{X}'V_0^{-1}, \tag{2.5}$$

in the right hand side of (2.4), Q_v and Q_i's can be simplified as

$$Q_i = \frac{1}{\sigma_{i0}^4}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + n_i^{-1}w_{i0}^{-2}(\bar{y}_i - \hat{\beta}_0'x_i)^2, \tag{2.6}$$

$$Q_v = \sum_{i=1}^{k} w_{i0}^{-2}(\bar{y}_i - \hat{\beta}_0'x_i)^2, \tag{2.7}$$

where

$$w_{i0} = \sigma_{v0}^2 + n_i^{-1}\sigma_{i0}^2,$$

and

$$\hat{\beta}_0 = (X'W_0^{-1}X)^{-1}X'W_0^{-1}y \tag{2.8}$$

$$W_0 = \mathrm{diag}_i(w_{i0}). \tag{2.9}$$

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 θ 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,

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 β using weights constructed from these variance component estimators. Also, this procedure may lead to negative estimators for σ_v^2 and σ_i^2 (especially for the former parameter). As a result, these estimators cannot be used in constructing a WLS estimator of β without some arbitrary modification to guarantee positive weights.

2.3 ASR Estimator

Let $\sigma_{v0}^2, \sigma_{10}^2, \ldots, \sigma_{k0}^2$ be a set of a priori values of $\sigma_v^2, \sigma_1^2, \ldots, \sigma_k^2$ 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 σ_v^2 and σ_i^2 as

$$\hat{\sigma}_v^2 = (\sigma_{v0}^4/k)Q_v,$$
$$\hat{\sigma}_i^2 = (\sigma_{i0}^4/n_i)Q_i,$$

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 β 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 (AMINQU) 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^*_{UI} in the sense that

$$H_{UI} = H^*_{UI} + O(k^{-1}),$$

where

$$H^*_{UI} = \begin{pmatrix} \left(tr(V_0^{-1}F_iV_0^{-1}F_j)\right)_{k\times k} & col_i\left(tr(V_0^{-1}F_iV_0^{-1}F_v)\right) \\ col_i'\left(tr(V_0^{-1}F_iV_0^{-1}F_v)\right) & tr(V_0^{-1}F_vV_0^{-1}F_v) \end{pmatrix}.$$

where $col_i(a_i)$ denotes a k-vector with a_i as i-th element, $col_i'(a_i) = (col_i(a_i))'$. Evaluating the traces in the above matrix, we obtain that

$$H^*_{UI} = \begin{pmatrix} \mathrm{diag}_i(n_i^{-2}w_{i0}^{-2}d_i^{-1})_{k\times k} & col_i(n_i^{-1}w_{i0}^{-2}) \\ col_i'(n_i^{-1}w_{i0}^{-2}) & \sum_{i=1}^k w_{i0}^{-2} \end{pmatrix},$$

where $w_{i0} = \sigma_{v0}^2 + \sigma_{i0}^2/n_i$ and

$$d_i = \frac{1}{1 + (n_i - 1)(1 + \rho_{i0}n_i)^2}. \tag{2.10}$$

with

$$\rho_{i0} = \sigma_{v0}^2/\sigma_{i0}^2.$$

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

$$H^*_{UI}\tilde{\theta} = q_{UI}, \tag{2.11}$$

with

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

The matrix H^*_{UI} 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 $H = A - B'CB$ 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 $A = \mathrm{diag}_i(n_i^{-2}w_{i0}^{-2}d_i^{-1})$, $B = col_i(n_i^{-1}w_{i0}^{-2})$ and $C = \sum_{i=1}^{k} w_{i0}^{-2}$. Then we have

$$(H_{UI}^*)^{-1} = \begin{pmatrix} \mathrm{diag}_i(n_i^2 w_{i0}^2 d_i) + \nu^{-1} col_i(n_i d_i) col_i'(n_i d_i) & -\nu^{-1} col_i(n_i d_i) \\ -\nu^{-1} col_i'(n_i d_i) & \nu^{-1} \end{pmatrix},$$

where

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

The solution to linear system (2.11) is

$$\tilde{\theta} = (H_{UI}^*)^{-1} q_{UI}\,.$$

Hence, our untruncated AMINQU estimators of σ_v^2 and σ_i^2 are obtained as

$$\tilde{\sigma}_v^2 = \frac{1}{\sum_i c_i}(Q_v - \sum_{i=1}^{k} n_i d_i Q_i)\,, \tag{2.13}$$

$$\tilde{\sigma}_i^2 = n_i^2 w_{i0}^2 d_i Q_i - n_i d_i \tilde{\sigma}_v^2\,. \tag{2.14}$$

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

$$\hat{\sigma}_v^2 = \max(0, \tilde{\sigma}_v^2)\,. \tag{2.15}$$

An estimator of σ_i^2 is obtained correspondingly as

$$\hat{\sigma}_i^2 = n_i^2 w_{i0}^2 d_i Q_i - n_i d_i \hat{\sigma}_v^2\,. \tag{2.16}$$

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 σ_i^2, e.g., the within group sample variance. But it can be shown that there does not exist a non-negative unbiased quadratic estimator for σ_v^2 (La Motte, 1973). Kleffe and Rao (1986) have shown that even asymptotically consistent nonnegative quadratic estimators for σ_v^2 do not exist.

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

$$\hat{w}_i = \hat{\sigma}_v^2 + n_i^{-1}\hat{\sigma}_i^2 = (1 - d_i)\hat{\sigma}_v^2 + n_i w_{i0}^{-2} d_i Q_i > 0\,. \tag{2.17}$$

The weights $\hat{w}_i$ can be applied directly to construct $\hat{W}$ in (1.31) for the WLS estimator of β. 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 σ_i^2 are equal, the corresponding F_i should be pooled together before calculating pooled Q_i. For the purpose of estimating regression coefficient β, 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 σ_i^2.

Remark 5. For the special case $\sigma_v^2 = 0$, our model reduced to model (1.2). Since we do not have the parameter σ_v^2, we should exclude it while applying our AMINQUE procedure. Let $\sigma_{i0}^2 = 1$ for all i. From (2.4), (2.5)–(2.11), the AMINQUE estimator of σ_i^2 can be obtained as

$$\hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij} - x_i'\hat{\beta}_0)^2,$$

which is identical to the estimator of σ_i^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, σ_v^2, σ_1^2, ..., σ_k^2, unlike the MINQUE, while the calculation of MINQU estimators is complicated for general design matrices, since it involves inverting $\mathcal{X}'V_0^{-1}\mathcal{X}$. 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 σ_i^2 unless σ_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 $\tilde{\sigma}_v^2$ and $\hat{\sigma}_v^2$ 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 σ_v^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 σ_v^2, we impose the following conditions on model (1.1):

(A) There are positive constants $\sigma_0^2, \sigma_\infty^2$, c_∞, and a positive integer n_∞ such that $\sigma_0^2 \le \sigma_i^2 \le \sigma_\infty^2$, $2 \le n_i \le n_\infty$ and $\|x_i\| \le c_\infty$ 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

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

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

$$E\,|v_i|^{4+2\delta} \le r_1 \quad \text{and} \quad E|e_{ij}|^{2+\delta} \le r_1 \quad \text{for all } i,\ j \tag{2.18}$$

for some positive constant δ.

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

$$\lambda_0 I_p \le k^{-1}X'X \le \lambda_\infty I_p \tag{2.19}$$

for all k and some positive constants λ_0, λ_∞.

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

$$E\frac{\bar{\varepsilon}_i}{[r_2 + r_3\sum_{j=1}^{n_i}(e_{ij} - \bar{e}_i)^2 + r_4\bar{e}_i^2]^s} = 0, \tag{2.20}$$

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

$$\bar{\varepsilon}_i = v_i + \bar{e}_{i\cdot}\,, \quad \bar{e}_{i\cdot} = \sum_{j=1}^{n_i} e_{ij}/n_i\,.$$

The property (2.20) is used in Chapter 3.

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

$$Var(\bar{\varepsilon}_i) = \sigma_v^2 + \sigma_i^2/n_i \le C$$

for some positive constant C. From (2.18), for the δ given in condition (C) by C_r-inequality (see Lemma 2.2)

$$E|\bar{\varepsilon}_i|^{2+\delta} \leq C_1 \tag{2.21}$$

for some positive constant C_1.

For the prior values σ^2_{v0}, $\sigma^2_{10}, \ldots, \sigma^2_{k0}$, we are able to choose them so that they are bounded away from both 0 and ∞. Therefore, from condition (A), c_i's in (2.12) and d_i's in (2.10) satisfy that

$$0 < c_0^* \leq c_i \leq c_\infty^* < +\infty, \quad 0 < d_0^* \leq d_i < \frac{1}{2} \tag{2.22}$$

for some positive constants c_0^*, c_∞^*, d_0^*.

From (2.19), $W_0 = \mathrm{diag}_i(w_{i0})$ in (2.9) satisfies that

$$\lambda_0^* I_p \leq k^{-1} X' W_0^{-1} X \leq \lambda_\infty^* I_p \tag{2.23}$$

for some positive constants λ_0^*, λ_∞^*.

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

Note that in the following context, all δ's refers to the one in condition (C) except in some lemmas and their proofs where a δ 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 ξ in distribution, denoted as $\xi_k \to_d \xi$, if the cumulative distribution function (cdf) of ξ_k converges to the cdf of ξ 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 $\xi_k = o_p(1)$ or ξ_k is of order $o_p(1)$. Also denote $\xi_k = o_p(\eta_k)$ (ξ_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_δ such that $P(|\xi_k| > M_\delta) < \delta$ for all $k > K_\delta$. It is denoted as $\xi_k = O_p(1)$ or ξ_k is of order $O_p(1)$. And $\xi_k = O_p(\eta_k)$ (ξ_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):

$$O_p(1) \cdot o_p(1) = o_p(1), \quad O_p(1) \cdot O_p(1) = O_p(1). \tag{2.24}$$

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

Lemma 2.2 *(C_r-Inequality) Let ξ and η be two random variables, and r be a positive real number. Then*

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

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 $(1+\epsilon)$-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 $k \to +\infty$, $\xi_k - \mu_k \to_p 0$.

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 $\{a_{ki} : 1 \le i \le k\}$ be a double array which satisfies $\max_i |a_{ki}| \le M/k$ for some positive constant M. If for a positive $\delta < 1$, $k^{-(1+\delta)} \sum_i E|\xi_i|^{1+\delta} \to 0$, then $\sum_i a_{ni}(\xi_i - \mu_i) \to_p 0$.* □

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 $|z_i| \le z_\infty$*

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

Hence

$$k^{-1} \sum_{i=1}^{k} z_i \bar{\varepsilon}_i x_i = o_p(1).$$

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

$$k^{-1/2} \sum_{i=1}^{k} z_i x_{it} \bar{\varepsilon}_i .$$

Since

$$var(k^{-1/2} \textstyle\sum_{i=1}^{k} z_i x_{it} \bar{\varepsilon}_i) = k^{-1} \displaystyle\sum_{i=1}^{k} z_i^2 x_{it}^2 (\sigma_v^2 + n_i^{-1} \sigma_i^2) \le C$$

for some constant $C > 0$, the result follows from Lemma 2.3 with $\epsilon = 1$. □

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

$$\hat{\beta}_0 - \beta = O_p(k^{-1/2}).$$

Proof. From (2.8)

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

where $\varepsilon = (\bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_k)'$. The result follows from (2.23) $\left(\text{viz., } k(X'W_0^{-1}X)^{-1} = O_p(1)\right)$ and Lemma 2.5 since

$$k^{1/2}(\hat{\beta}_0 - \beta) = k(X'W_0^{-1}X)^{-1}k^{-1/2}\sum_i(\sigma_{v0}^2 + n_i^{-1}\sigma_{i0}^2)^{-1}\bar{\varepsilon}_i x_i\,. \qquad \square$$

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

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

Hence

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

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

$$\begin{aligned}&\left|\textstyle\sum_i[\eta_i - E(\eta_i)][(\hat{\beta}_0 - \beta)'x_i]^m\right| \\ &\le\ k\sum_{t_1=1}^{p}\cdots\sum_{t_m=1}^{p}\left|k^{-1}\sum_{i=1}^{k}x_{it_1}\cdots x_{it_m}[\eta_i - E(\eta_i)]\right|\prod_{l=1}^{m}|\hat{\beta}_{0t_l} - \beta_{t_l}|\end{aligned}$$

where $\hat{\beta}_{0t_l}$ is the t_l-th element of $\hat{\beta}_0$. Since $|\hat{\beta}_{0t_l} - \beta_{t_l}| = O_p(k^{-1/2})$ from Lemma 2.6, (2.26) follows from

$$k^{-1}\sum_{i=1}^{k}(\eta_i - E\eta_i)x_{it_1}\cdots x_{it_m} \to_p 0\,,$$

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 $\dot{\sigma}_v^2$ first. Then the asymptotic normality of $\hat{\sigma}_v^2$, the truncation of $\tilde{\sigma}_v^2$ at 0, is just a corollary.

Substituting (2.6), (2.7) into (2.13), we have

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

Further noting that $y_{ij} = x_i'\beta + v_i + e_{ij}$, we can express $\tilde{\sigma}_v^2$ as

$$\tilde{\sigma}_v^2 = \frac{\sum_i c_i Z_i}{\sum_i c_i} - 2R_{1k} + R_{2k} = \tilde{\sigma}_{vp}^2 - 2R_{1k} + R_{2k} \tag{2.28}$$

where

$$\begin{aligned}
Z_i &= v_i^2 + 2v_i\bar{e}_i + \sum\sum_{j\neq l} e_{ij}e_{il}/(n_i^2 - n_i)\,, \\
\tilde{\sigma}_{vp}^2 &= (\textstyle\sum_{i=1}^k c_i)^{-1} \sum_{i=1}^k c_i Z_i\,, && (2.29)\\
R_{1k} &= (\textstyle\sum_{i=1}^k c_i)^{-1} \sum_{i=1}^k c_i \bar{e}_i (\hat{\beta}_0 - \beta)' x_i\,, && (2.30)\\
R_{2k} &= (\textstyle\sum_{i=1}^k c_i)^{-1} \sum_{i=1}^k c_i \left[(\hat{\beta}_0 - \beta)' x_i \right]^2 . && (2.31)
\end{aligned}$$

We have the following theorem on the asymptotic normality of $\tilde{\sigma}_v^2$.

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

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

where

$$var(Z_i) = M_4 - \sigma_v^4 + 4\sigma_v^2\sigma_i^2/n_i + \sigma_i^4/(n_i^2 - n_i)\,, \quad with\ \ M_4 = E(v_1^4)\,.$$

Proof. We show first that

$$\frac{1}{\sqrt{\sum_i c_i^2 var(Z_i)}} \sum_i c_i (Z_i - \sigma_v^2) \to_d N(0,1) \quad \text{as } k \to +\infty\,. \tag{2.32}$$

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

$$\begin{aligned}
&E|Z_i - \sigma_v^2|^{2+\delta} = E\left| v_i^2 + 2v_i\bar{e}_i + \sum\sum_{j\neq l} e_{ij}e_{il}/(n_i^2 - n_i) - \sigma_v^2 \right|^{2+\delta} \\
&\le 4^{2+\delta}\left\{ E|v_i|^{4+2\delta} + 2^{2+\delta}E|v_i|^{2+\delta}E|\bar{e}_i|^{2+\delta} + E|\sum\sum_{j\neq l} e_{ij}e_{il}|^{2+\delta}/(n_i^2 - n_i)^{2+\delta} + \sigma_v^{4+2\delta} \right\} \\
&\le 4^{2+\delta}\left\{ E|v_i|^{4+2\delta} + 2^{2+\delta}E|v_i|^{2+\delta}\textstyle\sum_{j=1}^{n_i} E|e_{ij}|^{2+\delta} + \sum\sum_{j\neq l} E|e_{ij}|^{2+\delta}E|e_{il}|^{2+\delta} + \sigma_v^{4+2\delta} \right\}.
\end{aligned}$$

Then from (2.18),

$$E|Z_i - \sigma_v^2|^{2+\delta} \leq C \tag{2.33}$$

for some positive constant C. From (2.22)

$$E|c_i(Z_i - \sigma_v^2)|^{2+\delta} \leq C_1\,,$$

for some constant C_1 by C_r-Inequality. Moreover,

$$\sqrt{\sum_i c_i^2 var(Z_i)} \;\geq\; c_0\sqrt{\sum_i 4\sigma_v^2\sigma_0^2/n_i} \;\geq\; \sqrt{k}/C_2 \tag{2.34}$$

for some constant $C_2 > 0$. Therefore as $k \to +\infty$

$$\left(\sum_i c_i^2 var(Z_i)\right)^{-(2+\delta)/2} \sum_i E|c_i(Z_i - \sigma_v^2)|^{2+\delta} \leq C_1 k \cdot k^{-(1+\delta/2)} C_2^{2+\delta} \to 0\,.$$

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 $R_{1k} = o_p(k^{-1/2})$ and $R_{2k} = o_p(k^{-1})$, since

$$\sqrt{k}R_{1k} = \frac{k}{\sum_i c_i} \cdot k^{-1/2} \sum_i c_i \bar{\varepsilon}_i (\hat{\beta}_0 - \beta)' x_i = k^{-1/2} o_p(k^{-1/2+1}) = o_p(1)\,,$$

and

$$\sqrt{k}R_{2k} = \frac{k}{\sum_i c_i} \cdot k^{-1/2} \sum_i c_i [(\hat{\beta}_0 - \beta)' x_i]^2 = k^{-1/2} O_p(k^{-2/2+1}) = o_p(k^{-1/2})\,,$$

by applying Lemma 2.7. Also, from (2.34),

$$\sum_{i=1}^k c_i \Big/ \left[\sqrt{\sum_{i=1}^k c_i^2 var(Z_i)}\right] \leq k c_\infty^* \Big/ \sqrt{C_2 k} = C_3 k^{1/2}.$$

for some constant C_3. Thus

$$\frac{\sum_i c_i}{\sqrt{\sum_i c_i^2 var(Z_i)}}(-2R_{1k} + R_{2k}) = o_p(1). \tag{2.35}$$

Finally, from (2.28), (2.32) and (2.35), by Slutsky's theorem (see, e.g., Bickel and Doksum, 1977, A.14.9),

$$\begin{aligned}\frac{\sum_i c_i}{\sqrt{\sum_i c_i^2 var(Z_i)}}(\hat{\sigma}_v^2 - \sigma_v^2) \;&=\; \frac{\sum_i c_i (Z_i - \sigma_v^2)}{\sqrt{\sum_i c_i^2 var(Z_i)}} + \frac{\sum_i c_i}{\sqrt{\sum_i c_i^2 var(Z_i)}}(-2R_{1k} + R_{2k}) \\ &\to_d N(0,1) \quad \text{as} \quad k \to +\infty\,.\end{aligned}$$

This completes the proof. □

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

$$V_a = var(\tilde{\sigma}_{vp}^2) = \frac{\sum_i c_i^2 var(Z_i)}{(\sum_i c_i)^2}$$
$$= \frac{\sum_i c_i^2 \left(M_4 - \sigma_v^4 + 4\sigma_v^2\sigma_i^2/n_i + \sigma_i^4/(n_i^2 - n_i)\right)}{(\sum_i c_i)^2}.$$

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, $\tilde{\sigma}_v^2 \to_p \sigma_v^2$.

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

$$P\{|\tilde{\sigma}_v^2 - \sigma_v^2| > \epsilon\} \to 0$$

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

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

as $k \to +\infty$. Therefore

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

as $k \to +\infty$, or

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

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

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

$$\hat{\sigma}_v^2 = \max(0, \tilde{\sigma}_v^2) \to_p \sigma_v^2,$$

and

$$\frac{\sum_i c_i}{\sqrt{\sum_i c_i^2 var(Z_i)}}(\hat{\sigma}_v^2 - \sigma_v^2) \to_d N(0,1)$$

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

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 $\tilde{\sigma}_v^2$ 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:

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

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

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

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

Proof. Note that

$$\begin{aligned}
k(\tilde{V}_a - V_a) &= \frac{k}{(\sum_i c_i)^2} \sum_i c_i^2 \left[(Z_i - \sigma_v^2)^2 - var(Z_i) \right] \\
&- 2(\tilde{\sigma}_v^2 - \sigma_v^2) \frac{k^2}{(\sum_i c_i)^2} \left[k^{-1} \sum_i c_i^2 (Z_i - \sigma_v^2) \right] + (\tilde{\sigma}_v^2 - \sigma_v^2)^2 \frac{k}{(\sum_i c_i)^2} \sum_i c_i^2 \\
&+ \frac{k}{(\sum_i c_i)^2} \sum_i c_i^2 \left\{ 4\bar{\varepsilon}_i^2 [(\hat{\beta}_0 - \beta)' x_i]^2 + [(\hat{\beta}_0 - \beta)' x_i]^4 \right. \\
&\quad - 4 Z_i \bar{\varepsilon}_i (\hat{\beta}_0 - \beta)' x_i + 2 Z_i [(\hat{\beta}_0 - \beta)' x_i]^2 \\
&\quad \left. + \tilde{\sigma}_v^2 \left[4\bar{\varepsilon}_i (\hat{\beta}_0 - \beta)' x_i - 2((\hat{\beta}_0 - \beta)' x_i)^2 \right] - 4\bar{\varepsilon}_i [(\hat{\beta}_0 - \beta)' x_i]^3 \right\} .
\end{aligned}$$

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:

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

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

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

Note that $\tilde{V}_a$ and $\hat{V}_a$ are the same when $\tilde{\sigma}_v^2 = \hat{\sigma}_v^2$, but when the true value of σ_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,

$$\tilde{V}_a^{-1/2}(\tilde{\sigma}_v^2 - \sigma_v^2) \to_d N(0,1). \tag{2.38}$$

Corollary 2.5 *Under the assumptions of Corollary* 2.3, *we have*

$$\hat{V}_a^{-1/2}(\hat{\sigma}_v^2 - \sigma_v^2) \to_d N(0,1). \tag{2.39}$$

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

$$(\tilde{\sigma}_v^2 - t_{1-\alpha/2,k-1}\tilde{V}_a^{1/2},\ \tilde{\sigma}_v^2 + t_{1-\alpha/2,k-1}\tilde{V}_a^{1/2}) \tag{2.40}$$

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 σ_v^2 and σ_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

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

From (1.31), the regression vector β 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.

3.1 Notations and Some Preliminary Results

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

$$\begin{aligned}
\hat{w}_i &= \hat{\sigma}_v^2 + n_i^{-1}\hat{\sigma}_i^2 \\
&= (1-d_i)\hat{\sigma}_v^2 + d_i\left[n_i^{-1}(1+\rho_{i0}n_i)^2\textstyle\sum_j(y_{ij}-\bar{y}_i)^2 + (\bar{y}_i - \hat{\beta}_0' x_i)^2\right], \\
w_{1i} &= (1-d_i)\sigma_v^2 + d_i\left[n_i^{-1}(1+\rho_{i0}n_i)^2\textstyle\sum_j(y_{ij}-\bar{y}_i)^2 + (\bar{y}_i - \hat{\beta}_0' x_i)^2\right], \\
u_i &= (1-d_i)\sigma_v^2 + d_i\left[n_i^{-1}(1+\rho_{i0}n_i)^2\textstyle\sum_j(e_{ij}-\bar{e}_i)^2 + \bar{e}_i^2\right],
\end{aligned}$$

and

$$\tau_{1i} = Eu_i^{-1}, \quad \tau_{2i} = Eu_i^{-1}\bar{\varepsilon}_i^2, \quad \tau_{3i} = Eu_i^{-2}\bar{\varepsilon}_i^2, \tag{3.1}$$

where d_i's are defined in (2.10) and $\rho_{i0} = \sigma_{v0}^2/\sigma_{i0}^2$.

Remark. When $\sigma_v^2 > 0$, from (2.22), we have

$$u_i^{-1} \le 2/\sigma_v^2. \tag{3.2}$$

Hence τ_{1i}, τ_{2i} and τ_{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* $\sigma_v^2 > 0$, *then for any non-random sequence* $\{z_i\}$ *satisfying* $|z_i| \le z_\infty$ *for all* i,

$$k^{-1}\sum_{i=1}^{k} z_i\left[u_i^{-m}\bar{\varepsilon}_i^r - E(u_i^{-m}\bar{\varepsilon}_i^r)\right] \to_p 0$$

for $r = 0,1,2$ *and any nonnegative integer* m.

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

$$E|z_i u_i^{-m}\bar{\varepsilon}_i^r|^{1+\delta/2} \le C_1 E|\bar{\varepsilon}_i^r|^{1+\delta/2} \le C$$

for some positive constants C and C_1. The result follows from Lemma 2.4. □

Lemma 3.2 *Suppose that assumptions* (A), (B) *and* (C) *hold. Assume* $\sigma_v^2 > 0$. *Let* ζ_1, ζ_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

$$k^{-3/2} \sum_i \frac{\eta_i}{[(1-d_i)\hat{\sigma}_v^2 + \zeta_1]^r [(1-d_i)\sigma_v^2 + \zeta_1]^m [(1-d_i)\sigma_v^2 + \zeta_2]^s} \to_p 0,$$

as $k \to +\infty$.

Proof. With $\sigma_v^2 > 0$, from Corollary 2.2, we have $\hat{\sigma}_v^2 \to_p \sigma_v^2$. Therefore

$$P(\hat{\sigma}_v^2 \leq \sigma_v^2/2) \to 0 \quad \text{as } k \to +\infty\,.$$

For any $\delta_1 > 0$,

$$\begin{aligned}
&P\left\{k^{-3/2}\left|\sum_i \frac{\eta_i}{[(1-d_i)\hat{\sigma}_v^2 + \zeta_1]^r [(1-d_i)\sigma_v^2 + \zeta_1]^m [(1-d_i)\sigma_v^2 + \zeta_2]^s}\right| \geq \delta_1\right\} \\
&\leq P\{\hat{\sigma}_v^2 \leq \sigma_v^2/2\} \\
&\quad + P\left\{k^{-3/2} \sum_i \frac{|\eta_i|}{[(1-d_i)\sigma_v^2/2]^r [(1-d_i)\sigma_v^2]^{m+s}} \geq \delta_1,\ \ \hat{\sigma}_v^2 > \sigma_v^2/2\right\} \\
&\leq P\{\hat{\sigma}_v^2 \leq \sigma_v^2/2\} + P\{Ck^{-3/2} \sum_i |\eta_i| \geq \delta_1\} \\
&\leq P\{\hat{\sigma}_v^2 \leq \sigma_v^2/2\} + Ck^{-3/2} \sum_i E|\eta_i|/\delta_1 \quad \text{(by Lemma 2.3)} \\
&\leq P\{\hat{\sigma}_v^2 \leq \sigma_v^2/2\} + C_1 k^{-1/2} \to 0
\end{aligned}$$

as $k \to +\infty$, where C and C_1 are some positive constants. The proof is complete. □

Lemma 3.3 *Suppose that assumptions* (A), (B) *and* (C) *hold. If* $\sigma_v^2 > 0$, *then for any non-random sequence* $\{z_i\}$ *satisfying* $|z_i| \leq z_\infty$

$$k^{-1} \sum_{i=1}^{k} z_i w_{1i}^{-1} \hat{w}_i^{-1} \bar{\varepsilon}_i = o_p(1)\,.$$

Proof. From the definition of $\hat{w}_i$, u_i and w_{1i},

$$\hat{w}_i - w_{1i} = (1-d_i)(\hat{\sigma}_v^2 - \sigma_v^2)\,, \tag{3.3}$$

$$u_i - w_{1i} = d_i\left(2\bar{\varepsilon}_i^2 - (\hat{\beta}_0 - \beta)'x_i\right)(\hat{\beta}_0 - \beta)'x_i\,. \tag{3.4}$$

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

$$\begin{aligned}
&k^{-1}\sum_{i=1}^{k} z_i w_{1i}^{-1}\hat{w}_i^{-1}\bar{\varepsilon}_i \\
&= k^{-1}\sum_{i=1}^{k} z_i w_{1i}^{-2}\bar{\varepsilon}_i - k^{-1}\sum_{i=1}^{k} z_i w_{1i}^{-2}\hat{w}_i^{-1}(\hat{w}_i - w_{1i})\bar{\varepsilon}_i \\
&= k^{-1}\sum_{i=1}^{k} z_i u_i^{-2}\bar{\varepsilon}_i + k^{-1}\sum_{i=1}^{k} z_i(u_i^{-2}w_{1i}^{-1} + u_i^{-1}w_{1i}^{-2})\bar{\varepsilon}_i\left(2\bar{\varepsilon}_i - (\hat{\beta}_0 - \beta)'x_i\right)(\hat{\beta}_0 - \beta)'x_i \\
&\quad - k^{1/2}(\hat{\sigma}_v^2 - \sigma_v^2)k^{-3/2}\sum_{i=1}^{k} z_i(1 - d_i)w_{1i}^{-2}\hat{w}_i^{-1}\bar{\varepsilon}_i\,.
\end{aligned}$$

In the right hand side of the last equality, the first term is $o_p(1)$ from Lemma 3.1 since $Eu_i^{-2}\bar{\varepsilon}_i = 0$ 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. □

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

$$H = \mathrm{diag}_i(z_i\hat{w}_i^{-1})\,,$$
$$G = \mathrm{diag}_i(z_i\tau_{1i})\,,$$

where z_i*'s satisfy* $z_0 \le z_i \le z_\infty < +\infty$. *Then*

$$k^{-1}X'HX - k^{-1}X'GX \to_p 0\,.$$

Proof. The (t,s)-elements of $k^{-1}X'HX$ and $k^{-1}X'GX$ are respectively

$$k^{-1}\sum_{i=1}^{k} z_i\hat{w}_i^{-1}x_{it}x_{is} \quad \text{and} \quad k^{-1}\sum_{i=1}^{k} z_i\tau_{1i}x_{it}x_{is}.$$

Using the identity (3.5) twice, their difference can be decomposed as

$$
\begin{aligned}
& k^{-1}\sum_{i=1}^{k} z_i \hat{w}_i^{-1} x_{it} x_{is} - k^{-1}\sum_{i=1}^{k} z_i \tau_{1i} x_{it} x_{is} \\
&= k^{-1}\sum_{i=1}^{k} z_i x_{it} x_{is} (u_i^{-1} - \tau_{1i}) \\
&\quad + k^{-1}\sum_{i=1}^{k} z_i x_{it} x_{is} u_i^{-1} w_{1i}^{-1} \left(2\bar{\epsilon}_i - (\hat{\beta}_0 - \beta)' x_i\right) (\hat{\beta}_0 - \beta)' x_i \\
&\quad - (\hat{\sigma}_v^2 - \sigma_v^2) k^{-1} \sum_{i=1}^{k} (1 - d_i) z_i x_{it} x_{is} w_{1i}^{-1} \hat{w}_i^{-1}.
\end{aligned}
$$

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. □

3.2 Asymptotic Distribution of $\hat{\beta}_w$

Note that $\hat{w}_i$'s are even functions of ϵ_{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

$$
\begin{aligned}
\Sigma_k = {} & X' D_3 X + 2X' D_2 W_0^{-1} X (X' W_0^{-1} X)^{-1} X' D_0 D_3 X \\
& + 2X' D_3 D_0 X (X' W_0^{-1} X)^{-1} X' W_0^{-1} D_2 X \\
& + 4X' D_3 D_0 X (X' W_0^{-1} X)^{-1} X' W_0^{-1} D W_0^{-1} X (X' W_0^{-1} X)^{-1} X' D_0 D_3 X\,,
\end{aligned}
$$

where

$$
D_0 = \mathrm{diag}_i(d_i)\,, \quad D = \mathrm{diag}_i(\sigma_v^2 + n_i^{-1}\sigma_i^2)\,, \quad D_t = \mathrm{diag}_i(\tau_{ti}) \;\; (t = 1, 2, 3)\,,
$$

and τ_{ti}'s are defined in (3.1).

The following theorem establishes the asymptotic distribution of the WLS estimator $\hat{\beta}_w$.

Theorem 3.1 *Suppose that assumptions* (A), (B)*and* (C) *hold. If* $\sigma_v^2 > 0$, *then*

$$V_k^{-1/2}(\hat{\beta}_w - \beta) \to_d N(0, I_p) \tag{3.6}$$

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

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

Proof. We first show that

$$\Sigma_k^{-1/2}T_k \to_d N(0, I_p), \tag{3.7}$$

where

$$T_k = X'\hat{W}_1^{-1}\varepsilon + 2(X'D_3D_0X)(X'W_0^{-1}X)^{-1}X'W_0^{-1}\varepsilon$$

with

$$\hat{W}_1 = \mathrm{diag}_i(u_i) \quad \text{and} \quad \varepsilon = (\bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_k)'.$$

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

$$\begin{aligned}
Var(T_k) &= E(T_kT_k') \\
&= E(X'\hat{W}_1^{-1}\varepsilon\varepsilon'\hat{W}_1^{-1}X) + 2E\left[X'\hat{W}_1^{-1}\varepsilon\varepsilon'W_0^{-1}X(X'W_0^{-1}X)^{-1}X'D_0D_3X\right] \\
&\quad + 2E\left[X'D_3D_0X(X'W_0^{-1}X)^{-1}X'W_0^{-1}\varepsilon\varepsilon'\hat{W}_1^{-1}X\right] \\
&\quad + 4E\left[X'D_3D_0X(X'W_0^{-1}X)^{-1}X'W_0^{-1}\varepsilon\varepsilon'W_0^{-1}X(X'W_0^{-1}X)^{-1}X'D_0D_3X\right] \\
&= \Sigma_k,
\end{aligned}$$

since

$$\hat{W}_1^{-1}\varepsilon\varepsilon'\hat{W}_1^{-1} = \left(\frac{\bar{\varepsilon}_i\bar{\varepsilon}_t}{u_iu_t}\right)_{(k\times k)},$$

and

$$\varepsilon\varepsilon'\hat{W}_1^{-1} = \left(\frac{\bar{\varepsilon}_i\bar{\varepsilon}_t}{u_t}\right)_{(k\times k)},$$

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

$$d_k^2 = var(\lambda_k'T_k) = (l'\Sigma_k^{-1}l)^{-1}l'\Sigma_k^{-1/2}\Sigma_k\Sigma_k^{-1/2}l = (l'\Sigma_k^{-1}l)^{-1}l'l.$$

Let λ_{ks} and η_{ks} be the s-th element of λ_k and $\eta_k = 2(X'W_0^{-1}X)^{-1}X'D_0D_3X\lambda_k$, respectively. Then

$$\lambda_k'T_k = \sum_{i=1}^{k}\sum_{s=1}^{p}\left[x_{is}\lambda_{ks}u_i^{-1}\bar{\varepsilon}_i + x_{is}\eta_{ks}w_{i0}^{-1}\bar{\varepsilon}_i\right].$$

From assumption (A) and (2.22) (due to choice of σ_{v0}^2, σ_{i0}^2)

$$\|\eta_k\|^2 = 4\lambda_k'X'D_3D_0X(X'W_0^{-1}X)^{-2}X'D_0D_3X\lambda_k \le C_1$$

where C_1 is some positive constant, and

$$E|x_{is}\lambda_{ks}u_i^{-1}\bar{\varepsilon}_i + x_{is}\eta_{ks}w_{i0}^{-1}\bar{\varepsilon}_i|^{2+\delta} \le C_2E\{[(1-d_i)^{-1}\sigma_v^{-2}+1]|\bar{\varepsilon}_i|\}^{2+\delta} \le C_3,$$

for some positive constants C_2 and C_3. Hence

$$d_k^{-(2+\delta)}\sum_i\sum_s E|x_{is}\lambda_{ks}u_i^{-1}\bar{\varepsilon}_i + x_{is}\eta_{ks}w_{i0}^{-1}\bar{\varepsilon}_i|^{2+\delta} \le C_3pkd_k^{-(2+\delta)}.$$

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

$$C_0^*I_p \le k^{-1}\Sigma_k \le C_1^*I_p. \tag{3.8}$$

Hence

$$kd_k^{-(2+\delta)} \le C_0^{*-(1-\delta/2)}k^{-\delta/2} \to 0,$$

and the Linderberg's condition holds. Therefore

$$d_k^{-1}\lambda_k'T_k = l'\Sigma_k^{-1/2}T_k/(l'l)^{1/2} \to_d N(0,1).$$

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 $\sum_{i=1}^k x_{it}\hat{w}_i^{-1}\bar{\varepsilon}_i$. By using (3.5) repeatedly

$$\begin{aligned}
&\sum_{i=1}^{k} x_{it}\hat{w}_i^{-1}\bar{\varepsilon}_i \\
&= \sum_{i=1}^{k} x_{it}u_i\bar{\varepsilon}_i + 2\sum_{i=1}^{k} d_i\tau_{3i}x_{it}(\hat{\beta}_0-\beta)'x_i + 2\sum_{i=1}^{k} d_ix_{it}(u_i^{-2}\bar{\varepsilon}_i^2-\tau_{3i})(\hat{\beta}_0-\beta)'x_i \\
&\quad - \sum_{i=1}^{k} d_iu_i^{-1}w_{1i}^{-1}x_{it}\bar{\varepsilon}_i\left((\hat{\beta}_0-\beta)'x_i\right)^2 - \sum_{i=1}^{k}(1-d_i)x_{it}w_{1i}^{-1}\hat{w}_i^{-1}\bar{\varepsilon}_i(\hat{\sigma}_v^2-\sigma_v^2) \\
&\quad + 2\sum_{i=1}^{k} d_i^2x_{it}u_i^{-2}w_{1i}^{-1}\bar{\varepsilon}_i^2\left((\hat{\beta}_0-\beta)'x_i\right)^2\left(2\bar{\varepsilon}_i-(\hat{\beta}_0-\beta)'x_i\right).
\end{aligned}$$

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 $|u_i^{-2}\bar{\varepsilon}_i^3| \leq |d_i^{-1}u_i^{-1}\bar{\varepsilon}_i|$); 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)

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

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

$$\begin{aligned}&\Sigma_k^{-1/2}(X'D_1X)(\hat{\beta}_w - \beta) \\ &\quad = \Sigma_k^{-1/2}(X'D_1X)(X'\hat{W}^{-1}X)^{-1}\Sigma_k^{1/2}\Sigma_k^{-1/2}X'\hat{W}^{-1}\varepsilon.\end{aligned} \tag{3.11}$$

Note that

$$\begin{aligned}&\Sigma_k^{-1/2}(X'D_1X)(X'\hat{W}^{-1}X)^{-1}\Sigma_k^{1/2} \\ &\quad = \Sigma_k^{-1/2}(X'D_1X)\left[(X'\hat{W}^{-1}X)^{-1} - (X'D_1X)^{-1}\right]\Sigma_k^{1/2} + I_p.\end{aligned}$$

From Lemma 3.4, $k[(X'\hat{W}^{-1}X)^{-1} - (X'D_1X)^{-1}] \to_p 0$. From (3.8) and and the fact $X'D_1X \leq 2\sigma_v^{-2}X'X$,

$$\Sigma_k^{-1/2}(X'D_1X)(X'\hat{W}^{-1}X)^{-1}\Sigma_k^{1/2} \to I_p.$$

Thus from (3.10) and (3.11)

$$\Sigma_k^{-1/2}(X'D_1X)(\hat{\beta}_w - \beta) \to_d N(0, I_p). \tag{3.12}$$

We have shown (3.6) when $(X'D_1X)^{-1}\Sigma_k^{1/2}$ 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. □

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

$$\hat{\theta} = g(\hat{\beta}_w).$$

Then we have the following results about $\hat{\theta}$.

Theorem 3.2 *Assume g is a continuously differentiable function from $R^p \to R^q$ and $\theta = g(\beta)$. Under conditions in Theorem 3.1,*

$$V_\theta^{-1/2}(\hat{\theta} - \theta) \to N(0, I_q),$$

where

$$V_\theta = \Delta g(\beta) V_k [\Delta g(\beta)]',$$

with $\Delta g(\beta)$ denoting the gradient of g at β.

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 $X'D_1X$, and an estimator of Σ_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

$$\begin{aligned}\hat{\Sigma}_k = {} & X'\hat{D}_3X + 2X'\hat{D}_2W_0^{-1}X(X'W_0^{-1}X)^{-1}X'D_0\hat{D}_3X \\ & + 2X'\hat{D}_3D_0X(X'W_0^{-1}X)^{-1}X'W_0^{-1}\hat{D}_2X \\ & + 4X'\hat{D}_3D_0X(X'W_0^{-1}X)^{-1}X'W_0^{-1}\hat{D}W_0^{-1}X(X'W_0^{-1}X)^{-1}X'D_0\hat{D}_3X\,,\end{aligned}$$

and

$$\begin{aligned}\hat{D} &= \mathrm{diag}_i\left((\bar{y}_i - \hat{\beta}_0'x_i)^2\right)\,, \\ \hat{D}_2 &= \mathrm{diag}_i\left(\hat{w}_i^{-1}(\bar{y}_i - \hat{\beta}_0'x_i)^2\right)\,, \\ \hat{D}_3 &= \mathrm{diag}_i\left(\hat{w}_i^{-2}(\bar{y}_i - \hat{\beta}_0'x_i)^2\right)\,.\end{aligned}$$

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

Theorem 3.3 *Under the assumptions in Theorem 3.1, we have*

$$k(\hat{V}_k - V_k) \to_p 0.$$

Proof. We prove first that

$$k^{-1}(\hat{\Sigma}_k - \Sigma_k) \to_p 0. \tag{3.13}$$

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

$$\begin{aligned}
&k^{-1}(X'\hat{D}_3X - X'D_3X) \to_p 0,\\
&k^{-1}(X'W_0^{-1}\hat{D}_2X - X'W_0^{-1}D_2X) \to_p 0,\\
&k^{-1}(X'D_0\hat{D}_3X - X'D_0D_3X) \to_p 0,\\
&k^{-1}(X'W_0^{-1}\hat{D}W_0^{-1}X - X'W_0^{-1}DW_0^{-1}X) \to_p 0.
\end{aligned} \tag{3.14}$$

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

$$\begin{aligned}
&k(\hat{V}_k - V_k)\\
&= k[(X'\hat{W}^{-1}X)^{-1} - (X'D_1X)^{-1}]\,\hat{\Sigma}_k\,[(X'\hat{W}^{-1}X)^{-1} - (X'D_1X)^{-1}]\\
&\quad + k[(X'\hat{W}^{-1}X)^{-1} - (X'D_1X)^{-1}]\,\hat{\Sigma}\,(X'D_1X)^{-1}\\
&\quad + k(X'D_1X)^{-1}\,\hat{\Sigma}_k\,[(X'\hat{W}^{-1}X)^{-1} - (X'D_1X)^{-1}]\\
&\quad + k(X'D_1X)^{-1}(\hat{\Sigma}_k - \Sigma_k)(X'D_1X)^{-1}.
\end{aligned}$$

From Lemma 3.4, $k[(X'\hat{W}^{-1}X)^{-1} - (X'D_1X)^{-1}] = o_p(1)$. From (3.13) and (3.8), $k^{-1}\hat{\Sigma}_k = O_p(1)$. Clearly $k(X'D_1X)^{-1} = O_p(1)$. This completes the proof. □

Similarly for the parameter θ in Section 3.2. The following variance estimator $\hat{V}_\theta$ is a consistent estimator of V_θ.

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

$$k(\hat{V}_\theta - V_\theta) \to_p 0.$$

where

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

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*

$$\hat{V}_k^{-1/2}(\hat{\beta}_w - \beta) \to_d N(0, I_p) \tag{3.15}$$

as $k \to \infty$.

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*

$$\hat{V}_\theta^{-1/2}(\hat{\theta} - \theta) \to_d N(0, I_q)$$

as $k \to \infty$.

Using the result in Corollary 3.1 we can construct a simultaneous confidence region for the regression coefficients in the form

$$\{\beta : \ (\hat{\beta}_w - \beta)'\hat{V}_k^{-1}(\hat{\beta}_w - \beta) \leq C\} \tag{3.16}$$

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 β. 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

$$(\hat{\theta} - t_{1-\alpha/2,k-1}\hat{V}_\theta^{1/2}, \ \hat{\theta} + t_{1-\alpha/2,k-1}\hat{V}_\theta^{1/2}) \tag{3.17}$$

as a $1 - \alpha$ confidence interval of θ 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 β

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 β with any invariant estimator of σ_i^2. One classical estima' .r of σ_i^2 is $s_i^2 =$

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

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 $\sum_{j=1}^{n_i} (y_{ij} - x_i'\hat{\beta})^2 - \hat{\sigma}_v^2$. Another reason for using s_i^2 is to guarantee positive weights for the WLS estimator of β. Since s_i^2 is always positive, $\hat{\sigma}_v^2 + s_i^2/n_i$ 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 σ_v^2 is small or close to 0, $\hat{\sigma}_v^2$ 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 β can be obtained from (1.31) as

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

with

$$\hat{W}_s = \mathrm{diag}_i(\hat{w}_{si}).$$

Denote

$$\begin{aligned} u_{si} &= \sigma_v^2 + n_i^{-1}s_i^2, \\ \tau_{4i} &= E(u_{si}^{-1}), \\ \tau_{5i} &= E(u_{si}^{-2}\bar{\varepsilon}_i^2), \\ D_4 &= \mathrm{diag}_i(\tau_{4i}), \\ D_5 &= \mathrm{diag}_i(\tau_{5i}). \end{aligned}$$

Clearly τ_{4i} and τ_{5i} are finite since $u_{si}^{-1} \leq \sigma_v^{-2}$ given $\sigma_v^2 > 0$.

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

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

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

$$k^{-1}\sum_{i=1}^{k} z_i\left[u_{si}^{-m}\bar{\varepsilon}_i^r - E(u_{si}^{-m}\bar{\varepsilon}_i^r)\right] \to_p 0$$

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 $k \to +\infty$,*

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

(iii)

$$k^{-1}(X'\hat{W}_s^{-1}X - X'D_4X) \to_p 0 \quad as \quad k \to +\infty. \qquad \square$$

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 $\sigma_v^2 > 0$, then*

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

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

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

Proof. Let

$$\hat{W}_{s1} = \mathrm{diag}_i(u_{si}) .$$

Then

$$\hat{W}_{s1}^{-1}\varepsilon = col_i\left(u_{si}^{-1}\bar{\varepsilon}_i\right) ,$$

$$\hat{W}_{s1}^{-1}\varepsilon\varepsilon'\hat{W}_{s1}^{-1} = \left(\frac{\bar{\varepsilon}_i\bar{\varepsilon}_t}{u_{si}u_{st}}\right)_{(k\times k)} .$$

From (2.20)

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

Thus

$$Var(X'\hat{W}_{s1}^{-1}\varepsilon) = E[X'\hat{W}_{s1}^{-1}\varepsilon\varepsilon'\hat{W}_{s1}^{-1}X] = X'D_5X .$$

Along the lines of the proof of (3.7) and (3.9), we can prove similarly by using Lemma 3.5 that

$$(X'D_5X)^{-1/2}X'\hat{W}_{s1}^{-1}\varepsilon \to_d N(0, I_p),$$

$$X'\hat{W}_s^{-1}\varepsilon - X'\hat{W}_{s1}^{-1}\varepsilon = o_p(k^{1/2}).$$

Hence

$$(X'D_5X)^{-1/2}X'\hat{W}_s^{-1}\varepsilon \to_d N(0, I_p) \quad \text{as} \quad k \to +\infty.$$

Then, similar to the proof of (3.12),

$$(X'D_5X)^{-1/2}(X'D_4X)(\hat{\beta}_{sw} - \beta) \to_d N(0, I_p) \quad \text{as} \quad k \to +\infty.$$

Similar argument to the last paragraph of the proof of Theorem 3.1 completes this proof. The details are omitted. □

Similar to the estimator of V_k in section 3.3, we have a consistent estimator of V_{sk} as

$$\hat{V}_{sk} = (X'\hat{W}_s^{-1}X)^{-1}\,(X'\hat{D}_5X)\,(X'\hat{W}_s^{-1}X)^{-1}$$

with

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

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

$$k(\hat{V}_{sk} - V_{sk}) \to_p 0 \quad \textit{as} \quad k \to +\infty.$$

The proof of the theorem is similar to that of Theorem 3.3, and is omitted for simplicity. □

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 σ_v^2 is discussed. Jackknifing the untruncated AMINQUE estimator $\tilde{\sigma}_v^2$ is proposed. A jackknife variance estimator of $\tilde{\sigma}_v^2$ (and $\hat{\sigma}_v^2$) 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 σ_v^2. We can also obtain a jackknife estimator of σ_v^2 by jackknifing $\hat{\sigma}_v^2$, and corresponding variance estimators. They are consistent if we have the prior knowledge $\sigma_v^2 > 0$.

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 β.

In Section 4.4, we outline the jackknife method to construct a consistent estimator of V_{sk}, 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, θ. 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 θ), 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 θ, 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 β, and $\hat{\beta}_{OLS}^{(-i)}$ be the OLS estimator after the observation (y_i, x_i) is removed. Denote $\hat{\theta} = f(\hat{\beta}_{OLS})$ and $\hat{\theta}_{-i} = f(\hat{\beta}_{OLS}^{(-i)})$. Then the pseudo-values are defined as

$$Q_i = k\hat{\theta} - (k-1)\hat{\theta}_{-i}.$$

A jackknife estimator of θ is

$$\hat{\theta}_J = \frac{1}{k}\sum_{i=1}^{k}\hat{\theta}_{-i},$$

and a jackknife variance estimator is defined as

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

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 μ. 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 σ_v^2

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

$$\tilde{\sigma}_v^2 = \tilde{\sigma}_{vp}^2 - 2R_{1k} + R_{2k}\,.$$

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

Let $\tilde{\sigma}_{v(-t)}^2$ denote the estimator of σ_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), $\tilde{\sigma}_{v(-t)}^2$ can be written as

$$\tilde{\sigma}_{v(-t)}^2 = \tilde{\sigma}_{vp(-t)}^2 - 2R_{1k(-t)} + R_{2k(-t)}$$

with

$$\tilde{\sigma}^2_{vp(-t)} = \left(\sum_{i\neq t} c_i\right)^{-1} \sum_{i\neq t} c_i Z_i,$$
$$R_{1k(-t)} = \left(\sum_{i\neq t} c_i\right)^{-1} \sum_{i\neq t} c_i \bar{\varepsilon}_i (\hat{\beta}_{0(-t)} - \beta)' x_i,$$
$$R_{2k(-t)} = \left(\sum_{i\neq t} c_i\right)^{-1} \sum_{i\neq t} c_i \left[(\hat{\beta}_{0(-t)} - \beta)' x_i\right]^2,$$

where

$$\hat{\beta}_{0(-t)} = (X'_{(-t)} W^{-1}_{0(-t)} X_{(-t)})^{-1} X'_{(-t)} W^{-1}_{0(-t)} y_{(-t)}, \tag{4.1}$$

$X_{(-t)}$ is the design matrix without design point x_t, $W_{0(-t)} = \mathrm{diag}(w_{10}, \ldots, w_{(t-1)0}, w_{(t+1)0}, \ldots, w_{k0})$, and $y_{(-t)} = (\bar{y}_1, \ldots, \bar{y}_{t-1}, \bar{y}_{t+1}, \ldots, \bar{y}_k)'$.

The pseudo-values are defined as

$$Q_t = k\tilde{\sigma}^2_v - (k-1)\tilde{\sigma}^2_{v(-t)}.$$

A jackknife estimator of σ^2_v is then given as

$$\bar{Q} = k^{-1} \sum_{t=1}^{k} Q_t = \tilde{\sigma}^2_{v(J)}.$$

And a jackknife estimator of variance of $\tilde{\sigma}^2_{v(J)}$ or $\tilde{\sigma}^2$ is given as

$$\hat{V}_{Jt} = (k(k-1))^{-1} \sum_{t=1}^{k} (Q_t - \bar{Q})^2 = k^{-1}(k-1) \sum_{t=1}^{k} \left[\tilde{\sigma}^2_{v(-t)} - k^{-1} \sum_t \tilde{\sigma}^2_{v(-t)}\right]^2. \tag{4.2}$$

We shall establish the asymptotic normality of $\tilde{\sigma}^2_{v(J)}$ 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*

$$k^{-1} \max_i |\xi_i| \to_p 0 \quad as \quad k \to +\infty.$$

Proof. For any $\delta_1 > 0$,

$$\begin{aligned} P\{k^{-1} \max_i |\xi_i| \geq \delta_1\} &\leq \sum_i P\{k^{-1}|\xi_i| \geq \delta_1\} \\ &\leq \sum_i k^{-(1+\delta)} E|\xi_i|^{1+\delta} / \delta_1^{1+\delta} \\ &\leq Ck^{-\delta} \to 0 \quad \text{as} \quad k \to \infty \end{aligned}$$

where C is some positive constant. □

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

$$\max_t |\xi| = (\max_t |\xi_1|, \ldots, \max_k |\xi_p|)'.$$

Lemma 4.2 *Suppose that assumptions* (A), (B) *and* (C) *hold. Then*

$$\max_t |\hat{\beta}_0 - \hat{\beta}_{0(-t)}| = o_p(k^{-1/2}). \tag{4.3}$$

Proof. From a standard matrix result,

$$\begin{aligned}(X'_{(-t)}W^{-1}_{0(-t)}X_{(-t)})^{-1} &= (X'W_0^{-1}X - w_{t0}^{-1}x_t x'_t)^{-1} \\ &= (X'W_0^{-1}X)^{-1} + w_{t0}^{-1}h_{0t}^{-1}(X'W_0^{-1}X)^{-1}x_t x'_t(X'W_0^{-1}X)^{-1}\end{aligned} \tag{4.4}$$

with

$$h_{0t} = 1 - w_{t0}^{-1}x'_t(X'W_0^{-1}X)^{-1}x_t. \tag{4.5}$$

Applying identity (4.4) in (4.1), we get

$$\begin{aligned}\hat{\beta}_{0(-t)} &= (X'W_0^{-1}X - w_{t0}^{-1}x_t x'_t)^{-1}(X'W_0^{-1}y - w_{t0}^{-1}x_t\bar{y}_t) \\ &= \hat{\beta}_0 - w_{t0}^{-1}h_{0t}^{-1}(X'W_0^{-1}X)^{-1}x_t(\bar{y}_t - x'_t\hat{\beta}_0).\end{aligned} \tag{4.6}$$

Or

$$\hat{\beta}_0 - \hat{\beta}_{0(-t)} = w_{t0}^{-1}h_{0t}^{-1}(X'W_0^{-1}X)^{-1}x_t(\bar{y}_t - x'_t\hat{\beta}_0). \tag{4.7}$$

In (4.5), from the choice of prior values and condition (A),

$$\max_t |h_{0t}| \to 1 \quad \text{as} \quad k \to +\infty.$$

Also due to the fact $k(X'W_0^{-1}X)^{-1} = O_p(1)$, (4.3) follows from

$$k^{-1/2}\max_t |\bar{y}_t - x'_t\hat{\beta}_0| = o_p(1). \tag{4.8}$$

Now

$$|\bar{y}_t - x'_t\hat{\beta}_0| = |\bar{\varepsilon}_t - (\hat{\beta}_0 - \beta)'x_t| \le |\bar{\varepsilon}_t| + |(\hat{\beta}_0 - \beta)'x_t|.$$

Since $E|\bar{\varepsilon}_t|^{2+\delta}$ is bounded, from Lemma 4.1

$$k^{-1}\max_t |\bar{\varepsilon}_t|^2 = o_p(1),$$

or

$$k^{-1/2} \max_t |\bar{\varepsilon}_t| = o_p(1). \tag{4.9}$$

Since $\hat{\beta}_0 - \beta = O_p(k^{-1/2})$ and it is independent of t,

$$k^{-1/2} \max_t |(\hat{\beta}_0 - \beta)' x_t| = o_p(1). \tag{4.10}$$

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

Remark. From Lemma 2.6 and (4.3),

$$\max_t |\hat{\beta}_{0(-t)} - \beta| = O_p(k^{-1/2}). \tag{4.11}$$

Lemma 4.3 *Suppose that conditions* (A), (B) *and* (C) *hold. Then as* $k \to +\infty$,

$$k \max_t |R_{1k} - R_{1k(-t)}| \to_p 0, \tag{4.12}$$

$$k \max_t |R_{2k} - R_{2k(-t)}| \to_p 0, \tag{4.13}$$

$$k^{1/2} \max_t |\check{\sigma}_v^2 - \check{\sigma}_{v(-t)}^2| \to_p 0. \tag{4.14}$$

Proof. Since the proof of (4.13) is similar to that of (4.12), the proof of (4.13) is omitted. Note that

$$\begin{aligned} k \max_t |R_{1k} - R_{1k(-t)}| \le \max_t \left(\frac{k}{\nu - c_t}\right) \Big[& \max_t \left|\textstyle\sum_i c_i \bar{\varepsilon}_i (\hat{\beta}_0 - \hat{\beta}_{0(-t)})' x_i\right| \\ & + \max_t \left|c_t/\nu \textstyle\sum_i c_i \bar{\varepsilon}_i (\hat{\beta}_0 - \beta)' x_i\right| + \max_t \left|c_t \bar{\varepsilon}_t (\hat{\beta}_{0(-t)} - \beta)' x_i\right| \Big], \end{aligned} \tag{4.15}$$

where $\nu = \sum_{i=1}^k c_i$. Since $\max_t k/(\nu - c_t) \le C$ 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

$$\max_t \left|c_t/\nu \textstyle\sum_i c_i \bar{\varepsilon}_i (\hat{\beta}_0 - \beta)' x_i\right| = o_p(1).$$

From (4.3) and Lemma 2.5 $\left(k^{-1/2} \sum_i c_i \bar{\varepsilon}_i x_{il} = O_p(1) \text{ for any } l\right)$

$$\begin{aligned} & \max_t \left|\textstyle\sum_i c_i \bar{\varepsilon}_i (\hat{\beta}_0 - \hat{\beta}_{0(-t)})' x_i\right| \\ & \quad \le \textstyle\sum_{l=1}^p \max_t \left|k^{1/2} (\hat{\beta}_{0_l} - \hat{\beta}_{0(-t)_l})'\right| \left|k^{-1/2} \sum_i c_i \bar{\varepsilon}_i x_{il}\right| = o_p(1), \end{aligned}$$

where $\hat{\beta}_{0_l}$ and $\hat{\beta}_{0(-t)_l}$ are the l-th elements of $\hat{\beta}_0$ and $\hat{\beta}_{0(-t)}$ respectively. From (4.11) and (4.9)

$$\max_t \left|c_t\bar{\varepsilon}_t(\hat{\beta}_{0(-t)} - \beta)'x_t\right| \le \max_t |k^{1/2}(\hat{\beta}_{0(-t)} - \beta)'| \max_t \left|k^{-1/2}c_t\varepsilon_t x_t\right| = o_p(1).$$

Therefore (4.12) is proved.

Since

$$\begin{aligned} k^{1/2}\max_t|\tilde{\sigma}^2_{v(-t)} - \tilde{\sigma}^2_v| &\le k^{1/2}\max_t|\tilde{\sigma}^2_{vp(-t)} - \tilde{\sigma}^2_{vp}| \\ &+ 2k^{1/2}\max_t|R_{1k} - R_{1k(-t)}| + k^{1/2}\max_t|R_{2k} - R_{2k(-t)}|, \end{aligned}$$

(4.14) follows from

$$k^{1/2}\max|\tilde{\sigma}^2_{vp(-t)} - \tilde{\sigma}^2_{vp}| = o_p(1).$$

In fact,

$$\begin{aligned} & k^{1/2}\max|\tilde{\sigma}^2_{vp(-t)} - \tilde{\sigma}^2_{vp}| \\ & \le \max_t[k(\nu - c_t)^{-1}]\max_t|k^{-1/2}c_tZ_t| + \max_t[k^{1/2}c_t(\nu - c_t)^{-1/2}]|\nu^{-1}\textstyle\sum_t c_tZ_t| \end{aligned}$$

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,

$$k\sum_t[R_{1k} - R_{1k(-t)}]^2 \le k^2\max_t|R_{1k} - R_{1k(-t)}|^2 = o_p(1), \tag{4.16}$$

$$k\sum_t[R_{2k} - R_{2k(-t)}]^2 \le k^2\max_t|R_{2k} - R_{2k(-t)}|^2 = o_p(1). \tag{4.17}$$

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

Theorem 4.1 *Suppose that conditions* (A), (B) *and* (C) *hold, then*

$$V_a^{-1/2}(\tilde{\sigma}^2_{v(J)} - \sigma^2_v) \to_d N(0,1) \quad as \quad k \to +\infty. \tag{4.18}$$

Proof. Since

$$\tilde{\sigma}^2_{v(J)} - \sigma^2_v = \tilde{\sigma}^2_v - \sigma^2_v + \frac{k-1}{k}\sum_{t=1}^{k}(\tilde{\sigma}^2_v - \tilde{\sigma}^2_{v(-t)}),$$

the result follows from

$$k^{1/2}\sum_t(\tilde{\sigma}_v^2 - \tilde{\sigma}_{v(-t)}^2) \to_p 0 \quad \text{as} \quad k \to +\infty\,. \tag{4.19}$$

by Theorem 2.1. Now

$$\begin{aligned} &k^{1/2}\sum_t(\tilde{\sigma}_v^2 - \tilde{\sigma}_{v(-t)}^2) = \\ &\quad k^{1/2}\sum_t(\tilde{\sigma}_{vp}^2 - \tilde{\sigma}_{vp(-t)}^2) - 2k^{1/2}\sum_t(R_{1k} - R_{1k(-t)}) + k^{1/2}\sum_t(R_{2k} - R_{2k(-t)})\,. \end{aligned} \tag{4.20}$$

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

$$\begin{aligned} k^{1/2}\sum_t(\tilde{\sigma}_{vp}^2 - \tilde{\sigma}_{vp(-t)}^2) &= k^{1/2}\sum_i\left(\frac{1}{\nu - c_i} - \nu^{-1}\sum_t\frac{c_t}{\nu - c_t}\right)c_i Z_i \\ &= k^{-1/2}\nu^{-1}\sum_i\left[\sum_t\frac{kc_t(c_i - c_t)}{(\nu - c_t)(\nu - c_t)}\right]c_i Z_i \;= o_p(1)\,. \end{aligned} \tag{4.21}$$

For the second term in the right hand side of (4.20)

$$\begin{aligned} &k^{1/2}\sum_{t=1}^{k}(R_{1k} - R_{1k(-t)}) \\ &\quad = k^{1/2}\sum_t(\nu - c_t)^{-1}\sum_i c_i\bar{\varepsilon}_i(\hat{\beta}_0 - \hat{\beta}_{0(-t)})'x_i \\ &\qquad + k^{1/2}\sum_t c_t(\nu - c_t)^{-1}\bar{\varepsilon}_t(\hat{\beta}_{0(-t)} - \beta)'x_t \\ &\qquad - k^{1/2}\nu^{-1}\sum_t c_t(\nu - c_t)^{-1}\sum_i c_i\bar{\varepsilon}_i(\hat{\beta}_0 - \beta)'x_i\,. \end{aligned} \tag{4.22}$$

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

$$k^{1/2}\sum_t c_t(\nu - c_t)^{-1}\bar{\varepsilon}_t(\hat{\beta}_{0(-t)} - \beta)'x_t = o_p(1)\,.$$

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

$$k^{1/2}\nu^{-1}\left(\sum_t c_t(\nu - c_t)^{-1}\right)\sum_i c_i\bar{\varepsilon}_i(\hat{\beta}_0 - \beta)'x_i = o_p(1)\,.$$

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

$$R_{3k} \triangleq \sum_t(\nu - c_t)^{-1}w_{i0}^{-1}h_{0t}^{-1}x_t\left(\bar{\varepsilon}_t - (\hat{\beta}_0 - \beta)'x_t\right) = o_p(1)\,.$$

Then from (4.6), the first term in (4.22) is

$$k^{1/2}\sum_t(\nu - c_t)^{-1}\sum_i c_i\tilde{\varepsilon}_i(\hat{\beta}_0 - \hat{\beta}_{0(-t)})'x_i$$
$$= \left(k^{-1/2}\sum_i c_i x_i\right)'(k(X'W_0^{-1}X)^{-1})R_{3k} = o_p(1).$$

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. □

We know from Section 2.5.2 that V_a is the asymptotic variance of $\tilde{\sigma}_v^2$ 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 $\tilde{\sigma}_{v(J)}^2$. The following theorem shows that the jackknife variance estimator $\hat{V}_{Jt}^2$ (defined in (4.2)) is consistent for V_a.

Theorem 4.2 *Suppose that conditions* (A), (B) *and* (C) *hold. Then*

$$k(\hat{V}_{Jt} - V_a) \to_p 0 \quad as \quad k \to +\infty. \tag{4.25}$$

Proof. We have

$$\hat{V}_{Jt} = (k(k-1))^{-1}\sum_{t=1}^{k}(Q_t - \bar{Q})^2$$
$$= (k-1)k^{-1}\sum_{t=1}^{k}\left[\tilde{\sigma}_v^2 - \tilde{\sigma}_{v(-t)}^2 - k^{-1}\sum_t(\tilde{\sigma}_v^2 - \tilde{\sigma}_{v(-t)}^2)\right]^2$$
$$= (k-1)k^{-1}\sum_{t=1}^{k}(\tilde{\sigma}_v^2 - \tilde{\sigma}_{v(-t)}^2)^2 - (k-1)k^{-2}\left[\sum_t(\tilde{\sigma}_v^2 - \tilde{\sigma}_{v(-t)}^2)\right]^2$$
$$= (k-1)k^{-1}\sum_{t=1}^{k}(\tilde{\sigma}_v^2 - \tilde{\sigma}_{v(-t)}^2)^2 + o_p(k^{-2}).$$

The last equality follows from (4.19). Hence

$$
\begin{aligned}
& k(\hat{V}_{Jt} - V_a) \\
&= k\Big[\frac{k-1}{k}\sum_{t=1}^{k}(\tilde{\sigma}_v^2 - \tilde{\sigma}_{v(-t)}^2)^2 - V_a\Big] + o_p(1) \\
&= k\Big[\frac{k-1}{k}\sum_{t=1}^{k}(\tilde{\sigma}_{vp}^2 - \tilde{\sigma}_{vp(-t)}^2)^2 - V_a\Big] \\
&\quad + \frac{k-1}{k}\Big[4k\sum_{t=1}^{k}(R_{1k} - R_{1k(-t)})^2 + k\sum_{t=1}^{k}(R_{2k} - R_{2k(-t)})^2\Big] \\
&\quad + (k-1)k^{-2}(\text{c.p. terms}) + o_p(1) \\
&= k\Big[\frac{k-1}{k}\sum_{t=1}^{k}(\tilde{\sigma}_{vp}^2 - \tilde{\sigma}_{vp(-t)}^2)^2 - V_a\Big] + (k-1)k^{-2}(\text{c.p. terms}) + o_p(1)
\end{aligned}
$$

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

$$
k\Big[\sum_{t=1}^{k}(\tilde{\sigma}_{vp}^2 - \tilde{\sigma}_{vp(-t)}^2)^2 - V_a\Big] = o_p(1), \tag{4.26}
$$

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

After some simplification,

$$
\begin{aligned}
& k\Big[\sum_{t=1}^{k}(\tilde{\sigma}_{vp}^2 - \tilde{\sigma}_{vp(-t)}^2)^2 - V_a\Big] \\
&= \frac{k^2}{\nu^2}k^{-1}\sum_{t=1}^{k}\Big[(Z_t - \sigma_v^2)^2 - var(Z_i)\Big] + \frac{k^2}{\nu^2}k^{-1}\sum_{t=1}^{k}\frac{(2\nu - c_t)c_t^3(Z_t - \sigma_v^2)^2}{(\nu - c_t)^2} \\
&\quad + k(\tilde{\sigma}_{vp}^2 - \sigma_v^2)^2\sum_{t=1}^{k}\frac{c_t^2}{(\nu - c_t)^2} - 2(\tilde{\sigma}_{vp}^2 - \sigma_v^2)k\sum_{i=1}^{k}\frac{c_t^2(Z_t - \sigma_v^2)}{(\nu - c_t)^2}.
\end{aligned}
$$

Note that k^2/ν^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 $\tilde{\sigma}_{vp}^2 \to_p \sigma_v^2$, and $k\sum_{t=1}^{k} c_t^2/(\nu - c_t)^2$ is bounded. The fourth term is $o_p(1)$ since $k^{1/2}(\tilde{\sigma}_{vp}^2 - \sigma_v^2) = O_p(1)$ from Theorem 2.1 and

$$
k^{1/2}\sum_{i=1}^{k}\frac{c_t^2(Z_t - \sigma_v^2)}{(\nu - c_t)^2} = o_p(1)
$$

from Lemma 2.4. The proof is complete. □

Instead of jackknifing $\tilde{\sigma}_v^2$, we can also jackknife $\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0)$, and get another estimator of σ_v^2 and a variance estimator for V_a.

Let

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

Hence a jackknife estimator of σ_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

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

We will see these estimators have the same properties as $\tilde{\sigma}_{v(J)}^2$ and V_{Jt} 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*

$$P\{\bigcup_t (\tilde{\sigma}_{v(-t)}^2 \le 0)\} \to 0 \quad as \quad k \to +\infty . \tag{4.27}$$

Proof. Note that

$$\begin{aligned}
&P\{\bigcup_t (\tilde{\sigma}_{v(-t)}^2 \le 0)\} \\
&= P\{\Big(\bigcup_t (\tilde{\sigma}_{v(-t)}^2 \le 0)\Big) \cap (\tilde{\sigma}_v^2 < \sigma_v^2/2)\} + P\{\Big(\bigcup_t (\tilde{\sigma}_{v(-t)}^2 \le 0)\Big) \cap (\tilde{\sigma}_v^2 \ge \sigma_v^2/2)\} \\
&\le P\{\tilde{\sigma}_v^2 < \sigma_v^2/2\} + P\{\bigcup_t \Big((\tilde{\sigma}_{v(-t)}^2 \le 0) \cap (\tilde{\sigma}_v^2 \ge \sigma_v^2/2)\Big)\} \\
&\le P\{\tilde{\sigma}_v^2 < \sigma_v^2/2\} + P\{\bigcup_t \Big(|\tilde{\sigma}_{v(-t)}^2 - \tilde{\sigma}_v^2| \ge \sigma_v^2/2\Big)\} \\
&= P\{\tilde{\sigma}_v^2 < \sigma_v^2/2\} + P\{\max_t |\tilde{\sigma}_{v(-t)}^2 - \tilde{\sigma}_v^2| \ge \sigma_v^2/2\} .
\end{aligned}$$

Hence (4.27) follows from (4.14) and Corollary 2.1. □

Define the set A_k as

$$A_k = \bigcup_t (\tilde{\sigma}_{v(-t)}^2 \le 0) \cup (\tilde{\sigma}_{v(J)}^2 \le 0) .$$

Then from Lemma 4.4 and Corollary 2.1,

$$P(A_k) \le P\{\bigcup_t (\tilde{\sigma}_{v(-t)}^2 \le 0)\} + P\{\tilde{\sigma}_{v(J)}^2 \le 0\} \to 0 \tag{4.28}$$

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*

$$V_a^{-1/2}(\hat{\sigma}_{v(J)}^2 - \sigma_v^2) \to_d N(0.1) \quad as \quad k \to +\infty .$$

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

Theorem 4.4 *Under the conditions of Theorem* 2.4,

$$k(\hat{V}_{Jh} - V_a) \to_p 0 \quad as \quad k \to +\infty .$$

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

Remark 1. Our simulation results indicate that both $\tilde{\sigma}_{v(J)}^2$ and $\hat{\sigma}_{v(J)}^2$ reduce the bias in $\tilde{\sigma}_v^2$ and $\hat{\sigma}_v^2$. The estimator $\hat{\sigma}_{v(J)}^2$ 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 σ_v^2:

$$\hat{V}_{Jt}^{-1/2}(\tilde{\sigma}_v^2 - \sigma_v^2).$$

Hence, a confidence interval of level $1-\alpha$ for σ_v^2 may be taken as

$$(\tilde{\sigma}_{v(J)}^2 - t_{1-\alpha/2,k-1}\hat{V}_{Jt}^{1/2}, \quad \tilde{\sigma}_{v(J)}^2 + t_{1-\alpha/2,k-1}\hat{V}_{Jt}^{1/2}), \tag{4.29}$$

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 σ_v^2 as follows

$$(\hat{\sigma}_{v(J)}^2 - t_{1-\alpha/2,k-1}\hat{V}_{Jh}^{1/2}, \quad \hat{\sigma}_{v(J)}^2 + t_{1-\alpha/2,k-1}\hat{V}_{Jh}^{1/2}). \tag{4.30}$$

But the latter approach leads to severe undercoverage as found in our simulation study in Section 4.5. Actually, when σ_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 β 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 σ_v^2 and σ_i^2, and then we get the corresponding weights $\hat{w}_{i(-t)}$ for $w_i (i \neq t)$. Using these weights we estimate β, and denote the estimator by $\hat{\beta}_{w(-t)}$. Then

$$\hat{\beta}_{w(-t)} = (X'_{(-t)}\hat{W}_{(-t)}X_{(-t)})^{-1}X'_{(-t)}\hat{W}_{(-t)}y_{(-t)} \tag{4.31}$$

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

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

with

$$\begin{aligned}\hat{w}_{i(-t)} &= (1-d_i)\hat{\sigma}^2_{v(-t)} + d_i\left[n_i^{-1}(1+\rho_{i0}n_i)^2\textstyle\sum_j(y_{ij}-\bar{y}_i)^2 + (\,y_i - \hat{\beta}'_{0t}x_i\,)^2\right],\\ &= (1-d_i)\hat{\sigma}^2_{v(-t)}\\ &\qquad + d_i\left[n_i^{-1}(1+\rho_{i0}n_i)^2\textstyle\sum_j(e_{ij}-\bar{e}_i)^2 + (\,\bar{\varepsilon}_i - (\hat{\beta}_{0(-t)}-\beta)'x_i\,)^2\right].\end{aligned}$$

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, $\hat{V}_J$ is consistent for V_k.

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

$$k(\hat{V}_J - V_k) \to_p 0 \quad as \quad k \to \infty$$

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*

$$k^{1/2} \max_{i \le k} |\hat{w}_i^{-1} u_i - 1| = O_p(1), \tag{4.32}$$

$$k^{1/2} \max_{i,t \le k, i \ne t} |\hat{w}_{i(-t)}^{-1} u_i - 1| = O_p(1), \tag{4.33}$$

$$k^{1/2} \max_{i,t \le k, i \ne t} |\hat{w}_i^{-1} u_i - \hat{w}_{i(-t)}^{-1} u_i| = o_p(1). \tag{4.34}$$

Proof. Note that

$$\begin{aligned} & k^{1/2}|\hat{w}_i u_i^{-1} - 1| \\ & \le (1 - d_i) k^{1/2} |\hat{\sigma}_v^2 - \sigma_v^2| u_i^{-1} + d_i |k^{1/2}(\hat{\beta}_0 - \beta)' x_i \, [\bar{\varepsilon}_i - (\hat{\beta}_0 - \beta)' x_i] \, u^{-1}| \\ & \le \sigma_v^2 k^{1/2} |\hat{\sigma}_v^2 - \sigma_v^2| + d_i \zeta_i (d_i^{-1/2} + k^{-1/2} \zeta_i) \end{aligned}$$

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

$$\max_i \zeta_i \le 2\sigma_v^{-1} |k^{1/2}(\hat{\beta}_0 - \beta)' x_i| = O_p(1)$$

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. □

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

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

Proof. From Lemma 3.4

$$k(X' \hat{W}^{-1} X)^{-1} = k\left(\textstyle\sum_i \hat{w}_i^{-1} x_i x_i' \right)^{-1} = O_p(1). \tag{4.35}$$

By Lemma 4.5

$$k(\sum_{i\neq t}\hat{w}_{i(-t)}^{-1}x_ix_i')^{-1} = O_p(1) \quad \text{uniformly in } t\,,$$
$$k(\sum_{i\neq t}\hat{w}_i^{-1}x_ix_i')^{-1} = O_p(1) \quad \text{uniformly in } t\,.$$

Note that

$$\begin{aligned}
&(\sum_{i\neq t}\hat{w}_{i(-t)}^{-1}x_ix_i')^{-1} - (\sum_i \hat{w}_i^{-1}x_ix_i')^{-1} \\
&= \left[(\sum_{i\neq t}\hat{w}_{i(-t)}^{-1}x_ix_i')^{-1} - (\sum_{i\neq t}\hat{w}_i^{-1}x_ix_i')^{-1}\right] \\
&\qquad + \left[(\sum_{i\neq t}\hat{w}_i^{-1}x_ix_i')^{-1} - (\sum_i \hat{w}_i^{-1}x_ix_i')^{-1}\right] \\
&= (\sum_{i\neq t}\hat{w}_{i(-t)}^{-1}x_ix_i')^{-1}(\sum_{i\neq t}(\hat{w}_i - \hat{w}_{i(-t)})x_ix_i')(\sum_{i\neq t}\hat{w}_i^{-1}x_ix_i')^{-1} \\
&\qquad + h_t(X'\hat{W}^{-1}X)^{-1}\,\hat{w}_t^{-1}x_tx_t'\,(X'\hat{W}^{-1}X)^{-1}\,, \qquad (4.36)
\end{aligned}$$

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

$$h_t = 1 - \hat{w}_t^{-1}x_t'(X'\hat{W}^{-1}X)^{-1}x_t\,.$$

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

$$\max_t |h_t|\hat{w}_t^{-1}x_tx_t' = O_p(1)\,.$$

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

$$k^{-1/2}\sum_{i\neq t}(\hat{w}_i - \hat{w}_{i(-t)})x_ix_i' = o_p(1) \quad \text{uniformly in } t.$$

From condition (A), we have

$$\|k^{-1/2}\sum_{i\neq t}(\hat{w}_i - \hat{w}_{i(-t)})x_ix_i'\| \leq Ck^{1/2}\max_{i,t\leq k, i\leq t}|\hat{w}_i^{-1}u_i - \hat{w}_{i(-t)}^{-1}u_i|k^{-1}\sum_i u_i^{-1}$$

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

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

$$\hat{\beta}_{w(-t)} - \hat{\beta}_w = -(X'D_1X)^{-1}(x_tu_t^{-1}\hat{\varepsilon}_t + 2(X'D_3D_0X)(X'W_0^{-1}X)^{-1}w_{t0}^{-1}x_t\hat{\varepsilon}_t)$$
$$+ o_p(k^{-1}) \quad \textit{uniformly in } t\,.$$

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

$$\begin{aligned}
&\hat{\beta}_{w(-t)} - \hat{\beta}_w \\
&\quad = \left(\textstyle\sum_{i\neq t} \hat{w}_{i(-t)}^{-1} x_i x_i'\right)^{-1} \textstyle\sum_{i\neq t} \hat{w}_{i(-t)}^{-1} \bar{\varepsilon}_i x_i - \left(\textstyle\sum_i \hat{w}_i^{-1} x_i x_i'\right)^{-1} \textstyle\sum_i \hat{w}_i^{-1} \bar{\varepsilon}_i x_i \\
&\quad = (X'\hat{W}^{-1}X)^{-1} \left[\textstyle\sum_{i\neq t} \hat{w}_{i(-t)}^{-1} \bar{\varepsilon}_i x_i - \textstyle\sum_i \hat{w}_i^{-1} \bar{\varepsilon}_i x_i\right] \\
&\qquad + \left[\left(\textstyle\sum_{i\neq t} \hat{w}_{i(-t)}^{-1} x_i x_i'\right)^{-1} - (X'\hat{W}^{-1}X)^{-1}\right] \textstyle\sum_{i\neq t} \hat{w}_{i(-t)}^{-1} \bar{\varepsilon}_i x_i \\
&\quad = (X'\hat{W}^{-1}X)^{-1} \left[\textstyle\sum_{i\neq t} \hat{w}_{i(-t)}^{-1} \bar{\varepsilon}_i x_i - \textstyle\sum_i \hat{w}_i^{-1} \bar{\varepsilon}_i x_i\right] + o_p(k^{-1})
\end{aligned}$$

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

$$k[(X'\hat{W}^{-1}X)^{-1} - (X'D_1^{-1}X)^{-1}] \to_p 0\,.$$

Hence the result follows if

$$\begin{aligned}
&\textstyle\sum_{i\neq t} \hat{w}_{i(-t)}^{-1} \bar{\varepsilon}_i x_i - \textstyle\sum_i \hat{w}_i^{-1} \bar{\varepsilon}_i x_i \\
&\quad = \quad -x_t u_t^{-1} \bar{\varepsilon}_t - 2(X'D_3D_0X)(X'W_0^{-1}X)^{-1} w_{t0}^{-1} x_t \bar{\varepsilon}_t + o_p(1) \quad \text{uniformly in } t.
\end{aligned}$$

The left hand side of the previous expression is equal to

$$\textstyle\sum_{i\neq t} (\hat{w}_{i(-t)}^{-1} - \hat{w}_i^{-1}) \bar{\varepsilon}_i x_i - \hat{w}_t^{-1} \bar{\varepsilon}_t x_t\,.$$

From (4.32)

$$\hat{w}_t^{-1} \bar{\varepsilon}_t x_t = u_t^{-1} \bar{\varepsilon}_t x_t + o_p(1) \quad \text{uniformly in } t\,.$$

Also from (4.7), by using Lemma 2.6,

$$\hat{\beta}_{0(-t)} - \hat{\beta}_0 = -w_{t0}^{-1} (X'W_0^{-1}X)^{-1} x_t \bar{\varepsilon}_t + O_p(k^{3/2}) \quad \text{uniformly in } t\,.$$

Hence it suffices to show

$$\textstyle\sum_{i\neq t} (\hat{w}_{i(-t)}^{-1} - \hat{w}_i^{-1}) \bar{\varepsilon}_i x_i = 2(X'D_3D_0X)(\hat{\beta}_{0(-t)} - \hat{\beta}_0) + o_p(1) \quad \text{uniformly in } t\,. \tag{4.37}$$

Since

$$\begin{aligned}
\hat{w}_{i(-t)} - \hat{w}_i \quad &= \quad (1 - d_i)(\hat{\sigma}_{v(-t)}^2 - \hat{\sigma}_v^2) \\
&\qquad + d_i \left[x_i'(\hat{\beta}_0 - \hat{\beta}_{0(-t)}) x_i'(2\bar{\varepsilon}_i - (\hat{\beta}_0 + \hat{\beta}_{0(-t)} - 2\beta)\right]\,,
\end{aligned}$$

we have

$$\begin{aligned}
&\textstyle\sum_{i\neq t}(\hat{w}_{i(-t)}^{-1} - \hat{w}_i^{-1})\bar{\varepsilon}_i x_i = \\
&\quad 2\textstyle\sum_{i\neq t} d_i \hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1}\bar{\varepsilon}_i^2 x_i x_i'(\hat{\beta}_{0(-t)} - \hat{\beta}_0) \\
&\quad + (\hat{\sigma}_{v(-t)}^2 - \hat{\sigma}_v^2)\textstyle\sum_{i\neq t}(1-d_i)\hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1}\bar{\varepsilon}_i x_i \\
&\quad + \textstyle\sum_{i\neq t} d_i \hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1}\bar{\varepsilon}_i x_i x_i'(\hat{\beta}_0 - \hat{\beta}_{0(-t)})x_i'(\hat{\beta}_0 + \hat{\beta}_{0(-t)} - 2\beta) \\
&\quad \triangleq \text{(I)} + \text{(II)} + \text{(III)} \overset{\text{claim}}{=} \text{(I)} + o_p(1).
\end{aligned} \tag{4.38}$$

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

$$k^{-1/2}\sum_{i\neq t} u_i^{-2}\bar{\varepsilon}_i x_i = O_p(1) \quad \text{uniformly in } t,$$

and Lemma 4.5, we have

$$k^{-1/2}\sum_{i\neq t}\hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1}x_i\bar{\varepsilon}_i = O_p(1) \quad \text{uniformly in } t\,.$$

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),

$$\|\textstyle\sum_{i\neq t} d_i \hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1}\bar{\varepsilon}_i x_i x_i'(\hat{\beta}_0 - \hat{\beta}_{0(-t)})x_i'(\hat{\beta}_0 + \hat{\beta}_{0(-t)} - 2\beta)\| \le o_p(k^{-1})\sum_{i\neq t}\hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1}|\bar{\varepsilon}_i|.$$

Note that, from (4.32), (4.33)

$$\begin{aligned}
&|k^{-1/2}\textstyle\sum_{i\neq t}|\bar{\varepsilon}|(u_i^2 - \hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1})| \\
&\le \Big[(\max_{i\neq t}|u_i\hat{w}_{i(-t)}^{-1} - 1| + 3)(k^{1/2}\max_{i\le k}|u_i\hat{w}_i^{-1} - 1|) + k^{1/2}\max_{i\neq t}|u_i\hat{w}_{i(-t)}^{-1} - 1|\Big] \\
&\quad \times Ck^{-1}\textstyle\sum_{i\neq t}u_i^{-2}|\bar{\varepsilon}_i| = O_p(1) \quad \text{uniformly in } t\,.
\end{aligned} \tag{4.39}$$

Hence

$$\sum_{i\neq t}\hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1}|\bar{\varepsilon}_i| = \sum_{i\neq t}u_i^{-2}|\bar{\varepsilon}_i| + O_p(k^{1/2}) = O_p(k) \quad \text{uniformly in } t\,.$$

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

Back to term (I) in (4.38). From Lemma 3.1, we have

$$k^{-1}(X'D_0D_3X - \sum_{i=1}^{k} d_i x_i x_i' u_i^{-2}\bar{\varepsilon}_i^2) \to_p 0\,.$$

And from Lemma 4.1,

$$k^{-1} \max_t \|d_t x_t x_t' u_t^{-2}\bar{\varepsilon}_t^2\| \to_p 0\,.$$

Hence

$$k^{-1}(X'D_0D_3X - \sum_{i\neq t}^{k} d_i x_i x_i' u_i^{-2}\bar{\varepsilon}_i^2) \to_p 0 \quad \text{uniformly in } t\,.$$

Using argument similar to (4.39), we have

$$k^{-1/2}\sum_{i\neq t} d_i x_i x_i' \bar{\varepsilon}_i^2 (u_i^2 - \hat{w}_{i(-t)}^{-1}\hat{w}_i^{-1}) = O_p(1) \quad \text{uniformly in } t.$$

Hence, in term (I)

$$k^{-1}\Big(X'D_0D_3X - \sum_{i\neq t} d_i \hat{w}_{i(-t)}^{-1}\hat{w}_{i(-t)}^{-1}\bar{\varepsilon}_i^2 x_i x_i'\Big) \to_p 0 \quad \text{uniformly in } t\,.$$

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

Proof of Theorem 4.5. From Lemma 4.7,

$$\begin{aligned}
k(\hat{V}_J - V_k) = {} & [k(X'D_1X)^{-1}]\Big[k^{-1}\sum_t (u_t^{-2}\bar{\varepsilon}_t^2 - \tau_{3i})x_t x_t' \\
& + 2(X'D_3D_0X)(X'W_0^{-1}X)^{-1}\Big(k^{-1}\sum_t w_{i0}^{-1}(u_i^{-1}\bar{\varepsilon}_i^2 - \tau_{2i})x_t x_t'\Big) \\
& + 2\Big(k^{-1}\sum_t w_{i0}^{-1}(u_i^{-1}\bar{\varepsilon}_i^2 - \tau_{2i})x_t x_t'\Big)(X'W_0^{-1}X)^{-1}(X'D_0D_3X) \\
& + 4(X'D_3D_0X)(X'W_0^{-1}X)^{-1}\Big(k^{-1}\sum_t w_{i0}^{-2}(\bar{\varepsilon}_t^2 \\
& \qquad - w_i)x_t x_t'\Big)(X'W_0^{-1}X)^{-1}(X'D_0D_3X)\Big] \times [k(X'D_1X)^{-1}] + o_p(1)\,.
\end{aligned}$$

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

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 β, using $\hat{\beta}_w$ and $\hat{V}_J$, by substituing $\hat{V}_J$ in place of $\hat{V}_k$ in interval (3.16). Also the confidence intervals on elements of β 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 $\hat{\sigma}^2_{v(-t)}$ of σ^2_v. Then we can construct weights $\hat{w}_{i(-t),s}$ for w_i $(i \neq t)$. Using these weights, we can compute a WLS estimator $\hat{\beta}_{sw(-t)}$ of β as follows:

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

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

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

with

$$\hat{w}_{i(-t),s} = \hat{\sigma}^2_{v(-t)} + n_i^{-1}s_i^2.$$

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

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

We now give the following theorem on the consistency of $\hat{V}_{sJ}$. 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*

$$k(\hat{V}_{sJ} - V_{sk}) \to_p 0 \qquad \square$$

We can construct confidence intervals on the regression coefficients using the estimator $\hat{V}_{sJ}$ of the covariance matrix. These confidence intervals performed well with regard to coverage probability, in a simulation study, but the variance estimator $\hat{V}_{sJ}$ is very unstable, and biased, when σ^2_v is relatively small or moderate, and the n_i'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 σ_v^2, viz., the untruncated and truncated AMINQU estimators $\tilde{\sigma}_v^2$, $\hat{\sigma}_v^2$ (given in Section 2.4), and the corresponding jackknife estimators $\tilde{\sigma}_{v(J)}^2$ and $\hat{\sigma}_{v(J)}^2$ (given in Section 4.2). We have shown that all these variance estimators are asymptotically normally distributed with common asymptotic variance V_a when $\sigma_v^2 > 0$. For estimation of V_a, we proposed two substitution estimators $\tilde{V}_a$ and $\hat{V}_a$ in Section 2.5.3 by using $\tilde{\sigma}_v^2$ and $\hat{\sigma}_v^2$, respectively. We also obtained two jackknife estimators $\hat{V}_{Jt}$ and $\hat{V}_{Jh}$ of V_a by delete–group jackknifing $\tilde{\sigma}_v^2$ and $\hat{\sigma}_v^2$, respectively, in Section 4.2. In this section, we investigate the relative biases of all four estimators of σ_v^2, as well as the coverage probabilities (CP) and lengths of confidence intervals (LCI) associated with these estimators of σ_v^2 and corresponding variance estimators.

For the estimation of regression parameters β, we focus on the behaviour of estimators of asymptotic covariance matrix since all WLS estimators of β 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 $\hat{V}_J^s$ 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 σ_v^2 of v_i are chosen to be 0.1, 0.25, 0.5, and 2.0. And the error variances σ_i^2 are chosen such that one third of the groups (10 groups) have error variance σ_1^2, one third have error variance σ_2^2, and the remaining one third have error variance σ_3^2. The selected patterns for $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ 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) $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$ and $e_{ij} \overset{iid}{\sim} N(0, \sigma_i^2)$ for each i.

(2) $v_i \overset{iid}{\sim}$ Double Exp$(\sigma_v/\sqrt{2})$, and $e_{ij} \overset{iid}{\sim}$ Double Exp$(\sigma_i/\sqrt{2})$ for each i, where Double Exp(λ), 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) $v_i \overset{iid}{\sim}$ uniform distribution on $(-\sigma_v\sqrt{3}, \sigma_v\sqrt{3})$, $e_{ij} \overset{iid}{\sim}$ uniform distribution on $(-\sigma_i\sqrt{3}, \sigma_i\sqrt{3})$ for each i.

The prior values chosen for the AMINQUE procedures are $\sigma_{v0}^2 = 1$ and $\sigma_{i0}^2 = 1$ for all i.

For each set of distributions, and each combination of selected values of σ_v^2 and $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, 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 $\tilde{\sigma}_v^2$ and $\hat{\sigma}_v^2$, and the corresponding variance estimators $\tilde{V}_a$ and $\hat{V}_a$; the coverage probability (CP) and length of confidence intervals (LCI) at level 90% associated with $\tilde{\sigma}_v^2$ and $\tilde{V}_a$, 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 $\tilde{\sigma}_{v(J)}^2$ and $\hat{\sigma}_{v(J)}^2$, and the corresponding jackknife variance estimators $\hat{V}_{J\ell}$ and $\hat{V}_{Jh}$; the CP and LCI at level 90% associated with $\tilde{\sigma}_{v(J)}^2$ and $\hat{V}_{J\ell}$, and the CP and LCI at level 90% associated with $\hat{\sigma}_{v(J)}^2$ and $\hat{V}_{Jh}$.

(3) The substitution variance estimator $\hat{V}_k$ and the jackknife variance estimator $\hat{V}_J$ associated with the WLS estimator $\hat{\beta}_w$ ($\beta = \mu$ here); the CP and LCI at level 90%

using $\hat{V}_k$ and $\hat{V}_J$ 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 $\hat{\beta}_{sw}$ of β, the CP, LCI, RB and CV associated with substitution variance estimator $\hat{V}_{sk}$.

(5) The CP, LCI, RB and CV associated with the WLS estimator $\hat{\beta}_w^{rs}$ 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 $\hat{\beta}_{sw}$ and Shao and Rao's WLS estimator β_w^{rs}.

Here the coverage probability (CP) and length of confidence interval (LCI) at level 90% associated with an estimator $\hat{\theta}$ and its variance estimator $\hat{V}_\theta$ (for a parameter θ), 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 $\hat{V}_\theta$, and the relative bias (RB_v) of $\hat{\theta}$, are defined as

$$\mathrm{CP} = 100 \times \sum_{r=1}^{1000} I\Big(\hat{\theta}^{(r)} - t_{.95,29}(\hat{V}_\theta^{(r)})^{1/2} \le \theta_0 \le \hat{\theta}^{(r)} + t_{.95,29}(\hat{V}_\theta^{(r)})^{1/2}\Big)/1000,$$

$$\mathrm{LCI} = 2\sum_{r=1}^{1000} t_{.95,29}(\hat{V}_\theta^{(r)})^{1/2}/1000,$$

$$\mathrm{EV}(\hat{\theta}) = \sum_{r=1}^{1000}(\hat{\theta}^{(r)})^2/1000 - \Big(\sum_{r=1}^{1000}\hat{\theta}^{(r)}/1000\Big)^2,$$

$$\mathrm{RB} = 100 \times \Big(\frac{1}{1000}\sum_{r=1}^{1000}\hat{V}_\theta^{(r)}/\mathrm{EV}(\hat{\theta}) - 1\Big),$$

$$\mathrm{CV} = 100 \times \sqrt{\mathrm{MSE}(\hat{V}_\theta)}/\mathrm{EV}(\hat{\theta}),$$

$$\mathrm{RB}_v = 100 \times \Big(\frac{1}{1000}\sum_{r=1}^{1000}\hat{\theta}^{(r)}/\theta_0 - 1\Big),$$

where

$$\mathrm{MSE}(\hat{V}_\theta) = \sum_{r=1}^{1000}[\hat{V}_\theta^{(r)} - \mathrm{EV}(\hat{\theta})]^2/1000,$$

θ_0 is the true value of the parameter θ, $I(a \le \theta_0 \le b) = 1$ if $a \le \theta_0 \le b$, and $= 0$ otherwise, $\hat{\theta}^{(r)}$ and $\hat{V}_\theta^{(r)}$ are the calculated values of $\hat{\theta}$ and $\hat{V}_\theta$ from the data in the r-th run ($r = 1, \ldots, 1000$), and $t_{.95,29}$ is the 95% percentile of a student t–variate with 29 degrees of freedom.

The percentage efficiency (E) of an estimator $\hat{\beta}^{(2)}$ relative to another estimator $\hat{\beta}^{(1)}$ of β is defined as

$$\mathrm{E}(\hat{\beta}^{(2)}, \hat{\beta}^{(1)}) = 100 \times \mathrm{EV}(\hat{\beta}^{(1)})/\mathrm{EV}(\hat{\beta}^{(2)}).$$

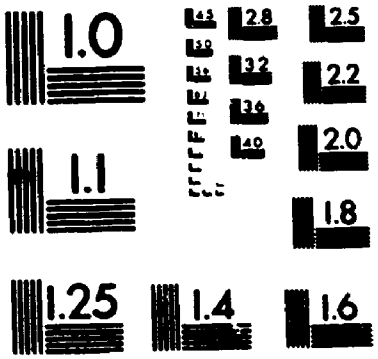
We found that the results when the errors follow a uniform distribution are quite similar to the results when the errors follow a normal distribution. Hence only results for errors following normal and double exponential distributions are presented in Tables 4.1–4.8.

4.5.2 Empirical results on estimating σ_v^2

Table 4.1 reports the values of the relative biases (RB_v) of four estimators of σ_v^2 under normal and double exponential distributions respectively. We have the following results: (1) Both AMINQUE estimators $\tilde{\sigma}_v^2$ and $\hat{\sigma}_v^2$ underestimate σ_v^2 slightly when σ_v^2 is moderate or large relative to σ_i^2's. (2) The relative biases decrease as σ_v^2 increases, while they increase as the heteroscedasticity increases. (3) Both jackknife estimators of σ_v^2 reduces bias. The truncated jackknife estimator $\hat{\sigma}_{vJ}^2$ is almost unbiased for all cases, while the untruncated jackknife estimator $\tilde{\sigma}_{vJ}^2$ is almost unbiased only for moderate or large σ_v^2 relative to σ_i^2'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 σ_v^2 and corresponding variance estimators. We have the following results: (1) The confidence intervals using truncated variance estimator $\hat{V}_a$ with truncated AMINQU estimator $\hat{\sigma}_v^2$ or using corresponding truncated jackknife estimator $\hat{V}_{Jh}$ with $\hat{\sigma}_{Jh}^2$ fail to give good coverage when σ_v^2 is small relative to σ_i^2's. From the lengths of confidence intervals, we can see that the first one severely overcovers σ_v^2 due to truncation, and the second one seriously undercovers σ_v^2. (2) The untruncated estimator $\tilde{\sigma}_v^2$ and the corresponding variance estimator $\tilde{V}_a$ by substitution lead to undercoverage which increases with σ_v^2. 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 $\tilde{\sigma}_{v(J)}^2$ and the variance esti-

2 of/de 2



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

mator $\hat{V}_{Jt}$ give reasonably good coverage in the case of normal distribution when σ_v^2 is small (≤ 0.5), but lead to undercoverage as σ_v^2 increases. The undercoverage is again more severe in the case of double exponential distribution.

Although our methods for inference on σ_v^2 are asymptotically correct, the above empirical results suggest the need for alternative methods using suitable transformations on σ_v^2 to reduce the undercoverage in finite samples, especially for long-tailed distributions and larger σ_v^2.

4.5.3 Empirical results on estimating β

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 $\hat{\beta}_w^{rs}$. 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 σ_v^2 is small or moderate relative to σ_i^2. (2) In the case of normal distribution, our estimator $\hat{\beta}_w$, compared to Shao and Rao's estimator $\hat{\beta}_w^{rs}$, gains efficiency when σ_v^2 is relatively moderate or large, and loses efficiency when σ_v^2 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 β 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 $\hat{\beta}_w^{rs}$ and associated delete-group jackknife variance estimator $\hat{V}_J^{rs}$, our method based on WLS estimator $\hat{\beta}_w$ and associated substitution variance estimator $\hat{V}_k$ 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 $\hat{V}_k$. (3) The alternative substitution variance estimator, $\hat{V}_{sk}$, performs well with respect to CP, but leads to wider LCI compared to $\hat{V}_k$ or $\hat{V}_J$ when σ_v^2 is small or moderate relative to σ_i^2's and errors follow normal distribution or for all σ_v^2

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 $\hat{V}_J^{r*}$ decrease as σ_v^2 increases. The relative biases of our variance estimators $\hat{V}_k$, $\hat{V}_J$ and $\hat{V}_{sk}$ are small ($|\text{RB}| < 10\%$) for moderate or large σ_v^2 relative to σ_i^2's, while they are over 20% for some cases when σ_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 σ_v^2. (2) The alternative substitution variance estimator $\hat{V}_{sk}$ is less stable than $\hat{V}_k$, especially for small or moderate σ_v^2. (3) Shao and Rao's variance estimator $\hat{V}_J^{r*}$ is less stable than $\hat{V}_k$, especially as σ_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 $\hat{V}_J^{r*}$. 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 σ_v^2

		Normal				Double Exponential			
σ_v^2	σ_i^2	$\hat{\sigma}_v^2$	$\tilde{\sigma}_v^2$	$\hat{\sigma}_{v(J)}^2$	$\tilde{\sigma}_{v(J)}^2$	$\hat{\sigma}_v^2$	$\tilde{\sigma}_v^2$	$\hat{\sigma}_{v(J)}^2$	$\tilde{\sigma}_{v(J)}^2$
	(i)	10.0	−15.3	0.2	−3.5	8.0	−17.5	−2.4	−2.8
0.1	(ii)	10.6	−16.3	0.2	−4.4	8.6	−18.5	−2.5	−2.0
	(iii)	43.3	−24.1	−1.2	14.3	34.6	−30.2	−7.9	7.7
	(i)	−6.0	−8.4	−0.4	−3.6	−6.0	−9.1	−1.1	−4.2
0.25	(ii)	−6.0	−8.8	−0.4	−3.6	−6.0	−9.5	−1.1	−4.3
	(iii)	0.4	−12.0	−0.8	7.2	−1.1	−14.3	−3.4	7.0
	(i)	−6.2	−6.4	−0.6	−1.2	−5.9	−6.2	−0.5	−1.3
0.5	(ii)	−6.2	−6.6	−0.6	−1.0	−6.0	−6.4	−0.5	−1.5
	(iii)	−6.6	−8.0	−0.8	4.4	−6.4	−8.9	−1.8	3.9
	(i)	−4.9	−4.9	−0.9	−0.9	−3.9	−3.9	0.0	0.0
2	(ii)	−4.9	−4.9	−0.9	−0.9	−3.9	−3.9	0.0	0.0
	(iii)	−5.2	−5.2	−1.0	1.0	−4.6	−4.6	−0.3	0.4

(1) Patterns for σ_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.

$\tilde{\sigma}_v^2$ = untruncated AMINQU estimator.

$\hat{\sigma}_{v(J)}^2$ = jackknife estimator corresponding $\hat{\sigma}_v^2$.

$\tilde{\sigma}_{v(J)}^2$ = jackknife estimator corresponding $\tilde{\sigma}_v^2$.

Table 4.2: Coverage Probability (CP %) and Length of Confidence Interval (LCI) for Estimators of σ_v^2 at Level $1-\alpha = 90\%$, when Errors Follow Normal Distribution

σ_v^2	σ_i^2	$(\hat{\sigma}_v^2,\ \hat{V}_a)$	$(\tilde{\sigma}_v^2,\ \tilde{V}_a)$	$(\hat{\sigma}_{v(J)}^2,\ \hat{V}_{Jh})$	$(\tilde{\sigma}_{v(J)}^2,\ \hat{V}_{Jt})$
	(i)	98.0 (0.242)	87.0 (0.241)	62.2 (0.184)	89.3 (0.251)
0.1	(ii)	98.0 (0.249)	87.3 (0.248)	61.3 (0.187)	89.3 (0.258)
	(iii)	97.9 (0.409)	87.9 (0.407)	54.9 (0.271)	88.9 (0.423)
	(i)	88.4 (0.289)	86.4 (0.289)	82.0 (0.278)	87.9 (0.302)
0.25	(ii)	88.9 (0.297)	86.4 (0.297)	81.5 (0.283)	88.3 (0.311)
	(iii)	96.2 (0.450)	86.1 (0.449)	69.3 (0.378)	88.0 (0.468)
	(i)	85.1 (0.376)	85.1 (0.376)	86.9 (0.390)	87.3 (0.394)
0.5	(ii)	85.4 (0.385)	85.4 (0.385)	87.1 (0.398)	87.4 (0.403)
	(iii)	86.7 (0.526)	85.3 (0.526)	83.0 (0.519)	87.6 (0.550)
	(i)	82.9 (0.942)	82.9 (0.942)	84.9 (0.990)	84.9 (0.990)
2	(ii)	83.0 (0.953)	83.0 (0.953)	85.2 (1.002)	85.2 (1.002)
	(iii)	83.6 (1.062)	83.6 (1.062)	86.1 (1.115)	86.1 (1.115)

(1) Note: The pair $(a,\ b)$ implies that the quantities in the columns are obtained by using a as estimator of σ_v^2, and b as estimator of asymptotic variance.

(2) Patterns for σ_i^2: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(3) $\hat{\sigma}_v^2$, $\tilde{\sigma}_v^2$, $\hat{\sigma}_{v(J)}^2$ and $\tilde{\sigma}_{v(J)}^2$ are the same those in Table 4.1.

(4) $\hat{V}_a =$ substitution estimator of V_a using $\hat{\sigma}_v^2$.

$\tilde{V}_a =$ substitution estimator of V_a using $\tilde{\sigma}_v^2$.

$\hat{V}_{Jh} =$ 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 σ_v^2 at level $1-\alpha = 90\%$, when Errors follow Double Exponential Dist·.bution

σ_v^2	σ_i^2	$(\hat{\sigma}_v^2,\ \hat{V}_a)$	$(\tilde{\sigma}_v^2,\ \tilde{V}_a)$	$(\hat{\sigma}_{v(J)}^2,\ \hat{V}_{Jh})$	$(\tilde{\sigma}_{v(J)}^2,\ \hat{V}_{Jt})$
	(1/2, 1, 2)	96.1 (0.231)	86.2 (0.230)	62.7 (0.176)	88.4 (0.240)
0.1	(1/4, 1, 2)	96.2 (0.237)	85.8 (0.236)	62.3 (0.179)	88.0 (0.246)
	(1/4, 1, 4)	97.6 (0.380)	88.2 (0.378)	55.3 (0.251)	90.8 (0.393)
	(1/2, 1, 2)	84.7 (0.294)	82.1 (0.294)	77.6 (0.281)	84.5 (0.308)
0.25	(1/4, 1, 2)	85.3 (0.301)	82.5 (0.301)	77.2 (0.285)	85.5 (0.314)
	(1/4, 1, 4)	93.1 (0.434)	86.1 (0.433)	69.4 (0.364)	88.1 (0.451)
	(1/2, 1, 2)	79.6 (0.416)	79.4 (0.416)	80.7 (0.429)	81.7 (0.436)
0.5	(1/4, 1, 2)	80.8 (0.423)	80.6 (0.423)	81.7 (0.434)	82.8 (0.443)
	(1/4, 1, 4)	84.5 (0.542)	82.2 (0.542)	79.5 (0.528)	84.0 (0.566)
	(1/2, 1, 2)	77.5 (1.226)	77.5 (1.226)	79.9 (1.289)	79.9 (1.289)
2	(1/4, 1, 2)	77.9 (1.232)	77.9 (1.232)	80.3 (1.295)	80.3 (1.295)
	(1/4, 1, 4)	76.6 (1.316)	76.6 (1.316)	80.1 (1.382)	80.1 (1.382)

(1) Note: The pair $(a,\ b)$ implies that the quantities in the columns are obtained by using a as estimator of σ_v^2, and b as estimator of asymptotic variance.

(2) Patterns for σ_i^2: (i) (1/2, 1, 2); (ii) (1/4, 1, 2); (iii) (1/4, 1, 4).

(3) $\hat{\sigma}_v^2$, $\tilde{\sigma}_v^2$, $\hat{\sigma}_{v(J)}^2$ and $\tilde{\sigma}_{v(J)}^2$ are the same those in Table 4.1.

(4) $\hat{V}_a =$ substitution estimator of V_a using $\hat{\sigma}_v^2$.

$\tilde{V}_a =$ substitution estimator of V_a using $\tilde{\sigma}_v^2$.

$\hat{V}_{Jh} =$ truncated jackknife variance estimator.

$\hat{V}_{Jt} =$ untruncated jackknife variance estimator.

Table 4.4: Percentage Efficiency of $\hat{\beta}_w$ Relative to $\hat{\beta}_{sw}$ and $\hat{\beta}_w^{rs}$

		Normal		Double Exponential	
σ_v^2	σ_i^2	$E(\hat{\beta}_w, \hat{\beta}_{sw})$	$E(\hat{\beta}_w, \hat{\beta}_w^{rs})$	$E(\hat{\beta}_w, \hat{\beta}_{sw})$	$E(\hat{\beta}_w, \hat{\beta}_w^{rs})$
	(i)	191	89	173	101
0.1	(ii)	192	84	166	96
	(iii)	213	97	188	113
	(i)	121	102	123	102
0.25	(ii)	124	99	122	101
	(iii)	181	103	167	107
	(i)	105	108	112	100
0.5	(ii)	105	105	113	99
	(iii)	114	107	121	104
	(i)	100	108	109	95
2	(ii)	100	109	109	95
	(iii)	101	108	110	97

(1) Patterns for σ_i^2: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{\beta}_w =$ AMINQU WLS estimator.

$\hat{\beta}_{sw} =$ alternative WLS estimator using sample variance s_i^2 as estimator of σ_i^2.

$\hat{\beta}_w^{rs} =$ Shao and Rao's (1993) WLS estimator.

Table 4.5: Coverage Probability (CP %) of Confidence Intervals on β for Four Variance Estimators at level $1-\alpha = 90\%$

		Normal				Double Exponential			
σ_v^2	σ_i^2	$\hat{V}_J^{rs}$	$\hat{V}_k$	$\hat{V}_J$	$\hat{V}_{sk}$	$\hat{V}_J^{rs}$	$\hat{V}_k$	$\hat{V}_J$	$\hat{V}_{sk}$
	(i)	86.8	90.8	91.4	90.9	87.9	90.7	90.4	89.6
0.1	(ii)	87.0	90.5	90.3	90.1	87.7	89.4	89.6	89.9
	(iii)	87.8	91.0	91.8	91.0	88.3	90.6	90.9	89.4
	(i)	86.1	89.0	91.0	90.1	87.8	91.1	92.1	91.4
0.25	(ii)	86 6	89.4	91.1	90.2	87.4	90.8	92.2	91.0
	(iii)	87.5	89.6	91.7	90.6	87.4	91.5	92.1	91.2
	(i)	86.9	88.6	88.9	89.1	88.2	91.2	91.3	91.8
0.5	(ii)	86.7	88.7	89.6	89.2	87.9	91.8	92.0	91.7
	(iii)	86.3	88.7	89.8	89.9	87.7	91.9	92.7	92.0
	(i)	84.1	87.9	88.1	87.5	89.3	91.5	92.1	91.6
2	(ii)	84.7	87.5	88.0	87.5	88.5	91.6	92.1	91.7
	(iii)	84.0	88.1	88.3	88.1	88.5	91.9	92.4	92.1

(1) Patterns for σ_i^2: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{V}_J^{rs}$ = 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 4.6: Average Lengths of Confidence Intervals (LCI) for Four Variance Estimators at Level $1-\alpha = 90\%$

		Normal				Double Exponential			
σ_v^2	σ_i^2	$\hat{V}_J^{rs}$	$\hat{V}_k$	$\hat{V}_J$	$\hat{V}_{sk}$	$\hat{V}_J^{rs}$	$\hat{V}_k$	$\hat{V}_J$	$\hat{V}_{sk}$
	(i)	0.414	0.437	0.469	0.523	0.366	0.366	0.386	0.422
0.1	(ii)	0.446	0.487	0.520	0.585	0.390	0.396	0.421	0.451
	(iii)	0.462	0.472	0.511	0.584	0.416	0.404	0.425	0.474
	(i)	0.493	0.486	0.512	0.519	0.436	0.434	0.449	0.466
0.25	(ii)	0.517	0.519	0.549	0.556	0.457	0.455	0.473	0.486
	(iii)	0.540	0.540	0.577	0.632	0.485	0.472	0.494	0.540
	(i)	0.594	0.570	0.583	0.576	0.529	0.525	0.533	0.544
0.5	(ii)	0.613	0.590	0.605	0.598	0.542	0.540	0.550	0.560
	(iii)	0.637	0.618	0.641	0.649	0.573	0.561	0.573	0.601
	(i)	0.987	0.935	0.946	0.931	0.879	0.891	0.897	0.916
2	(ii)	0.999	0.943	0.954	0.938	0.886	0.898	0.905	0.923
	(iii)	1.022	0.973	0.984	0.969	0.913	0.918	0.925	0.944

(1) Patterns for σ_i^2: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{V}_J^{rs}$ = 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

		Normal				Double Exponential			
σ_v^2	σ_i^2	$\hat{V}_J^{r*}$	$\hat{V}_k$	$\hat{V}_J$	$\hat{V}_{sk}$	$\hat{V}_J^{r*}$	$\hat{V}_k$	$\hat{V}_J$	$\hat{V}_{sk}$
	(i)	15.12	10.63	32.33	6.60	15.26	10.77	26.51	3.23
0.1	(ii)	2.06	9.62	29.14	5.52	12.17	6.76	25.51	0.53
	(iii)	18.26	15.77	39.70	9.29	18.27	20.47	37.58	7.43
	(i)	10.67	2.26	15.39	3.34	16.56	9.26	18.93	9.88
0.25	(ii)	8.10	1.97	16.20	2.86	14.93	7.88	18.12	8.47
	(iii)	12.75	8.96	28.82	4.91	18.53	10.94	24.80	2.43
	(i)	8.32	−1.88	3.87	−3.19	18.65	7.06	10.69	4.97
0.5	(ii)	7.90	−2.21	4.33	−3.17	16.71	6.06	10.48	4.30
	(iii)	8.99	1.59	11.27	4.87	19.58	9.60	14.99	9.61
	(i)	5.36	−6.30	−4.03	−7.63	19.81	6.98	8.70	4.21
2	(ii)	4.71	−6.46	−4.13	−7.74	19.57	6.78	8.54	4.13
	(iii)	6.49	−5.01	−2.82	−6.41	20.48	7.69	9.42	4.89

(1) Patterns for σ_i^2: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{V}_J^{r*}$ = 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

		Normal				Double Exponential			
σ_v^2	σ_i^2	$\hat{V}_J^{rs}$	$\hat{V}_k$	$\hat{V}_J$	$\hat{V}_{sk}$	$\hat{V}_J^{rs}$	$\hat{V}_k$	$\hat{V}_J$	$\hat{V}_{sk}$
	(i)	74.70	59.83	104.22	215.24	80.82	59.40	90.80	193.02
0.1	(ii)	68.59	60.24	98.10	212.80	73.24	56.18	96.87	189.03
	(iii)	80.48	65.84	106.83	216.78	88.84	78.31	121.08	189.06
	(i)	70.16	35.57	64.25	126.78	82.95	44.90	70.96	122.62
0.25	(ii)	65.93	37.47	65.85	131.96	76.64	43.54	64.75	138.25
	(iii)	74.18	51.02	98.70	227.63	88.53	52.74	83.19	174.83
	(i)	67.93	26.58	45.96	39.26	85.72	38.11	43.51	79.28
0.5	(ii)	63.59	26.73	56.95	37.84	81.79	36.80	44.05	95.77
	(iii)	70.94	32.77	55.37	132.16	87.52	43.73	51.22	108.31
	(i)	67.66	25.79	26.12	25.42	90.57	39.88	41.34	41.18
2	(ii)	64.07	25.78	26.09	25.40	88.88	39.34	40.84	40.74
	(iii)	70.68	26.01	26.50	25.63	93.60	40.21	41.84	41.11

(1) Patterns for σ_i^2: (i) $(1/2, 1, 2)$; (ii) $(1/4, 1, 2)$; (iii) $(1/4, 1, 4)$.

(2) $\hat{V}_J^{rs}$ = 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 $\hat{\beta}_{sw}$.

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:

$$y_{ijk} = x_i'\beta + \varepsilon_{ijk} \quad \text{with} \quad \varepsilon_{ijk} = v_i + u_{ij} + e_{ijk}\,, \tag{5.1}$$

$$i = 1, \ldots, m, \quad j = 1, \ldots, m_i, \quad k = 1, \ldots, n_{ij},$$

$$\sum_{j=1}^{m_i} n_{ij} = n_{i.}, \quad \sum_{i=1}^{m} n_{i.} = n\,,$$

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 σ_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, $e_{ijk} \overset{iid}{\sim} (0, \sigma_i^2)$ 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

$$\begin{aligned}
y_{ij} &= (y_{ij1}, \ldots, y_{ijn_{ij}})'\,,\\
\varepsilon_{ij} &= (\varepsilon_{ij1}, \ldots, \varepsilon_{ijn_{ij}})'\,,\\
\mathcal{Y} &= (y_{11}' : \cdots : y_{1m_1}' : \cdots : y_{m1}' : \cdots : y_{mn_m}')',\\
\mathcal{X} &= (x_1 1_{n_1}' : x_2 1_{n_2}' : \cdots : x_m 1_{n_m}')',\\
\mathcal{E} &= (\varepsilon_{11}' : \cdots : \varepsilon_{1n_1}' : \cdots : \varepsilon_{m1}' : \cdots : \varepsilon_{mn_m}')'.
\end{aligned}$$

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

$$\mathcal{Y} = \mathcal{X}\beta + \mathcal{E}, \tag{5.2}$$

with dispersion matrix

$$V = Var(\mathcal{E}) = \text{block diag}_i(V_i)\,,$$

where

$$\begin{aligned} V_i &= \sigma_i^2 I_{n_i} + \sigma_u^2 \text{ block diag}_j(J_{n_{ij}}) + \sigma_v^2 J_{n_i} \\ &= \text{block diag}_j(\sigma_i^2 I_{n_{ij}} + \sigma_u^2 J_{n_{ij}}) + \sigma_v^2 J_{n_i}. \end{aligned}$$

with $J_m = 1_m 1_m'$.

Our principle interest is the estimation of σ_v^2, σ_u^2 and the regression parameters β. For simplicity, $\mathcal{X}$ is assumed to be of full column rank (and so is the matrix X defined below). If σ_v^2, σ_u^2 and σ_i^2's are known, the following weighted least squares (WLS) estimator is usually used for estimating regression coefficient β:

$$\tilde{\beta}_w = (\mathcal{X}'V^{-1}\mathcal{X})^{-1}\mathcal{X}V^{-1}\mathcal{Y}, \tag{5.3}$$

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

$$V_i^{-1} = C_i - w_i d_i d_i' / \Big(\sum_{j=1}^{m_i} w_{ij}\Big). \tag{5.4}$$

where

$$\begin{aligned} C_i^{-1} &= \text{block diag}_j\Big(\sigma_i^{-2}(I_{n_{ij}} - \frac{(\sigma_u^2/\sigma_i^2)J_{n_{ij}}}{1 + n_{ij}\sigma_u^2/\sigma_i^2})\Big), \\ d_i &= col_{1\le j\le m_i}\left(n_{ij}^{-1} w_{ij} 1_{n_{ij}}\right), \\ w_{ij}^{-1} &= \sigma_u^2 + \sigma_i^2/n_{ij}, \\ w_i &= \frac{\sum_j w_{ij}}{1 + \sigma_v^2 \sum_j w_{ij}}. \end{aligned}$$

Using (5.4) in (5.3), $\tilde{\beta}_w$ can be written as

$$\tilde{\beta}_w = \Big(\sum_i w_i x_i x_i'\Big)^{-1} \sum_i w_i x_i \bar{y}_{i..} \tag{5.5}$$

where

$$\bar{y}_{ij\cdot} = \sum_{k=1}^{n_{ij}} y_{ijk}/n_i \tag{5.6}$$
$$\bar{y}_{i\cdot\cdot} = \sum_{j=1}^{m_i} w_{ij}\bar{y}_{ij\cdot}/\sum_{j=1}^{m_i} w_{ij}.$$

Since σ_v^2, σ_u^2, and σ_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 σ_u^2 and common σ_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 σ_v^2 and σ_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 σ_v^2, σ_u^2 and σ_i^2 are given for a special set of prior values for these variances. The estimators of σ_v^2 and σ_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

$$\mathcal{Y} = \mathcal{X}\beta + U_v\xi_v + U_u\xi_u + U_1\xi_1 + \cdots + U_m\xi_m$$

where

$$U_i' = (0_{n_{i\cdot}\times\sum_{i'=1}^{i-1} n_{i'}} : I_{n_i} : 0_{n_{i\cdot}\times\sum_{i'=i+1}^{n} n_{i'}})_{n_{i\cdot}\times n},$$
$$U_v = \text{block diag}_i(1_{n_{i\cdot}}), \quad U_u = \text{block diag}_i(\text{block diag}_j(1_{n_{ij}})),$$

ξ_i is an $n_{i\cdot}$-random vector with mean 0 and dispersion matrix $\sigma_i^2 I_{n_i}$, ξ_v is an m-random vector with mean 0 and dispersion matrix $\sigma_v^2 I_m$, and ξ_u is an $\sum_{i=1}^m m_i \stackrel{\text{def}}{=} m_{\cdot}$-random vector with mean 0 and dispersion matrix $\sigma_u^2 I_{m_+}$.

Let σ_{v0}^2, σ_{u0}^2 and σ_{i0}^2 be prior values of σ_v^2, σ_u^2 and σ_i^2 respectively. And denote

$$\begin{aligned}
&w_{ij0} = (\sigma_{v0}^2 + \sigma_{i0}^2/n_{ij})^{-1}, \quad w_{i\cdot 0} = \sum_{j=1}^{m_i} w_{ij0}, \\
&w_{i0} = w_{i\cdot 0}/(1 + \sigma_{v0}^2 w_{i\cdot 0}), \\
&\bar{y}_{i\cdot\cdot 0} = \sum_{j=1}^{m_i} w_{ij0}\bar{y}_{ij\cdot}/w_{i\cdot 0}, \qquad (5.7)\\
&V_0 = \sigma_{10}^2 F_1 + \cdots + \sigma_{m0}^2 F_m + \sigma_{v0}^2 F_v + \sigma_{u0}^2 F_u, \\
&\hat{\beta}_0 = (\mathcal{X}'V_0^{-1}\mathcal{X})^{-1}\mathcal{X}'V_0^{-1}\mathcal{Y}, \qquad (5.8)
\end{aligned}$$

where $\bar{y}_{ij\cdot}$ 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 σ_v^2, σ_u^2 and σ_i^2 are the solutions to the following consistent linear system (referred to as MINQUE equations, C. R. Rao and Kleffe, 1988):

$$H_{UI}\hat{\theta} = q_{UI} \qquad (5.9)$$

where $H_{UI} = (h_{ij})_{(m+2)\times(m+2)}$, $\hat{\theta} = (\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2, \hat{\sigma}_u^2, \hat{\sigma}_v^2)'$, $q_{UI} = (Q_1, \ldots, Q_m, Q_u, Q_v)'$

with

$$h_{ji} = h_{ij} = \begin{cases} tr(MV_0M)^+F_i(MV_0M)^+F_j & \text{for } 1 \le i,j \le m \\ tr(MV_0M)^+F_i(MV_0M)^+F_u & \text{for } 1 \le i \le m, j = m+1 \\ tr(MV_0M)^+F_i(MV_0M)^+F_v & \text{for } 1 \le i \le m, j = m+2 \\ tr(MV_0M)^+F_u(MV_0M)^+F_u & \text{for } i = j = m+1 \\ tr(MV_0M)^+F_u(MV_0M)^+F_v & \text{for } i = m+1, j = m+2 \\ tr(MV_0M)^+F_v(MV_0M)^+F_v & \text{for } i = j = m+2\ , \end{cases}$$

and

$$\begin{aligned} Q_v &= \mathcal{Y}'(MV_0M)^+F_v(MV_0M)^+\mathcal{Y}, \\ Q_u &= \mathcal{Y}'(MV_0M)^+F_u(MV_0M)^+\mathcal{Y}, \\ Q_i &= \mathcal{Y}'(MV_0M)^+F_i(MV_0M)^+\mathcal{Y}, \\ M &= I_n - \mathcal{X}(\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}', \end{aligned}$$

where A^+ denotes the Moore–Penrose inverse of A. Using the following identity (same as (2.5))

$$(MV_0M)^+ = V_0^{-1} - V_0^{-1}X_1(X_1'V_0^{-1}X_1)^{-1}X_1'V_0^{-1}, \tag{5.10}$$

we obtain that

$$\begin{aligned} Q_i = \sigma_{i0}^{-4}\sum_j\sum_k(y_{ijk} - \bar{y}_{ij\cdot})^2 \qquad (5.11) \\ + \sum_j n_{ij}^{-1}w_{ij0}^2[\bar{y}_{ij\cdot} - x_i'\hat{\beta}_0 - \sigma_{v0}^2 w_{i0}(\bar{y}_{i\cdot\cdot 0} - x_i'\hat{\beta}_0)]^2, \end{aligned}$$

$$Q_u = \sum_i\sum_j w_{ij0}^2[\bar{y}_{ij\cdot} - x_i'\hat{\beta}_0 - \sigma_{v0}^2 w_{i0}(\bar{y}_{i\cdot\cdot 0} - x_i'\hat{\beta}_0)]^2, \tag{5.12}$$

$$Q_v = \sum_i[w_{i0}(\bar{y}_{i\cdot\cdot 0} - x_i'\hat{\beta}_0)]^2. \tag{5.13}$$

For general design points, x_i's, explicit expressions for the elements of the matrix H_{UI} 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

$$\hat{\sigma}_v^2 = (\sigma_{v0}^4/m)Q_v,$$
$$\hat{\sigma}_u^2 = \Big(\sigma_{u0}^4 \Big/ \sum_{i=1}^{m} m_i\Big)Q_u,$$
$$\hat{\sigma}_i^2 = (\sigma_{i0}^4/n_{i\cdot})Q_i.$$

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 σ_{v0}^2, σ_{u0}^2 and σ_{i0}^2:

$$R_0^2 = (\mathcal{Y} - \mathcal{X}\hat{\beta}_0)'V_0^{-1}(\mathcal{Y} - \mathcal{X}\hat{\beta}_0).$$

Denote

$$S_i = (n_{i\cdot} - m_i)s_i^2 = \sum_{j=1}^{m_i}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{ij\cdot})^2, \tag{5.14}$$

$$S_u = \sum_{i=1}^{m}\sum_{j=1}^{m_i} w_{ij0}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot 0})^2, \tag{5.15}$$

$$S_v = \sum_{i=1}^{m} w_{i0}(\bar{y}_{i\cdot\cdot 0} - x_i'\hat{\beta}_0)^2. \tag{5.16}$$

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 σ_i^2, σ_u^2 and σ_v^2 can be obtained from the following expressions:

$$\hat{\sigma}_i^2 = s_i^2 = S_i/(n_{i\cdot} - m_i), \tag{5.17}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{m_i} B_{ij}\hat{\sigma}_u^2 = Q_i - \sum_{i=1}^{m}\hat{\sigma}_i^2\sum_{j=1}^{m_i} B_{ij}/n_{ij}, \tag{5.18}$$

$$\sum_{i=1}^{m} A_i\hat{\sigma}_v^2 = Q_v^2 - \hat{\sigma}_u^2\sum_{i=1}^{m} A_i w_{i\cdot 0}^{-2}\sum_{j=1}^{m_i} w_{ij0}^2 - \sum_{i=1}^{m} A_i\hat{\sigma}_i^2 w_{i\cdot 0}^{-2}\sum_{j=1}^{m_i} w_{ij0}^2/n_{ij}, \tag{5.19}$$

where

$$B_{ij} = w_{ij0} - w_{i\cdot 0}^{-1}w_{ij0}^2,$$
$$A_i = w_{i0}\Big(1 - w_{i0}x_i'\Big(\sum_{i=1} w_{i0}x_i x_i'\Big)^{-1}x_i\Big).$$

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 $\sigma_{v0}^2 = \sigma_{u0}^2 = 0$ and $\sigma_{i0}^2 = 1$. With these prior values, $\bar{y}_{ij\cdot}$, $\bar{y}_{i\cdot\cdot 0}$ and $\hat{\beta}_0$ are obtained, and then S_i, S_u and S_v are computed from (5.14)–(5.16). Then the ANOVA estimators of σ_i^2, σ_u^2 and σ_v^2 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 σ_{v0}^2, σ_{u0}^2 and σ_{i0}^2's, we have

$$(MV_0M)^+ = V_0^{-1} + O(m^{-1})$$

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

$$h^*_{ji} = h^*_{ij} = \begin{cases} tr(V_0^{-1}F_iV_0^{-1}F_j) & \text{for } 1 \le i,j \le m, \\ tr(V_0^{-1}F_iV_0^{-1}F_u) & \text{for } 1 \le i \le m, j = m+1, \\ tr(V_0^{-1}F_iV_0^{-1}F_v) & \text{for } 1 \le i \le m, j = m+2, \\ tr(V_0^{-1}F_uV_0^{-1}F_u) & \text{for } i = j = m+1, \\ tr(V_0^{-1}F_uV_0^{-1}F_v) & \text{for } i = m+1, j = m+2, \\ tr(V_0^{-1}F_vV_0^{-1}F_v & \text{for } i = j = m+2. \end{cases}$$

It can be shown that

$$\begin{aligned}
tr(V_0^{-1}F_iV_0^{-1}F_j) &= 0 \quad \text{for } i \ne j, \\
tr(V_0^{-1}F_iV_0^{-1}F_i) &= \sum_{j=1}^{m_i} n_{ij}^{-1}w_{ij0}^2\Big[1 + 2(n_{ij}-1)\sigma_{u0}^2/\sigma_{i0}^2 + n_{ij}(n_{ij}-1)(\sigma_{u0}^2/\sigma_{i0}^2)^2\Big] \\
&\quad - 2\sigma_{v0}^2\textstyle\sum_{j=1}^{m_i} n_{ij}^{-2}w_{ij0}^3/(1+\sigma_{v0}^2w_{i\cdot 0}) + \sigma_{v0}^4\Big(\sum_j n_{ij}^{-1}w_{ij0}^2/(1+\sigma_{v0}^2w_{i\cdot 0})\Big)^2, \\
tr(V_0^{-1}F_iV_0^{-1}F_u) &= \textstyle\sum_j n_{ij}^{-1}w_{ij0}^2 - 2\sigma_{v0}^2\sum_j n_{ij}^{-1}w_{ij0}^3/(1+\sigma_{v0}^2w_{i\cdot 0}) \\
&\quad + \sigma_{v0}^4\textstyle\sum_j w_{ij0}^2\sum_j n_{ij}^{-1}w_{ij0}^2/(1+\sigma_{v0}^2w_{i\cdot 0})^2, \\
tr(V_0^{-1}F_iV_0^{-1}F_v) &= \textstyle\sum_j n_{ij}^{-1}w_{ij0}^2 - 2\sigma_{v0}w_{i0}\sum_j n_{ij}^{-1}w_{ij0}^2 + \sigma_{v0}^4w_{i0}^2\sum_j n_{ij}^{-1}w_{ij0}^2, \\
tr(V_0^{-1}F_uV_0^{-1}F_v) &= \sum_i\Big[\textstyle\sum_j w_{ij0}^2 - \sigma_{v0}^2w_{i0}\sum_j w_{ij0}^2\Big]/(1+\sigma_{v0}^2w_{i\cdot 0}), \\
tr(V_0^{-1}F_uV_0^{-1}F_u) &= \sum_i\Big[\textstyle\sum_j w_{ij0}^2 - 2\sigma_{v0}^2\sum_j w_{ij0}^3/(1+\sigma_{v0}^2w_{i\cdot 0}) \\
&\quad + \sigma_{v0}^4\Big(\textstyle\sum_j w_{ij0}^2/(1+\sigma_{v0}^2w_{i\cdot 0})\Big)^2\Big], \\
tr(V_0^{-1}F_vV_0^{-1}F_v) &= \sum_i w_{i0}^2.
\end{aligned}$$

The above expressions are still quite complicated for general prior values of σ_v^2, σ_u^2 and σ_i^2, so an explicit expression for H^*_{UI} is not available. To avoid this problem, we use $\sigma_{v0}^2 = \sigma_{u0}^2 = 0$ and $\sigma_{i0}^2 = 1$, like the prior values corresponding to the ANOVA procedure. In this case, $w_{ij0} = n_{ij}$, and $w_{i0} = n_{i\cdot}$, and Q_i, Q_u and Q_v in (5.11)–(5.13) can be rewritten as

$$Q_i = \sum_{j=1}^{m_i}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{ij\cdot})^2 + \sum_{j=1}^{m_i} n_{ij}(\bar{y}_{ij\cdot} - x_i'\hat{\beta}_0)^2, \tag{5.20}$$

$$Q_u = \sum_{i=1}^{m}\sum_{j=1}^{m_i} n_{ij}^2(\bar{y}_{ij\cdot} - x_i'\hat{\beta}_0)^2\,, \tag{5.21}$$

$$Q_v = \sum_{i=1}^{m} n_{i\cdot}^2(\bar{y}_{i\cdot\cdot 0} - x_i'\hat{\beta}_0)^2, \tag{5.22}$$

where $\bar{y}_{ij\cdot}$ is the same as that in (5.6), and $\hat{\beta}_0$ is the ordinary least square estimator of β from model (5.1). Further after computing the entries in H_{UI}^*, we find that

$$H_{UI}^* = \begin{pmatrix} (\mathrm{diag}_i(n_{i\cdot}))_{(m\times m)} & \mu & \mu \\ \mu' & \sum_i\sum_j n_{ij}^2 & \sum_i\sum_j n_{ij}^2 \\ \mu' & \sum_i\sum_j n_{ij}^2 & \sum_i(\sum_j n_{ij})^2 \end{pmatrix}$$

with

$$\mu = (n_{1\cdot}, \ldots, n_{m\cdot})'\,.$$

Applying Lemma 2.1 twice, we can find the inverse of H_{UI}^* as

$$(H_{UI}^*)^{-1} = \begin{pmatrix} (\mathrm{diag}_i(n_{i\cdot}^{-1}) + N_1^{-1}1_m1_m')_{(m\times m)} & -N_1^{-1}1_m & 0 \\ -N_1^{-1}1_m' & N_1^{-1} + N_2^{-1} & N_2^{-1} \\ 0 & N_2^{-1} & N_2^{-1} \end{pmatrix}$$

with

$$N_1 = \sum_{i=1}^{m}\sum_{j=1}^{m_i}(n_{ij}^2 - n_{ij}), \quad N_2 = \sum_{i=1}^{m} n_{i\cdot}^2 - \sum_{i=1}^{m}\sum_{j=1}^{m_i} n_{ij}{}^2\,.$$

The AMINQUE estimators of σ_i^2, σ_u^2 and σ_v^2 can be obtained from the linear system

$$H_{UI}^*\tilde{\theta} = q_{UI} \quad \text{or} \quad \tilde{\theta} = (H_{UI}^*)^{-1}q_{UI}\,,$$

where

$$\tilde{\theta} = (\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_m^2, \tilde{\sigma}_u^2, \tilde{\sigma}_v^2)'\,,$$

which gives

$$\tilde{\sigma}_i^2 = n_{i\cdot}^{-1}Q_i - N_1^{-1}(Q_u - \textstyle\sum_i Q_i)\,,$$

$$\tilde{\sigma}_u^2 = N_1^{-1}(Q_u - \textstyle\sum_i Q_i) - N_2^{-1}(Q_v - Q_u)\,, \tag{5.23}$$

$$\tilde{\sigma}_v^2 = N_2^{-1}(Q_v - Q_u)\,. \tag{5.24}$$

Note that both $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$ may assume negative values. Non-negative AMINQUE estimators of σ_v^2 and σ_u^2 are obtained by truncating them at zero, viz.

$$\hat{\sigma}_v^2 = \max(0, \tilde{\sigma}_v^2), \quad \text{and} \quad \hat{\sigma}_u^2 = \max(0, \tilde{\sigma}_u^2).$$

5.2.3 Asymptotic properties of AMINQU estimators

In this section, the asymptotic normality of estimators $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ in (5.23) and (5.24) are established. First some regularity conditions are imposed on model (5.1):

(A1) There are positive constants σ_0^2, σ_∞^2, c_∞, and positive integers n_∞, m_∞ such that $\sigma_0^2 \le \sigma_i^2 \le \sigma_\infty^2$, $2 \le n_{ij} \le n_\infty$, $2 \le m_i \le m_\infty$ and $\|x_i\| \le c_\infty$ for all i.

(B1) There is a positive constant c_0 so that

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

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

$$E\ |v_i|^{4+2\delta} \le b, \quad E\ |u_{ij}|^{4+2\delta} \le b \ \text{ and } \ E|e_{ijk}|^{2+\delta} \le b \ \text{ for all } i,\ j,\ k \tag{5.25}$$

for some positive constant δ.

Assumption (B1) together with the boundedness of $\|x_i\|$ implies that

$$\lambda_0 I_p \le I^{-1}X'X \le \lambda_\infty I_p \quad \text{for all } m \tag{5.26}$$

for some positive constants λ_0, λ_∞.

The symmetry of the distributions in condition (C1) implies that $Ev_i = 0$, $Ev_i^3 = 0$, $Eu_i = 0$, $Eu_i^3 = 0$, $Ee_{ij} = 0$, and

$$E\frac{\bar{\varepsilon}_{ij.}}{[b_1 + b_2\sum_j\sum_k(e_{ijk} - \bar{e}_{ij.})^2 + b_3\bar{\varepsilon}_{ij.}^2]^r} = 0 \tag{5.27}$$

with

$$\bar{\varepsilon}_{ij.} = v_i + u_{ij} + \bar{e}_{ij.}, \quad \bar{e}_{ij.} = \sum_{k=1}^{n_{ij}} e_{ijk}/n_{ij},$$

for any positive constants b_1, b_2, b_3, and positive integer r.

By the independence between v_i, u_{ij}, e_{ijk}, and from (A1)

$$var(\bar{\epsilon}_{ij.}) = \sigma_v^2 + \sigma_u^2 + \sigma_i^2/n_{ij} \leq C$$

for some positive constant C, and from (5.25), for the δ given in condition (C1)

$$E|\bar{\epsilon}_{ij.}|^{2+\delta} \leq C_1 \tag{5.28}$$

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 $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$ first. By substituting (5.20)-(5.22) into (5.23) and (5.24), $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$ can be rewritten as

$$\tilde{\sigma}_v^2 = N_2^{-1} \sum_i Z_{2i} - 2R_4 + R_5, \tag{5.29}$$

$$\tilde{\sigma}_u^2 = N_1^{-1} \sum_i Z_{3i} - \tilde{\sigma}_v^2 - 2R_6 + R_7, \tag{5.30}$$

where

$$Z_{2i} = \sum\sum_{j \neq j'} n_{ij} n_{ij_1} \Big[v_i^2 + 2v_i(u_{ij} + \bar{e}_{ij.}) + (u_{ij} + \bar{e}_{ij.})(u_{ij'} + \bar{e}_{ij'.}) \Big],$$

$$Z_{3i} = \sum_j (n_{ij}^2 - n_{ij}) \Big[v_i^2 + u_{ij}^2 + 2v_i u_{ij} + 2(v_i + u_{ij})\bar{e}_{ij.}$$

$$+ \sum\sum_{k \neq k'} e_{ijk} e_{ijk'} / (n_{ij}^2 - n_{ij}) \Big],$$

$$R_4 = N_2^{-1} \sum_i \sum_j n_{ij}(n_{i.} - n_{ij}) \bar{\epsilon}_{ij.} (\hat{\beta}_0 - \beta)' x_i,$$

$$R_5 = N_2^{-1} \sum_i \Big(n_{i.}^2 - \sum_j n_{ij}^2 \Big) \Big((\hat{\beta}_0 - \beta)' x_i \Big)^2,$$

$$R_6 = N_1^{-1} \sum_i \sum_j (n_{ij}^2 - n_{ij}) \bar{\epsilon}_{ij.} (\hat{\beta}_0 - \beta)' x_i,$$

$$R_7 = N_1^{-1} \sum_i \sum_j (n_{ij}^2 - n_{ij}) \left((\hat{\beta}_0 - \beta)' x_i \right)^2.$$

Using the independence between v_i, u_{ij} and e_{ijk}, we get

$$E(Z_{2i}) = \sigma_v^2 \Big(n_{i.}^2 - \sum_j n_{ij}^2 \Big) = N_2 \sigma_v^2, \tag{5.31}$$

$$E(Z_{3i}) = (\sigma_v^2 + \sigma_u^2) \Big(\sum_j n_{ij}^2 - n_{i.} \Big) = N_1 (\sigma_v^2 + \sigma_v^2), \tag{5.32}$$

and

$$
\begin{aligned}
var(Z_{2i}) &= \sum\sum_{j\neq j'} n_{ij}^2 n_{ij'}^2 \Big[M_{v4} - \sigma_v^4 \\
&\qquad + 4\sigma_v^2(\sigma_u^2 + n_{ij}^{-1}\sigma_i^2) + (\sigma_u^2 + n_{ij'}^{-1}\sigma_i^2)(\sigma_u^2 + n_{ij'}^{-1}\sigma_i^2)\Big], \\
var(Z_{3i}) &= \sum_j (n_{ij}^2 - n_{ij})\Big[M_{v4} - \sigma_v^4 + M_{u4} - \sigma_u^4 + 4\sigma_v^2\sigma_u^2 \\
&\qquad + 4(\sigma_v^2 + \sigma_u^2)\sigma_i^2/n_{ij} + \sigma_i^4/(n_{ij}^2 - n_{ij})\Big], \\
cov(Z_{2i}, Z_{3i}) &= (M_{v4} - \sigma_v^4)\sum_j (n_{ij}^2 - n_{ij})\Big(n_{i.}^2 - \sum_j n_{ij}^2\Big) \\
&\qquad + 4\sigma_v^2\sigma_u^2 \sum_j n_{ij}(n_{ij}^2 - n_{ij})(n_{i.} - n_{ij}),
\end{aligned}
$$

with

$$M_{v4} = Ev_i^4 \quad \text{and} \quad M_{u4} = Eu_{ij}^4.$$

Denote

$$V_1 = N_1^{-2} \sum_i var(Z_{2i}) \tag{5.33}$$

$$V_2 = N_1^{-2} N_2^{-2} \sum_i \Big(N_1^2 var(Z_{2i}) - 2N_1 N_2 cov(Z_{2i}, Z_{3i}) + N_2^2 var(Z_{3i}) \Big) \tag{5.34}$$

Now we have the following theorem on the asymptotic normality of $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$.

Theorem 5.1 *Suppose that assumptions* (A1), (B1) *and* (5.25) *hold. Then as* $m \to +\infty$

$$V_1^{-1/2}(\tilde{\sigma}_v^2 - \sigma_v^2) \to_d N(0,1), \tag{5.35}$$

$$V_2^{-1/2}(\tilde{\sigma}_u^2 - \sigma_u^2) \to_d N(0,1). \tag{5.36}$$

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 $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$ 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*

$$\tilde{\sigma}_v^2 \to_p \sigma_v^2, \quad \tilde{\sigma}_u^2 \to_p \sigma_u^2.$$

Using arguments similar to those for Corollary 2.2, we have the following results for AMINQUE after truncation.

Corollary 5.2 *Suppose $\sigma_v^2 > 0$ and $\sigma_u^2 > 0$. Then under the conditions in Theorem 5.1,*

$$\begin{aligned}
\hat{\sigma}_v^2 &= \max(0, \tilde{\sigma}_v^2) \to_p \sigma_v^2, \\
\hat{\sigma}_v^2 &= \max(0, \tilde{\sigma}_u^2) \to_p \sigma_v^2, \\
V_1^{-1/2}(\hat{\sigma}_v^2 - \sigma_v^2) &\to_d N(0,1), \\
V_2^{-1/2}(\hat{\sigma}_u^2 - \sigma_u^2) &\to_d N(0,1).
\end{aligned}$$

Therefore the asymptotic variances of $\hat{\sigma}_v^2$ 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:

$$\tilde{V}_1 = N_2^{-2} \sum_i \Big[\sum\sum_{j \neq j'} n_{ij} n_{ij'} \Big((\bar{y}_{ij\cdot} - x_i'\hat{\beta}_0)(\bar{y}_{ij'\cdot} - x_i'\hat{\beta}_0) - \tilde{\sigma}_v^2 \Big) \Big]^2, \tag{5.37}$$

$$\begin{aligned}
\tilde{V}_2 = N_1^{-2} N_2^{-2} \sum_i \Big[& N_2 \textstyle\sum_j (n_{ij}^2 - n_{ij})(\bar{y}_{ij\cdot} - x_i'\hat{\beta}_0)^2 \\
& - N_2 \textstyle\sum_j \sum_k (y_{ijk} - \bar{y}_{ij\cdot})^2 - N_2(\tilde{\sigma}_v^2 + \tilde{\sigma}_u^2) \sum_j (n_{ij}^2 - n_{ij}) \\
& - N_1 \textstyle\sum\sum_{j \neq j'} n_{ij} n_{ij'} ((\bar{y}_{ij\cdot} - x_i'\hat{\beta}_0)(\bar{y}_{ij'\cdot} - x_i'\hat{\beta}_0) + N_1 \tilde{\sigma}_v^2 \Big(n_{i\cdot} - \textstyle\sum_j n_{ij}^2 \Big) \Big]^2.
\end{aligned} \tag{5.38}$$

These estimators are consistent as shown in the following theorem.

Theorem 5.2 *Suppose that assumptions* (A1), (B1), *and* (5.25) *hold. Then*

$$\begin{aligned}
m(\hat{V}_1 - V_1) &\to_p 0, \\
m(\hat{V}_2 - V_2) &\to_p 0.
\end{aligned}$$

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 $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$ instead of using $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$ in (5.37) and (5.38), especially when the true value of σ_v^2 and σ_u^2 are significantly away from 0 relative to values of σ_i^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 $\sigma_u^2+n_{ij}^{-1}\sigma_i^2$ using AMINQU estimators of σ_u^2 and σ_i^2. Therefore we want to estimate the regression vector β using the approach in Section 3.4.

We proved the consistency and asymptotic normality of $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$ in section 5.2.3. The within group sample variance $s_i^2 = (\sum_{j=1}^{m_i}(n_{ij}-1))^{-1}\sum_{j=1}^{m_i}\sum_{k=1}^{n_{ij}}(y_{ijk}-\bar{y}_{ij\cdot})^2$ is an unbiased estimator of σ_i^2. From (5.5), a weighted least square estimator $\hat{\beta}_w$ of β can be obtained from

$$\hat{\beta}_w = (X'WX)^{-1}\sum_{i=1}^{m}\hat{w}_i x_i \hat{\bar{y}}_{i\cdot\cdot}. \tag{5.39}$$

where

$$\begin{aligned}
W &= \mathrm{diag}_i(\hat{w}_i),\\
\hat{w}_{ij} &= (\hat{\sigma}_u^2 + n_{ij}^{-1}s_i^2)^{-1},\\
\hat{w}_i &= \sum_{j=1}^{m_i}\hat{w}_{ij}\Big/\Big(1+\hat{\sigma}_v^2\sum_{j=1}^{m_i}\hat{w}_{ij}\Big),\\
\hat{\bar{y}}_{i\cdot\cdot} &= \sum_{j=1}^{m_i}\hat{w}_{ij}\bar{y}_{ij\cdot}\Big/\Big(\sum_{j=1}^{m_i}\hat{w}_{ij}\Big),
\end{aligned}$$

where $\bar{y}_{ij\cdot}$ 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

$$\begin{aligned}
&\hat{w}_{ij}^{(1)} = (\sigma_u^2 + n_{ij}^{-1}s_i^2)^{-1}, \quad \hat{w}_i^{(1)} = \Big(1+\sigma_v^2\sum_{j=1}^{m_i}\hat{w}_{ij}^{(1)}\Big)^{-1}\sum_{j=1}^{m_i}\hat{w}_{ij}^{(1)}\\
&\tau_{6i} = \sum\nolimits_j E\hat{w}_i^{(1)},\\
&\tau_{7i} = E[\hat{w}_i^{(1)}\bar{\varepsilon}_{i\cdot\cdot}^{(1)}]^2, \quad \text{where} \quad \bar{\varepsilon}_{i\cdot\cdot}^{(1)} = \sum_j w_{ij}^{(1)}\bar{\varepsilon}_{ij\cdot}\Big/\sum_j \hat{w}_{ij}^{(1)},\\
&D_6 = \mathrm{diag}_i(\tau_{6i}),\\
&D_7 = \mathrm{diag}_i(\tau_{7i}).
\end{aligned}$$

It is easy to check that when $\sigma_u^2 > 0$, both τ_{6i} and τ_{7i} are finite.

We have the following theorem on the asymptotic normality of $\hat{\beta}_w$.

Theorem 5.3 *Suppose that assumptions* (A1), (B1) *and* (C1) *hold. Then*

$$V_\beta^{-1/2}(\hat{\beta}_w - \beta) \to_p N(0, I_p) \quad as \quad m \to +\infty$$

where $V_\beta^{-1/2} = (V_\beta^{1/2})^{-1}$ *and* $V_\beta^{1/2}$ *is a square root of*

$$V_\beta = (X'D_6X)^{-1}X'D_7X(X'D_6X)^{-1}.$$

(*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

$$\sum_i \hat{w}_i x_i x_i' = \sum_i \tau_{6i} x_i x_i' + o_p(m) = (X'D_6X) + o_p(m), \tag{5.40}$$

$$\sum_i x_i \hat{w}_i \sum_j \hat{w}_{ij}\bar{\varepsilon}_{ij\cdot} \Big/ \sum_j \hat{w}_{ij} = \sum_i x_i \hat{w}_i^{(1)} \hat{\varepsilon}_{i\cdot\cdot}^{(1)} + o_p(m^{1/2})$$
$$= X'col_i(\hat{w}_i^{(1)}\hat{\varepsilon}_{i\cdot\cdot}^{(1)}) + o_p(m^{1/2}). \tag{5.41}$$

From condition (C1),

$$Ecol_i(\hat{w}_i^{(1)}\hat{\varepsilon}_{i\cdot\cdot}^{(1)}) = 0,$$

and

$$var[X'col_i(\hat{w}_i^{(1)}\hat{\varepsilon}_{i\cdot\cdot}^{(1)})] = X'D_7X.$$

Along the lines of the proof of (3.7), we can show that as $m \to +\infty$

$$(X'D_7X)^{-1/2}X'col_i(\hat{w}_i^{(1)}\hat{\varepsilon}_{i\cdot\cdot}^{(1)}) \to_d N(0, I_p). \tag{5.42}$$

Hence,

$$(X'D_7X)^{-1/2}\sum_i x_i\hat{w}_i \sum_j \hat{w}_{ij}\bar{\varepsilon}_{ij\cdot} \Big/ \sum_j \hat{w}_{ij} \to_d N(0, I_p) \quad \text{as} \quad m \to +\infty.$$

Then from (5.40) we can prove that

$$(X'D_7X)^{-1/2}(X'D_6X)(\hat{\beta}_w - \beta) \to_p N(0, I_p) \quad \text{as} \quad m \to +\infty,$$

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. □.

An estimator of V_3 can be obtained by the substitution method as follows:

$$\hat{V}_3 = (X'WX)^{-1}X'\hat{D}_7X(X'WX)^{-1},$$

where

$$D_7 = \mathrm{diag}_i(\hat{w}_i^2(\hat{\bar{y}}_{i..} - x_i'\hat{\beta}_0)^2).$$

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 θ_{it}, the i-th small area mean for the period t ($i = 1, \ldots, k; t = 1, \ldots, T$). We assume that

$$\begin{aligned} y_{it} &= \theta_{it} + e_{it} = x_{it}'\beta + v_i + u_{it} + e_{it}, \\ u_{it} &= \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1, \quad \text{for } t > 1, \text{ and } u_{i1} = \varepsilon_{i1}/\sqrt{1-\rho^2}, \end{aligned} \tag{6.1}$$

where $x_{it} = (x_{it1}, x_{it2}, \ldots, x_{itp})'$, are vectors of fixed concomitant variables related to θ_{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 σ_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 ρ over time period $t = 1, \ldots, T$; ε_{it} are random errors associated with u_{it} and $\varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$; $e_i = (e_{i1}, e_{i2}, \ldots, e_{iT})'$) is the vector of sampling errors for area i, following a multivariate normal distribution with mean 0, and known covariance matrix Σ_i ($e_i \overset{iid}{\sim} N(0, \Sigma_i)$); $\{v_i\}$, $\{\varepsilon_{it}\}$ and $\{e_i\}$ are mutually independent. We are interested in estimating the small area means θ_{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 $y = (y_{11}, \ldots, y_{1T}; \ldots; y_{k1}, \ldots, y_{kT})' =$

$(y_1', \ldots, y_k')'$, the proposed model (6.1) may be written in matrix form as

$$y = X\beta + Zv + u + e \tag{6.2}$$

where

$$\begin{aligned}
X &= (X_1', \ldots, X_k')', \qquad X_i' = (x_{i1}, \ldots, x_{iT}),\\
Z &= I_k \otimes 1_T,\\
v &= (v_1, \ldots, v_k)',\\
u &= (u_1', \ldots, u_k')', \qquad u_i' = (u_{i1}, \ldots, u_{iT})\\
e &= (e_1', \ldots, e_k')', \qquad e_i' = (e_{i1}, \ldots, e_{iT})
\end{aligned}$$

with I_k denoting the identity matrix of order k, 1_T denoting a T-vector of 1's, and $\otimes$ denoting the direct Kroneckor product. Further,

$$\begin{aligned}
E(v) &= 0, \quad Cov(v) = \sigma_v^2 I_k,\\
E(u) &= 0, \quad Cov(u) = \sigma^2 I_k \otimes \Gamma = \sigma^2 R,\\
E(e) &= 0, \quad Cov(e) = \Sigma = \text{block diag}(\Sigma_1, \ldots, \Sigma_k),
\end{aligned} \tag{6.3}$$

and v, u and e are mutually independent, where Γ is a $T \times T$ matrix with elements $\rho^{|i-j|}/(1-\rho^2)$. It follows from (6.2) and (6.3) that

$$\begin{aligned}
Cov(y) &= V = \Sigma + \sigma^2 R + \sigma_v^2 ZZ'\\
&= \text{block diag}_i(\Sigma_i + \sigma^2\Gamma + \sigma_v^2 J_T) = \text{block diag}_i(V_i) \quad \text{say},
\end{aligned}$$

with $J_T = 1_T 1_T'$.

In this chapter, we follow the standard EBLUP approach by first deriving the BLUP estimator of θ_{iT} in Section 6.1. In Section 6.2, assuming known autocorrelation ρ, we give "method of moments" type estimators of variance components σ_v^2 and σ^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 ρ is considered in Section 6.4. We provide an estimator of ρ. 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 ρ 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 i–th small area $\theta_{iT} = x'_{iT}\beta + v_i + u_{iT}$ is a special case of the linear combination $\tau = l'\beta + l'_1 v + l'_2 u$ 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 θ_{iT} can be obtained from Henderson's (1950) general results for arbitrary linear combinations τ.

Assuming first that σ_u^2, σ_v^2 and ρ are known, the BLUP estimator of τ is given by

$$\tilde{\tau} = l'\tilde{\beta} + (\sigma_v^2 l'_1 Z' + \sigma^2 l'_2 R)V^{-1}(y - X\tilde{\beta}), \tag{6.4}$$

where $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ is the weighted least squares estimator of β. Using the special structures of l_1, l_2, Z, R and V, it is easily seen that (6.4) reduces to

$$\begin{aligned}\tilde{\theta}_{iT} &= \varphi(\sigma^2, \sigma_v^2, \rho, y)\\ &= x'_{iT}\tilde{\beta} + (\sigma_v^2 1_T + \sigma^2\gamma_T)'(\Sigma_i + \sigma^2\Gamma + \sigma_v^2 J_T)^{-1}(y_i - X_i\tilde{\beta}),\end{aligned} \tag{6.5}$$

where γ_T is the T-th row of Γ. 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, $y_{it} - x'_{it}\tilde{\beta}$, $t = 1, \ldots, T-1$:

$$\tilde{\theta}_{iT} = w^*_{iT} y_{iT} + (1 - w^*_{iT})x'_{iT}\tilde{\beta} + \sum_{t=1}^{T-1} w^*_{it}(y_{it} - x'_{it}\tilde{\beta}),$$

where

$$(w_{i1}^*, \ldots, w_{iT}^*) = (\sigma_v^2 1_T + \sigma^2 \gamma_T)' V_i^{-1} .$$

6.2 Two–Stage Estimator: ρ Known

In practice, the parameters, σ^2, σ_v^2 and ρ are usually unknown. We first assume that ρ is known and replace σ^2 and σ_v^2 in (6.5) by their consistent estimators $\hat{\sigma}^2(\rho)$, $\hat{\sigma}_v^2(\rho)$ to obtain a two-stage estimator $\hat{\theta}_{iT}(\rho)$. The case of unknown ρ is studied in Section 6.4.

6.2.1 Estimation of σ^2

Pantula and Pollock (1985) estimated σ^2, σ_v^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 ρ is known.

We first obtain an unbiased estimator of σ^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

$$z_i = PX_i\beta + fv_i + P(u_i + e_i), \quad i = 1, \ldots, k. \tag{6.6}$$

where $f = (f_1, \ldots, f_T)'$ with $f_1 = (1-\rho^2)^{1/2}$ and $f_t = 1-\rho$ for $2 \le t \le T$. 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:

$$z_i^{(1)} = H_i^{(1)}\beta + e_i^*, \quad i = 1, \ldots, k, \tag{6.7}$$

where $H_i^{(1)} = (I_T - D)PX_i$ and $e_i^* = (I_T - D)P(u_i + e_i)$.

Since

$$Cov(e_i^*) = (I_T - D)(\sigma^2 I_T + P\Sigma_i P')(I_T - D)' \tag{6.8}$$

does not involve σ_v^2, we can estimate σ^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 σ^2 is then given by

$$\begin{aligned}\tilde{\sigma}^2(\rho) = &[\hat{e}'\hat{e} - tr\{(\text{block diag}_i(I_T - D) - H^{(1)}(H^{(1)\prime}H^{(1)})^- H^{(1)\prime}) \cdot \\ &\text{block diag}_i(P\Sigma_i P')\}][k(T-1) - rank(H^{(1)})]^{-1}\end{aligned} \tag{6.9}$$

where A^- is a generalized inverse of A. The unbiasedness of $\tilde{\sigma}^2(\rho)$ follows by noting that $\hat{e}'\hat{e} = e^{*\prime}(I_{kT} - H^{(1)}(H^{(1)\prime}H^{(1)})^- H^{(1)\prime})e^*$, 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 Ω, then $E(y'Gy) = tr(G\Omega)$ for any symmetric matrix G.*

Since $\tilde{\sigma}^2(\rho)$ can take negative value, we truncate it at zero and use

$$\hat{\sigma}^2 = \max\{0, \tilde{\sigma}^2(\rho)\}. \tag{6.10}$$

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 σ_v^2

Turning to the estimation of σ_v^2, we transform (6.6) by changing z_i to $c^{-1/2}f'z_i$ such that $u_i^* = c^{-1/2}f'Pu_i$ has mean 0 and variance σ^2. The transformed model is given by

$$c^{-1/2}f'z_i = c^{-1/2}f'PX_i\beta + c^{1/2}v_i + u_i^* + c^{-1/2}f'Pe_i$$

with error variance $c\sigma_v^2 + \sigma^2 + c^{-1}f'P\Sigma_i P'f$. 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 $\tilde{\sigma}_v^2$ is then given by

$$\tilde{\sigma}_v^2(\rho) = c^{-1}[k - rank(F)]^{-1}[\hat{u}'\hat{u} - tr\{(I_k - F(F'F)^- F')\mathrm{diag}_i(c^{-1} f' P\Sigma_i P' f)\}] - c^{-1}\tilde{\sigma}^2(\rho), \quad (6.11)$$

where $F = (X_1' P' f, \ldots, X_k' P' f)'$. The unbiasedness of $\tilde{\sigma}_v^2(\rho)$ follows by noting that

$$E(\hat{u}'\hat{u}) = (c\sigma_v^2 + \sigma^2)(k - rank(F)) + tr[(I_k - F(F'F)^- F')\mathrm{diag}_i(c^{-1} f' P\Sigma_i P' f)].$$

Also, noting that $\tilde{\sigma}_v^2(\rho)$ can take negative value, we truncate it at zero and use

$$\hat{\sigma}_v^2 = \max\{0, \tilde{\sigma}_v^2(\rho)\}. \quad (6.12)$$

The truncated estimator $\hat{\sigma}_v^2(\rho)$ is not unbiased, but it is also asymptotically consistent as $k \to \infty$.

Remark. The consistency of $\hat{\sigma}^2$ and $\hat{\sigma}_v^2$, 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 θ_{iT} is now obtained from (6.5) by substituting $\hat{\sigma}^2(\rho)$ and $\hat{\sigma}_v^2(\rho)$ for σ^2 and σ_v^2 respectively:

$$\hat{\theta}_{iT}(\rho) = \varphi[\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho, y]. \quad (6.13)$$

The two-stage estimator, $\hat{\theta}_{iT}(\rho)$, remains unbiased, noting that $\hat{\sigma}^2(\rho)$ and $\hat{\sigma}_v^2(\rho)$ 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

$$MSE[\hat{\theta}_{iT}(\rho)] = MSE(\tilde{\theta}_{iT}) + E[\tilde{\theta}_{iT} - \hat{\theta}_{iT}(\rho)]^2, \tag{6.14}$$

where $MSE(\tilde{\theta}_{iT}) = E(\tilde{\theta}_{iT} - \theta_{iT})^2$. Further, using Henderson's (1975) general result, an exact expression for $MSE(\tilde{\theta}_{iT})$ is given by

$$MSE(\tilde{\theta}_{iT}) = g_{1iT}(\sigma^2, \sigma_v^2, \rho) + g_{2iT}(\sigma^2, \sigma_v^2, \rho), \tag{6.15}$$

where

$$g_{1iT}(\sigma^2, \sigma_v^2, \rho) = \sigma_v^2 + \frac{\sigma^2}{1-\rho^2} - (\sigma_v^2 1_T + \sigma^2 \gamma_T)' V_i^{-1} (\sigma_v^2 1_T + \sigma^2 \gamma_T), \tag{6.16}$$

and

$$\begin{aligned} & g_{2iT}(\sigma^2, \sigma_v^2, \rho) \\ & \quad = [x_{iT} - X_i' V_i^{-1}(\sigma_v^2 1_T + \sigma^2 \gamma_T)]' (X' V^{-1} X)^{-1} [x_{iT} - X_i' V_i^{-1}(\sigma_v^2 1_T + \sigma^2 \gamma_T)]. \end{aligned} \tag{6.17}$$

The second term $g_{2iT}(\sigma^2, \sigma_v^2, \rho)$ in (6.15), due to estimating β, 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 $E[\tilde{\theta}_{iT} - \hat{\theta}_{iT}(\rho)]^2$ in (6.14). Following Kackar and Harville (1984), we propose a Taylor approximation to this term:

$$E[\tilde{\theta}_{iT} - \hat{\theta}_{iT}(\rho)]^2 \approx E[(\partial \tilde{\theta}_{iT}/\partial \sigma^2)(\tilde{\sigma}^2(\rho) - \sigma^2) + (\partial \tilde{\theta}_{iT}/\partial \sigma_v^2)(\tilde{\sigma}_v^2(\rho) - \sigma_v^2)]^2.$$

Following Prasad and Rao (1990), a further approximation is obtained as

$$E[\tilde{\theta}_{iT} - \hat{\theta}_{iT}(\rho)]^2 \approx tr[\Delta' V \Delta \Sigma^*] = g_{3iT}(\sigma^2, \sigma_v^2, \rho), \quad \text{say} \tag{6.18}$$

where Σ^* is the 2×2 covariance matrix of the unbiased estimators $\tilde{\sigma}^2(\rho)$ and $\tilde{\sigma}_v^2(\rho)$. Further $\Delta = (\partial b/\partial\sigma^2, \partial b/\partial\sigma_v^2)$ with $b' = (\sigma_v^2 1_T + \sigma^2\gamma_T)'V_i^{-1}$. 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 2×2 symmetric matrix with diagonal elements

$$a_{11} = [\gamma_T - \Gamma V_i^{-1}(\sigma_v^2 1_T + \sigma^2\gamma_T)]' V_i^{-1} [\gamma_T - \Gamma V_i^{-1}(\sigma_v^2 1_T + \sigma^2\gamma_T)],$$
$$a_{22} = [1_T - J_T V_i^{-1}(\sigma_v^2 1_T + \sigma^2\gamma_T)]' V_i^{-1} [1_T - J_T V_i^{-1}(\sigma_v^2 1_T + \sigma^2\gamma_T)],$$

and off-diagonal elements

$$a_{12} = a_{21} = [\gamma_T - \Gamma V_i^{-1}(\sigma_v^2 1_T + \sigma^2\gamma_T)]' V_i^{-1} [1_T - J_T V_i^{-1}(\sigma_v^2 1_T + \sigma^2\gamma_T)].$$

The term $g_{3iT}(\sigma^2, \sigma_v^2, \rho)$ is of the same order as $g_{2iT}(\sigma^2, \sigma_v^2, \rho)$. Combining (6.15) and (6.18), we get a second order approximation to $MSE[\hat{\theta}_{iT}(\rho)]$ as

$$MSE[\hat{\theta}_{iT}(\rho)] \approx g_{1iT}(\sigma^2, \sigma_v^2, \rho) + g_{2iT}(\sigma^2, \sigma_v^2, \rho) + g_{3iT}(\sigma^2, \sigma_v^2, \rho). \tag{6.19}$$

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 Σ^*. For this purpose we rewrite $\tilde{\sigma}^2(\rho)$ and $\tilde{\sigma}_v^2(\rho)$ as

$$\tilde{\sigma}^2(\rho) = [(k-1)T - rank(H^{(1)})]^{-1} a' C_1 a + \text{const.} \tag{6.20}$$
$$\tilde{\sigma}_v^2(\rho) = c^{-1}[k - rank(F)]^{-1} a' C_2 a$$
$$- c^{-1}[(k-1)T - rank(H^{(1)})]^{-1} a' C_1 a + \text{const.} \tag{6.21}$$

where $a = Zv + u + e \sim N(0, V)$,

$$C_1 = C'[I_n - CX(X'CC'X)^{-1}X'C']C$$

with $C = \text{block diag}_i[(I_T - D)P]$ and

$$C_2 = C^{*\prime}[I_k - C^{*\prime}X(X'C^{*\prime}C^*X)^{-1}X'C^{*\prime}]C^*$$

with $C^* = \text{block diag}_i(c_i^{-1/2} f' P)$. We can now evaluate the elements of Σ^* 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 $Cov(y'G_1y, y'G_2y) = 2tr(G_1\Omega G_2\Omega)$, 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

$$E[g_{1iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho)] \approx g_{1iT}(\sigma^2, \sigma_v^2, \rho) - g_{3iT}(\sigma^2, \sigma_v^2, \rho), \tag{6.22}$$

$$E[g_{2iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho)] \approx g_{2iT}(\sigma^2, \sigma_v^2, \rho), \tag{6.23}$$

and

$$E[g_{3iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho)] \approx g_{3iT}(\sigma^2, \sigma_v^2, \rho). \tag{6.24}$$

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 $MSE[\hat{\theta}_{iT}(\rho)]$ with expectation correct to $O(k^{-1})$ is given by

$$\text{mse}[\hat{\theta}_{iT}(\rho)] \approx g_{1iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho) + g_{2iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho) + 2g_{3iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho). \tag{6.25}$$

A naive estimator of MSE, which ignores the uncertainty in the estimators $\hat{\sigma}^2(\rho)$ and $\sigma_v^2(\rho)$ is given by

$$\text{mse}_N[\hat{\theta}_{iT}(\rho)] = g_{1iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho) + g_{2iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho). \tag{6.26}$$

The naive MSE estimator (6.26) is computationally simpler than (6.25) since it does not require the calculation of $g_{3iT}(\hat{\sigma}^2(\rho), \hat{\sigma}_v^2(\rho), \rho)$ 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: ρ Unknown

In Sections 6.2 and 6.3 we obtained a two–stage estimator of θ_{iT} and an estimator of its MSE, assuming known ρ. However, in practice ρ is seldom known. In this

section, we introduce an estimator $\hat{\rho}$ of ρ. Using $\hat{\rho}$, we obtain the estimators $\hat{\sigma}^2(\hat{\rho})$ and $\hat{\sigma}_v^2(\hat{\rho})$. Hence, an EBLUP estimator of θ_{iT} can be obtained by substituting $\hat{\sigma}^2(\hat{\rho})$, $\hat{\sigma}_v^2(\hat{\rho})$, and $\hat{\rho}$ for σ^2, σ_v^2 and ρ, 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 ρ 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 ρ

We now obtain a consistent moment estimator of ρ by taking into account the sampling errors. Let $a_{it} = v_i + u_{it} + e_{it}$. Then we have

$$E\Big[k^{-1}(T-2)^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T-2} a_{it}(a_{it} - a_{i,t+1})\Big]$$
$$= \sigma_u^2/(1+\rho) + k^{-1}(T-2)^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T-2}(\sigma_{tt}^{(i)} - \sigma_{t,t+1}^{(i)}) \tag{6.27}$$

and

$$E\Big[k^{-1}(T-2)^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T-2} a_{it}(a_{i,t+1} - a_{i,t+2})\Big]$$
$$= \sigma_u^2\rho/(1+\rho) + k^{-1}(T-2)^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T-2}(\sigma_{t,t+1}^{(i)} - \sigma_{t,t+2}^{(i)}), \tag{6.28}$$

where $\sigma_{t,t+s}^{(i)} = cov(e_{it}, e_{i,t+s})$ and $\sigma_{tt}^{(i)} = var(e_{it})$. It follows from (6.27) and (6.28) that a moment estimator of ρ, assuming known errors a_{it}, is given by

$$\rho^* = \frac{\sum_{i=1}^{k}\sum_{t=1}^{T-2}[a_{it}(a_{i,t+1} - a_{i,t+2}) - (\sigma_{t,t+1}^{(i)} - \sigma_{t,t+2}^{(i)})]}{\sum_{i=1}^{k}\sum_{t=1}^{T-2}[a_{it}(a_{i,t} - a_{i,t+1}) - (\sigma_{t,t}^{(i)} - \sigma_{t,t+1}^{(i)})]}. \tag{6.29}$$

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 $\tilde{\rho}$. Under certain regularity conditions $\tilde{\rho}$ is a consistent estimator of ρ as $k \to \infty$. Since the absolute value $|\tilde{\rho}|$ may be greater than or equal to 1, we need to truncate $\tilde{\rho}$:

$$\hat{\rho} = sign(\tilde{\rho})\ \min(1-\delta, |\tilde{\rho}|). \tag{6.30}$$

where $sign(x) = x/|x|$ 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 ρ^*.

6.4.2 Two–stage or EBLUP estimator

With the estimator $\hat{\rho}$ from (6.30), estimators of σ^2 and σ_v^2 can be obtained from (6.10) and (6.12) as

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

where $\tilde{\sigma}^2$ and $\tilde{\sigma}_v^2$ are obtained by substituting $\hat{\rho}$ for ρ in equations (6.9) and (6.11) respectively. Hence a two–stage estimator of θ_{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 $MSE[\hat{\theta}_{iT}(\hat{\rho})]$. Following Kackar and Harville (1984), we have

$$MSE[\hat{\theta}_{iT}(\hat{\rho})] = MSE[\tilde{\theta}_{iT}] + E[\hat{\theta}_{iT}(\hat{\rho}) - \tilde{\theta}_{iT}]^2, \tag{6.31}$$

where $MSE(\tilde{\theta}_{iT})$ 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

$$\begin{aligned} &E[\hat{\theta}_{iT}(\hat{\rho}) - \tilde{\theta}_{iT}]^2 \\ &\quad = E\{[\hat{\theta}_{iT}(\hat{\rho}) - \varphi(\tilde{\sigma}^2(\rho), \tilde{\sigma}_v^2(\rho), \rho, y)] + [\varphi(\tilde{\sigma}^2(\rho), \tilde{\sigma}_v^2(\rho), \rho, y) - \varphi(\sigma^2, \sigma_v^2, \rho, y)]\}^2 \end{aligned}$$

where $\varphi(\sigma^2, \sigma_v^2, \rho, y)$ is defined in (6.5). We have

$$\begin{aligned} E[\hat{\theta}_{iT}(\hat{\rho}) - \tilde{\theta}_{iT}]^2 \approx\ & E[(da(\tilde{\sigma}^2(\rho), \tilde{\sigma}_v^2(\rho), \rho)/d\rho)(\tilde{\rho} - \rho) \\ & + (\partial\varphi(\sigma^2, \sigma_v^2, \rho)/\partial\sigma^2)(\tilde{\sigma}^2(\rho) - \sigma^2) + (\partial t(\sigma^2, \sigma_v^2, \rho)/\partial\sigma_v^2)(\tilde{\sigma}_v^2(\rho) - \sigma_v^2)]^2, \end{aligned} \tag{6.32}$$

where the dependence on y is suppressed for simplicity. Note that the truncated estimators $\hat{\sigma}^2$, $\hat{\sigma}_v^2$ and $\hat{\rho}$ are replaced by their untruncated counterparts $\tilde{\sigma}^2(\rho)$, $\tilde{\sigma}_v^2(\rho)$ 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)

$$\frac{\partial\varphi(\sigma^2,\sigma_v^2,\rho)}{\partial\sigma^2} = \left(f + \frac{\partial b_1'}{\partial\sigma^2}\right)(Zv + u + e) \approx \frac{\partial b_1'}{\partial\sigma^2}(Zv + u + e) \tag{6.33}$$

since $f = O(k^{-1})$, where

$$\begin{aligned} b_1' &= (\sigma_v^2 l_1' Z' + \sigma^2 l_2' R)V^{-1}, \\ f &= (l' - b_1' X)(X'V^{-1}X)^{-1}X'\left(\frac{\partial V^{-1}}{\partial\sigma^2}\right)A - \left(\frac{\partial b_1'}{\partial\sigma^2}\right)X(X'V^{-1}X)^{-1}X'V^{-1}, \\ A &= I_{nT} - X(X'V^{-1}X)^{-1}X'V^{-1}, \end{aligned}$$

and

$$\partial b_1'/\partial\sigma^2 = [l_2' - (\sigma_v^2 l_1' Z' + \sigma_u^2 l_2' R)V^{-1}]RV^{-1}.$$

Similarly,

$$\frac{\partial\varphi(\sigma^2,\sigma_v^2,\rho)}{\partial\sigma_v^2} \approx \frac{\partial b_1'}{\partial\sigma_v^2}(Zv + u + e), \tag{6.34}$$

$$\frac{\partial\varphi(\tilde{\sigma}^2(\rho),\tilde{\sigma}_v^2(\rho),\rho)}{\partial\rho} \approx b_2'(Zv + u + e), \tag{6.35}$$

where

$$\begin{aligned} \partial b_1'/\partial\sigma_v^2 &= [l_1' - (\sigma_v^2 l_1' Z' + \sigma^2 l_2' R)V^{-1}Z]Z'V^{-1}, \\ b_2' &= \frac{\partial b_1'}{\partial\sigma_v^2}\cdot\frac{d\tilde{\sigma}_v^2(\rho)}{d\rho} + \frac{\partial b_1'}{\partial\sigma^2}\cdot\frac{d\tilde{\sigma}^2(\rho)}{d\rho} + \sigma^2[l_2' - (\sigma_v^2 l_1' Z' + \sigma^2 l_2' R)V^{-1}]\frac{dR}{d\rho}V^{-1}. \end{aligned}$$

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 $d\tilde{\sigma}_v^2(\rho)/d\rho$ involves the derivatives of matrices in (6.9) and (6.11) with respect to ρ. 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

$$E[\hat{\theta}_{iT}(\hat{\rho}) - \tilde{\theta}_{iT}]^2 \approx tr(\Delta_1' V \Delta_1 \Sigma^a), \tag{6.36}$$

where $\Delta_1 = (b_2,\ \partial b_1/\partial\sigma^2, \partial b_1/\partial\sigma_v^2)$ and Σ^a is the 3×3 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

$$MSE(\hat{\theta}_{iT}(\hat{\rho})) \approx g_{1iT}(\sigma^2, \sigma_v^2, \rho) + g_{2iT}(\sigma^2, \sigma_v^2, \rho) + g_{3iT}^a(\sigma^2, \sigma_v^2, \rho), \tag{6.37}$$

where g_{1iT}, g_{2iT} are defined in (6.16) and (6.17), and

$$g_{3iT}^a = tr(\Delta_1' V \Delta_1 \Sigma^a). \tag{6.38}$$

The neglected terms in (6.37) are of order $o(k^{-1})$.

For the evaluation of covariance matrix Σ^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

$$\hat{\rho} - \rho \approx [k(T-2)\sigma^2]^{-1}(1+\rho)a'[(\bar{G}_2 - \bar{G}_3) - \rho(\bar{G}_1 - \bar{G}_2)]a + \text{constant}, \tag{6.39}$$

where $a = Zv + u + e$, $\bar{G}_i = G_i \otimes I_k$, 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 $Var(\hat{\rho})$, $Cov(\hat{\rho}, \hat{\sigma}_v^2(\rho))$, and $Cov(\hat{\rho}, \hat{\sigma}^2(\rho))$.

Now we turn to the estimation of MSE. Following Prasad and Rao (1990), we have

$$E[g_{1iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_{1iT}(\sigma^2, \sigma_v^2, \rho) + g_{3iT}^a(\sigma^2, \sigma_v^2, \rho), \tag{6.40}$$

$$E[g_{2iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_{2iT}(\sigma^2, \sigma_v^2, \hat{\rho}), \tag{6.41}$$

$$E[g_{3iT}^a(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_{3iT}^a(\sigma^2, \sigma_v^2, \hat{\rho}). \tag{6.42}$$

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

$$mse(\hat{\theta}_{iT}) \approx g_{1iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho}) + g_{2iT}(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho}) + 2g_{3iT}^a(\hat{\sigma}^2, \hat{\sigma}_v^2, \hat{\rho}). \tag{6.43}$$

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 $\tilde{\rho}$ often took values outside the admissible range, $(-1, 1)$, especially for small T or small σ^2 relative to sampling variation.

6.4.4 Two other alternatives

Although $\tilde{\rho}$ is theoretically consistent, the probability of obtaining $\tilde{\rho}$ with $|\tilde{\rho}| > 1$ 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 $\tilde{\rho}$) is too close to 1. Since $1 - \rho^2$ appears in the denominator in the formula for $\text{MSE}(\tilde{\theta}_{iT})$, 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 ρ_0 for ρ from past experience. A two-stage estimator $\hat{\theta}_{iT}(\rho_0)$ based on the prior guess ρ_0 is used. And its MSE is estimated by substituting ρ_0 for ρ 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 ρ, along the lines of Pantula and Pollock (1985). This naive estimator of ρ is given by

$$\hat{\rho}_N = \frac{\sum_{i=1}^{k} \sum_{t=1}^{T-2} \hat{a}_{it}(\hat{a}_{i,t+1} - \hat{a}_{i,t+2})}{\sum_{i=1}^{k} \sum_{t=1}^{T-2} \hat{a}_{it}(\hat{a}_{i,t} - \hat{a}_{i,t+1})}, \quad T > 2, \tag{6.44}$$

where $\hat{a}_{it} = y_{it} - x'_{it}X(X'X)^{-1}X'y$ is the (it)-th ordinary least squares residual. The estimator $\hat{\rho}_N$ is inconsistent and typically underestimates ρ (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 ρ in (6.25). Denote this estimator of MSE as $\text{mse}[\hat{\theta}_{iT}(\hat{\rho}_N)]$.

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:

$$\begin{aligned} y_{it} &= v_i + u_{it} + e_{it} \\ u_{it} &= \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1 \end{aligned} \tag{6.45}$$

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 est$(\tilde{\rho}) = 0.36$ 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 (σ_v^2, σ^2). Note that σ_v^2 and σ^2 represent between small area variation relative to sampling variation and between time variation relative to sampling variation respectively since 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 θ_{1T} since MSE values are the same for all θ_{iT}, $i = 1, \ldots, 40$. 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:

$$\mathrm{MSE}(\hat{\theta}_{1T}) = \frac{1}{5000}\sum_{s=1}^{5000}(\hat{\theta}_{1T}^s - \theta_{1T}^s)^2,$$

$$\mathrm{RB[mse]} = \frac{1}{\mathrm{MSE}(\hat{\theta}_{1T})}\left[\frac{1}{5000}\sum_{s=1}^{5000}(\mathrm{mse}_s) - \mathrm{MSE}(\hat{\theta}_{1T})\right],$$

where $\hat{\theta}_{1T}^s$, and θ_{1T}^s and mse_s are the values of $\hat{\theta}_{1T}$, θ_{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

$$\mathrm{GE} = [\{\mathrm{MSE(FH\ estimator)/MSE(two\text{-}stage\ estimator)}\} - 1] \times 100.$$

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 σ^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 σ_v^2, GE decreases as σ^2 increases, whereas it increases with σ_v^2 for a fixed σ^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 σ^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 ρ_0, i.e., the two-stage estimator $\hat{\theta}_{1T}(\rho_0)$ retains its efficiency even when the prior guess ρ_0 deviates significantly from ρ.

We have also computed the RB–values for $\text{mse}[\hat{\theta}_{1T}(\rho_0)]$, the estimator of MSE obtained by substituting ρ_0 for ρ 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 ρ_0 is significantly larger than ρ and σ^2 is small. For example RB $\doteq 10\%$ 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 ρ_0 is not significantly larger than the true ρ.

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 ρ; 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 ρ which increases with ρ.

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 ρ 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 ρ, but it can lead to underestimation which increases with ρ. (For $T = 5$ and $\rho = 0.7$ we obtained RB $\doteq -10\%$). Thus, our method 2 may be satisfactory when ρ 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 σ^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: $\rho = 0.2$

σ_v^2	$\sigma^2 =$ 0.25	0.5	1.0	2.0
	$T = 5$			
0.25	32	19	10	5
0.5	51	29	15	7
1.0	77	45	22	10
2.0	107	64	32	15
	$T = 10$			
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: $\rho = 0.4$

σ_v^2	$\sigma^2 =$ 0.25	0.5	1.0	2.0
	$T = 5$			
0.25	33	21	14	8
0.5	50	30	17	10
1.0	74	44	24	12
2.0	101	62	33	16
	$T = 10$			
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 mse$[\hat{\theta}_{1T}(\rho)]$ and mse$_N[\hat{\theta}_{1T}(\rho)]$ (in brackets): $\rho = 0.2$

σ_v^2	$\sigma^2 =$ 0.25	0.5	1.0	2.0
		$T = 5$		
0.25	2.4(−7.2)	1.7(−4.1)	1.8(−1.5)	1.6(−0.1)
0.5	2.5(−5.6)	1.8(−3.3)	1.7(−1.2)	1.4(−0.2)
1.0	2.7(−4.3)	1.9(−2.5)	1.7(−0.8)	1.3(−0.1)
2.0	2.7(−3.6)	1.9(−2.1)	1.7(−0.6)	1.2(−0.0)
		$T = 10$		
0.25	4.5(−0.6)	4.1(1.2)	3.5(2.0)	2.8(2.0)
0.5	4.4(−0.1)	4.0(1.5)	3.5(2.1)	2.8(2.0)
1.0	4.2(0.2)	3.9(1.6)	3.4(2.1)	2.7(2.0)
2.0	4.1(0.3)	3.8(1.7)	3.3(2.1)	2.6(2.0)

Table 6.4: Relative bias (%) of mse$[\hat{\theta}_{1T}(\rho)]$ and mse$_N[\hat{\theta}_{1T}(\rho)]$ (in brackets): $\rho = 0.4$

σ_v^2	$\sigma^2 =$ 0.25	0.5	1.0	2.0
		$T = 5$		
0.25	2.9(−9.1)	1.3(−5.5)	1.6(−1.8)	1.7(−0.1)
0.5	2.8(−7.7)	1.2(−4.9)	1.4(−1.8)	1.4(−0.3)
1.0	2.9(−6.5)	1.3(−4.2)	1.4(−1.5)	1.2(−0.3)
2.0	2.9(−5.7)	1.3(−3.7)	1.4(−1.2)	1.2(−0.2)
		$T = 10$		
0.25	4.3(−2.6)	3.9(0.4)	3.5(1.7)	2.9(2.1)
0.5	4.3(−1.9)	3.9(0.7)	3.4(1.9)	2.8(2.0)
1.0	4.2(−1.6)	3.8(0.9)	3.4(2.0)	2.7(2.0)
2.0	4.1(−1.5)	3.7(1.0)	3.3(2.0)	2.6(2.0)

Table 6.5: Relative bias (%) of mse$[\hat{\theta}_{1T}(\rho_0)]$: $\rho_0 = 0.1, 0.3$, $\rho = 0.2$; $T = 5$

σ_v^2	$\sigma^2 =$ 0.25	0.5	1.0	2.0
	$\rho_0 = 0.1$			
0.25	1.0	0.5	0.8	0.8
0.5	1.1	0.5	0.6	0.6
1.0	1.1	0.5	0.5	0.5
2.0	1.1	0.4	0.4	0.3
	$\rho_0 = 0.3$			
0.25	4.6	3.4	3.0	2.5
0.5	4.7	3.5	2.9	2.1
1.0	4.8	3.7	3.0	2.0
2.0	4.9	3.8	3.1	2.1

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

σ_v^2	$\sigma^2 =$ 0.25	0.5	1.0	2.0
		$\rho_0 = 0.2$		
0.25	−1.4	−2.0	−1.1	−0.4
0.5	−1.5	−2.2	−1.3	−0.5
1.0	−1.5	−2.3	−1.5	−0.7
2.0	−1.5	−2.4	−1.7	−1.0
		$\rho_0 = 0.3$		
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	2.9	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 mse$[\hat{\theta}_{1T}(\hat{\rho}_N)]$: $\rho = 0.2$

σ_v^2	$\sigma^2 =$ 0.25	0.5	1.0	2.0
	$T = 5$			
0.25	−1.0	−0.7	0.3	0.7
0.5	−0.8	−0.7	0.1	0.5
1.0	−0.6	−0.6	0.0	0.4
2.0	−0.4	−0.5	0.0	0.3
	$T = 10$			
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 mse$[\hat{\theta}_{1T}(\hat{\rho}_N)]$: $\rho = 0.4$

σ_v^2	$\sigma^2 =$ 0.25	0.5	1.0	2.0
	$T = 5$			
0.25	−4.0	−3.8	−1.8	−0.3
0.5	−4.0	−3.9	−2.1	−0.5
1.0	−3.9	−4.0	−2.3	−0.7
2.0	−3.7	−4.0	−2.4	−0.9
	$T = 10$			
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 ρ known. Two methods for implementing the HB approach, the numerical integration and Gibbs sampler, are studied. In Section 7.2, the case of ρ unknown is briefly discussed.

7.1 HB Estimator: ρ Known

Consider our model (6.1). Similar to (6.2), the model can be written as

$$y_i = \theta_i + e_i = X_i\beta + v_i 1_T + u_i + e_i \tag{7.1}$$

where $\theta_i = (\theta_{i1}, \ldots, \theta_{iT})'$, y_i, X_i, e_i are the same as those for (6.2). With ρ known, we use the following four-stage hierarchical Bayes model.

(I) Conditional on θ_{it}'s, $y_i \overset{ind}{\sim} N(\theta_i,\ \Sigma_i)$.

(II) Conditional on v_i, β, and r_1 $(= 1/\sigma^2)$, $\theta_i \overset{ind}{\sim} N(X_i\beta + v_i 1_T,\ r_1^{-1}\Gamma)$.

(III) Conditional on r_2 $(= 1/\sigma_v^2)$, $v_i \sim N(0, r_2^{-1} I_T)$.

(IV) The marginal distribution of β, r_1, and r_2 are assumed independent with

$\beta \sim uniform(R^p)$,

$r_1 \sim Gamma(a_1/2, b_1/2)$,

$r_2 \sim Gamma(a_2/2, b_2/2)$,

where $Gamma(a, b)$ is a Gamma random variable with density function

$$g(r) = a^b r^{b-1} e^{-ar}/\Gamma(b), \quad \text{for } r > 0,\ a > 0 \text{ and } b > 0.$$

Note that model (7.1) (with normal errors) is the combination of stages (I), (II) and (III). In stage (IV), the parameter β 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 θ_{iT} given the observations $\{y_{it}\}$, say $g(\theta_{iT} \mid y)$. The parameter of interest θ_{iT} is estimated by $E(\theta_{iT} \mid y)$ and its precision is measured by $V(\theta_{iT} \mid y)$. Unfortunately, an explicit form of $g(\theta_{iT} \mid y)$ is not available due to model complexity. There are at least two approaches to obtaining the posterior expectation and posterior variance of θ_{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, θ, v, β, r_1 and r_2, as

$$\begin{aligned} & g(y, \theta, v, \beta, r_1, r_2) \\ & \quad \propto \exp\{-(y-\theta)'\Sigma^{-1}(y-\theta)/2\} \\ & \qquad \times r_1^{kT/2} \exp\{-r_1(\theta - X\beta - Zv)'R^{-1}(\theta - X\beta - Zv)/2\} \\ & \qquad \times r_2^{-k/2} \exp\{-r_2 v'v/2\} \times r_1^{b_1/2-1} e^{-a_1 r_1/2} \times r_2^{b_2/2-1} e^{-a_2 r_2/2}, \end{aligned} \tag{7.2}$$

where $\theta = (\theta_{11}, \ldots, \theta_{1T}, \ldots, \theta_{k1}, \ldots, \theta_{kT})'$, y, X, v, R, Z and Σ are the same as those defined in (6.2).

Integrating over v, we can find that the joint p.d.f. of y, θ, β, r_1 and r_2 is

$$\begin{aligned} & g(y, \theta, \beta, r_1, r_2) \\ & \quad \propto \exp\{-(y-\theta)'\Sigma^{-1}(y-\theta)/2\} \\ & \qquad r_1^{kT/2} r_2^{k/2} |r_1 1_T' \Gamma^{-1} 1_T + r_2|^{-k/2} \times \exp\{-r_1(\theta - X\beta)'U^{-1}(\theta - X\beta)/2\} \\ & \qquad \times r_1^{b_1/2-1} e^{-a_1 r_1/2} \times r_2^{b_2/2-1} e^{-a_2 r_2/2}, \end{aligned} \tag{7.3}$$

where $U = \text{block diag}_i(\sigma_v^2 1_T 1_T' + \sigma_u^2 \Gamma)$, Further, denote

$$U_1 = X'U^{-1}X.$$

From the following identity,

$$
\begin{aligned}
&(\theta - X\beta)'U^{-1}(\theta - X\beta) \\
&= (\beta - U_1^{-1}X'U^{-1}\theta)'U_1(\beta - U_1^{-1}X'U^{-1}\theta) + \theta'(U^{-1} - U^{-1}XU_1^{-1}X'U^{-1})\theta,
\end{aligned}
$$

we can obtain the joint p.d.f. of y, θ, r_1 and r_2 (by integrating over β in (7.3)) as

$$
\begin{aligned}
&g(y, \theta, r_1, r_2) \qquad (7.4) \\
&\propto r_1^{(nT-p)/2} r_2^{k/2} |r_1 1_T' \Gamma^{-1} 1_T + r_2|^{-k/2} |U_1|^{1/2} \\
&\quad \times \exp\{(-(y-\theta)'\Sigma^{-1}(y-\theta)/2\} \\
&\quad \times \exp\{-r_1\theta'(U^{-1} - U^{-1}XU_1^{-1}X'U^{-1})\theta/2\} \ . \\
&\quad \times r_1^{b_1/2-1} e^{-a_1 r_1/2} \times r_2^{b_2/2-1} e^{-a_2 r_2/2}.
\end{aligned}
$$

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 $(r_1 1_T' \Gamma^{-1} 1_T + r_2)$ and U_1. Nonetheless, noting that

$$
\begin{aligned}
&(y-\theta)'\Sigma^{-1}(y-\theta) + r_1\theta'(U^{-1} - U^{-1}XU_1^{-1}X'U^{-1})\theta \\
&= (\theta - W\Sigma^{-1}y)'W^{-1}(\theta - W\Sigma^{-1}y)' + y'(\Sigma^{-1} - \Sigma^{-1}W\Sigma^{-1})y
\end{aligned}
$$

where

$$W^{-1} = \Sigma^{-1} + r_1 U^{-1} - r_1 U^{-1}XU_1^{-1}X'U^{-1},$$

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 θ given y, r_1 and r_2 are

$$M(y, r_1, r_2) = E(\theta \mid y, r_1, r_2) = W\Sigma^{-1}y, \qquad (7.5)$$

$$V(y, r_1, r_2) = V(\theta \mid y, r_1, r_2) = W. \qquad (7.6)$$

It is interesting to note that $W\Sigma^{-1}y$ is exactly the BLUP we obtained in Section 6.1 for θ due to the following fact

$$W = \Sigma - \Sigma V^{-1}\Sigma + \Sigma V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\Sigma,$$

(note that $V = \Sigma + \sigma_v^2 ZZ' + \sigma_u^2 R$ in Chapter 6).

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 θ as

$$E(\theta \mid y) = \iint M(y, r_1, r_2) g(r_1, r_2 \mid y) dr_1 dr_2$$

and

$$\begin{aligned} Cov(\theta \mid y) = & \iint V(y, r_1, r_2) g(r_1, r_2 \mid y) dr_1 dr_2 \\ & + \iint [M(y, r_1, r_2) - E(\theta \mid y)][M(y, r_1, r_2) \\ & \quad - E(\theta \mid y)]' g(r_1, r_2 \mid y) dr_1 dr_2. \end{aligned}$$

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 $[U_i]$, $i = 1, \ldots, m$. The implementation of the Gibbs sampling scheme is as follows. Given an arbitrary starting set of values $U_1^{(0)}, U_2^{(0)}, \ldots, U_m^{(0)}$, draw $U_1^{(1)}$ from $[U_1 \mid U_2^{(0)}, U_3^{(0)} \ldots, U_m^{(0)}]$, $U_2^{(1)}$ from $[U_2 \mid U_1^{(1)}, U_3^{(0)} \ldots, U_m^{(0)}]$, and so on, up to $U_m^{(1)}$ from $[U_m \mid U_1^{(1)}, U_2^{(1)} \ldots, U_{m-1}^{(1)}]$. Thus each variable is visited once in this order, and a cycle is completed after m random number generations. This process is continued with $U_1^{(1)}, U_2^{(1)}, \ldots, U_m^{(1)}$. After a large number, B, of iterations, we would obtained an m-tuple $U_1^{(B)}, U_2^{(B)}, \ldots, U_m^{(B)}$. 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 $U_1^{(B+tg)}, U_2^{(B+tg)}, \ldots, U_m^{(B+tg)}$, 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

$$\frac{1}{G} \sum_{g=1}^{G} [U_i \mid U_s^{(B+tg)},\ s \neq i].$$

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, θ, 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 θ given

y can be approximated by

$$\frac{1}{G}\sum_{g=1}^{G}[\theta \mid v = v^{(B+tg)}, \beta = \beta^{(B+tg)}, r_1 = r_1^{(B+tg)}, r_2 = r_2^{(B+tg)}, y].$$

The posterior expectation and posterior covariance matrix of θ, can be approximated from the above conditional distributions by

$$E(\theta \mid y) \approx \frac{1}{G}\sum_{g=1}^{G}\Lambda_g, \tag{7.7}$$

and

$$\begin{aligned} &COV(\theta \mid y) \\ &\approx \frac{1}{G}\sum_{g=1}^{G} COV(\theta \mid v = v^{(B+tg)}, \beta = \beta^{(B+tg)}, r_1 = r_1^{(B+tg)}, r_2 = r_2^{(B+tg)}, y) \\ &\quad + \frac{1}{G-1}\left[\sum_{g=1}^{G}\Lambda_g\Lambda_g' - \frac{1}{G}\sum_{i=1}^{G}\Lambda_g\sum_{i=1}^{G}\Lambda_g'\right], \end{aligned} \tag{7.8}$$

where

$$\Lambda_g = E(\theta \mid v = v^{(B+tg)}, \beta = \beta^{(B+tg)}, r_1 = r_1^{(B+tg)}, r_2 = r_2^{(B+tg)}, y).$$

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: ρ Unknown

The HB estimation for the case of unknown ρ turns out to be much more difficult due to the inclusion of unknown correlation ρ. 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, β, r_1 and ρ, $\theta_i \overset{ind}{\sim} N(X_i\beta + v_i 1_T,\ r_1^{-1}\Gamma)$.

(III)′ is same as (III).

(IV)′ Assume the same prior distributions on β, r_1 and r_2 as in (IV).

(V)′ Note that difficulty arises regarding the choice of prior distribution for ρ. Most papers in the literature avoid this difficulty by assuming a known ρ (the case in Section 7.1), e.g., random walk. A plausible prior distribution for ρ 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 ρ and the posterior distribution of r_1, r_2 and ρ 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 $g(r_1, r_2, \rho \mid y)$, we can compute the posterior mean and the posterior covariance matrix of θ as

$$E(\theta \mid y) = \iiint M(y, r_1, r_2, \rho) g(r_1, r_2, \rho \mid y) dr_1 dr_2 d\rho,$$

and

$$\begin{aligned} Cov(\theta \mid y) = & \iiint V(y, r_1, r_2, \rho) g(r_1, r_2, \rho \mid y) dr_1 dr_2 d\rho \\ & + \iiint [M(y, r_1, r_2, \rho) - E(\theta \mid y)][M(y, r_1, r_2, \rho) \\ & \quad - E(\theta \mid y)]' g(r_1, r_2, \rho \mid y) dr_1 dr_2 d\rho. \end{aligned}$$

However, closed-form expression for $E(\theta \mid y)$ and $Cov(\theta \mid y)$ 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 θ given y can be obtained through an iterative Monte–Carlo procedure by sampling from the full conditional distributions $g(\theta \mid \beta, v, r_1, r_2, \rho, y)$, $g(v \mid \theta, \beta, r_1, r_2, \rho, y)$, $g(\beta \mid \theta, v, r_1, r_2, \rho, y)$, $g(r_1 \mid \theta, \beta, v, r_2, \rho, y)$,

$g(r_2 \mid \theta, \beta, v, r_1, \rho, y)$ and $g(\rho \mid \theta, \beta, v, r_1, r_2, y)$ which determine the joint distribution of $\theta, v, \beta, r_1, r_2, \rho$ 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') $r_1 \mid \theta, v, \beta, r_2, \rho, y \sim Gamma[\frac{1}{2}\{a_1 + (\theta - Xt\beta - Zv)'R^{-1}(\theta - X\beta - Zv)\}, \frac{1}{2}(kT + b_1)]$ (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, θ, r_1),

(vi') $\rho \mid \theta, v, \beta, r_1, r_2, y = [c(\beta, \theta, v, r_1)]^{-1} |R|^{-1/2} \exp[-\frac{r_1}{2}(\theta - X\beta - Zv)'R^{-1} \times (\theta - X\beta - Zv)] g(\rho)$ (independent of y, r_2) where $g(\rho)$ is the prior probability density imposed on ρ, and

$$c(\beta, \theta, v, r_1) = \int |R|^{-1/2} \exp\{-\tfrac{r_1}{2}(\theta - X\beta - Zv)'R^{-1}(\theta - X\beta - Zv)\} g(\rho) d\rho.$$

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 σ_v^2 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 σ_v^2 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 β, unlike ANOVA or MINQU estimators. Using the truncated AMINQU estimators, we have obtained a WLS estimator of regression parameters β in Chapter 3. The asymptotic normality of this WLS estimator of β 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 σ_v^2 is moderate or large relative to error variances σ_i^2. We have also constructed confidence intervals on the elements of β using the WLS estimator and the corresponding variance estimator. These confidence intervals lead to accurate coverage probability. An alternative WLS estimator of β is also considered, using the AMINQU estimator of σ_v^2 and the within-group sample variances as estimators of error variances σ_i^2. Its asymptotic properties are established. But in the case of a small number of observations within a group, and small or moderate σ_v^2 relative to σ_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 AMINQU estimators to reduce the bias in Chapter 4. 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. In a limited simulation study, we have found that the jackknife method reduces the bias in the untruncated and truncated AMINQU estimators. We have also found that jackknife variance estimators for the elements of β 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 β, alternative WLS estimators are proposed using within-group sample variances as estimators of error variances. Asymptotic normality of the resulting WLS estimator of β 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, ρ, 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 ρ, we have obtained a consistent estimator of ρ, 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 ρ, while method 2 uses a naive estimator of ρ 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 ρ is not significantly larger than the

true ρ. Our method 2 of estimating MSE may also be satisfactory when ρ 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 ρ 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 ρ is unknown, remains to be done. Also, we need to study the sensitivity of HB inference to choice of priors, especially when ρ is unknown.

Bibliography

Anderson, T. W. and Hsiao, C. (1981). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, **18**, 67–82.

Arvesen, J. N. (1969). Jackknifing U-statistics. *Annals of Mathematical Statistics*, **40**, 2076–2100.

Arvesen, J. N. and Layard M. W. J. (1975). Asymptotically robust tests in unbalanced variance component models. *Annals of Statistics*, **3**, 1122–1134.

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **80**, 28–36.

Bement, T. R. and Williams J. S. (1969). Variance of weighted regression estimators when sampling errors are independent and heteroscedastic. *Journal of the American Statistical Association*, **64**, 1369–1382.

Bickel P. J. and Doksum, K. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden–Day Inc.

Binder, D. A. and Dick, J. P. (1989). Modelling and estimation for repeated surveys. *Survey Methodology*, **15**, 29–45.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press.

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *Annals of Statistics* **10**, 1224–1233.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*, Chapman & Hall, New York.

Chen J. H. and Shao J. (1993). Iterative weighted least squares estimators. *Annals*

of Statistics, **21**, 1071–1092.

Choudhry, G. H. and Rao, J. N. K. (1989). Small area estimation using models that combine time series and cross-sectional data. *Proceedings of Statistics Canada Symposium on Analysis of Data in Time* (Eds. A. C. Singh and P. Whitridge), 67–74.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society, Supplement* **4**, 102–118.

Cochran, W. G. and Carroll, S. P. (1953). A sampling investigation of the efficiency of weighting inversely as the estimated variance. *Biometrics*, **9**, 447–459.

Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression*. Chapman and Hall.

Cressie, N. (1992). REML estimation in empirical Bayes smoothing of census undercount. *Survey Methodology*, **18**, 75–94.

Davidian M. and Carroll R. J. (1987). Variance function estimation. *Journal of the Ameican Statisical Assoiation*, **82**, 1079–1091.

Datta, G. S. and Ghosh, M. (1991). Bayesian prediction in linear models: application to small area estimation. *Annals of Statistics*, **19**, 1748–1770.

Datta, G. S., Fay, R. E. and Ghosh, M. (1991). Hierarchical and empirical Bayes analysis in small area estimation. *Proceedings of the Annual Research Conference, U.S. Bureau of the Census*, 63–78.

Drew, J. D., Singh, M. P. and Choudhry, G. H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Survey Methodology*, **8**, 17–47.

Ericksen, E. P. (1974). A regression method for estimating population of local areas. *Journal of the American Statistical Association*, **69**, 867–875.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, *Annals of Statistics* **13**, 342–348.

Fay, R. E. (1987). Application of multivariate regression to small domain estimation. 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.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

Fuller, W. A. (1987). *Measurement Error Models*, John Wiley & Sons, New York.

Fuller W. A. and Battese P. E. (1973). Transformations for estimation of linear models with nested-error structure. *Journal of the the American Statistical Association*, **68**, 626–632.

Fuller, W. A. and Harter, R. M. (1987). The multivariate components of variance model for small area estimation, In *Small Area Statistics* (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh eds) 103–123. John Wiley & Sons, New York.

Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Annals of Statistics*, **5**, 1149–1158.

Gelfand, A. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples. *Journal of the American Statistical Association*, **82**, 1153–1162.

Ghosh M. and Nangia N. (1993). Estimation of median income of four-person families: A Baysian time series approach. Technical Report, #429, Dept. of Statistics, University of Florida.

Ghosh, M. and Rao, J. N. K. (1993). Small area estimation: an appraisal. *Statistical Science*, **8**, (to appear).

Gonzalez, M. E. (1973). Use and evaluation of synthetic estimators. *Proceedings of the American Statistical Association, Social Statistics Section*, 33-36.

Hartley, H. O., Rao, J. N. K. and Kiefer, G. (1969). Variance estimation with one unit per stratum. *Journal of the American Statistical Association*, **64**, 841–851.

Henderson, C. R. (1950). Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics*, **21**, 309–310.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, **9**, 226–252.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–447.

Hinkley D. V. (1977). Jackknifing in unbalanced situations, *Technometrics*, **19**, 285–292.

Jones, R. G. (1980). Best linear unbiased estimators for repeated surveys. *Journal of the Royal Statistical Society*, Series B, **42**, 221–226.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T. (1985). *The Theory and Practice of Econometrics* (2nd Ed.). John Wiley & Sons, New York.

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statisical Association*, **79**, 853–862.

Kakwani, N. C. (1967). The unbiasedness of Zellner's seemingly unrelated regression equations estimators. *Journal of the American Statistical Association*, **62**, 141–142.

Kass, R. E. and Steffey, D. (1989). Approximations for standard errors of estimators for fixed and random effects in mixed models (parametric empirical Bayes models). *Journal of the American Statistical Association*, **84**, 717–726.

Jacquez, J. A., Mather, F. J. and Crawford, C. R. (1968). Linear regression model with nonconstant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics*, **24**, 607–626.

La Motte, L. R. (1976). Invariant quadratic estimators in the random, one–way ANOVA model. *Biometrics*, **32**, 793–804.

Lahiri, P. and Rao, J. N. K. (1992). Robust estimation of mean square error of small

area estimators. Unpublished manuscript.

Laird, N. M. and Louis, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association*, **82**, 739–750.

Lee, H. (1990). Estimation of panel correlations for the Canadian Labour Force Survey. *Survey Methodology*, **16**, 283–292.

Levy, P. (1970). Combining independent estimators—An empirical sampling study. *Technometrics*, **12**, 162–165.

MacGibbon, B. and Tomberlin, T. J. (1989). Small area estimation of proportions via empirical Bayes techniques. *Survey Methodology*, **15**, 237–252.

Meier, P. (1953). Variance of a weighted mean. *Biometrics*, **9**, 59–73.

Miller R. G. (1964). A trustworthy jackknife. *Annals of Mathematical Statistics*, **35**, 1594–1605.

Miller R. G. (1968). Jackknifing variances. *Annals of Mathematical Statistics*, **39**, 567–582.

Miller R. G. (1974a). The jackknife–a review. *Biometrika*, **61**, 1–15.

Miller R. G. (1974b). An unbalanced jackknife. *Annals of Statistics*, **2**, 880–891.

National Center for Health Statistics (1979). *Small Area Estimation: an Empirical Comparison of Conventional and Synthetic Estimators for States.* Vital and Health Statistics. Series 2, No. 82, U. S. Government Printing Office, Washington, D.C.

Pantula, S. G. and Pollock, K. H. (1985). Nested analysis of variance with autocorrelated errors. *Biometrics*, **41**, 909–920.

Pfeffermann, D. and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, **16**, 217–237.

Prasad, N. G. N. and Rao, J. N. K. (1986). On the estimation of mean squared error of small area predictors. In Tech. Report 97, Carleton University, Laboratory for Research in Statistics and Probability, Ottawa.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*,

85, 163–171.

Purcell, N. J. and Kish, L. (1980). Postcensal estimates for local areas (or domains). *International Statistical Review*, **48**,3–18.

Quenouille, M. H. (1949). Approximate Tests of Correlatio.ı in Time-Series. *Journal of the Royal Statistics Society*, Series B, **11**, 68-74.

Quenouille, M. H. (1956). Notes on Bias in Estimation, *Biometrika*, **43**, 353-360.

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statisical Association*, **65**, 161–172.

Rao, C. R. (1971). Estimation of variance and covariance components—MINQUE theory. *Journal of Multivariate Analysis*, **3**, 257–275.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, **67**, 112–115.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd Edition, John Wiley and Sons, New York.

Rao, C. R. (1984). Optimization with Applications to Statistical Problems, in *W. G. Cochran's Impact on Statistics*, P. S. R. S. Rao and J. Sedransk (Ed.), John Wiley and Sons, New York, 191–202.

Rao, C. R. and Kleffe, J. (1988). *Estimation of Variance Components and Applications*. North-Holland Publishing Company.

Rao, J. N. K. (1971). Combining independent estimators and estimation in linear regression with unequal variances. *Biometrics*, **27**, 971–990.

Rao, J. N. K. (1973). On the estimation of heteroscedastic variances. *Biometrics*, **29**, 11–24.

Rao, J. N. K. (1980). Estimating the common mean of possibly different normal populations: a simulation study. *Journal of the American Statistical Association*, **75**, 447–453.

Rao, J. N. K. (1986). Synthetic estimators, SPREE and best model-based predictors of small area means. In Tech. Report 97, Carleton University, Laboratory for Research in Statistics and Probability, Ottawa.

Rao, P. S. R. S. (1977). Theory of MINQUE–a review. *Sankyā*, **39**, 201–210.

Rao, P. S. R. S. and Chaubey Y. P. (1978). Three modifications of the principle of the MINQUE. *Communications in Statististics-Theory and Methods*, **A7**(8), 767–778.

Rao, P. S. R. S. and Heckler, C. E. (1992). Estimators for the three-fold nested random effect model. Unpublished manuscript.

Rao, P. S. R. S. and Kuranchie, P. (1987). Estimators for the random intercept regression model. *Communications in Statistics-Theory and Methods*, **17**, 1011–1026.

Rao, P. S. R. S., Kaplan, J. and Cochran, W. G. (1981). Estimators for the one-way random effects model with unequal error variances. *Journal of the American Statistical Association.*, **76**, 89–97.

Särndal, C. E. and Hidiroglou, M. A. (1989). Small domain estimation: a conditional analysis. *Journal of the American Statistical Association*, **84**, 266–275.

Scott, A. J. and Smith, T. M. P. (1974). Analysis of repeated surveys using time series methods. *Journal of the American Statistical Association*, **69**, 674–678.

Searle, S. R. (1971). *Linear Models.* John Wiley & Sons, New York.

Seber, G. A. F. (1977). *Linear Regression Analysis.* John Wiley & Sons, New York.

Shao J. (1989a). Asymptotic distribution of the weighted least squares estimator. *Annals of the Institute of Statistical Mathematics*, **41**, 365–382.

Shao J. (1989b). Jackknifing weighted least squares estimators. *Journal of the Royal Statistical Society*, Series B, **51**, 139–156.

Shao J. (1991). Empirical bayes estimation of heteroscedastic variances. *Statistica Sinica*, **2**, 495–518.

Shao J. and Rao J. N. K. (1993). A jackknife method for heteroscedastic linear regression models with replication and possibly correlated error within groups. *Canadian Journal of Statistics*, **21**.

Singh, A. C., Mantel, H. and Thomas, B. W. (1991). Time series generalizations of Fay-Herriot estimator for small areas. *Proceedings of the Survey Research Methods Section, the American Statistical Association*, 455–460.

Stukel, D. (1991). *Small Area Estimation Under One and Two-Fold Nested Error*

Regression Model. Ph.D. Thesis, Carleton University, Ottawa.

Tiller, R. B. (1992). Time Series modelling of sample survey data from the U. S. current Population Survey. *Journal of Official Statistics*, **8**, 149–166.

Tukey J. (1958). Bias and confidence in not quite large sample (abstract). *Annals of Mathematical Statistics*, **29**, 614.

William J. S. (1967). The variance of weighted regression estimators, *Journal of the American Statistical Association*, **62**, 1290–1301.

Wolter, K. (1985). *Introduction to Variance Estimation.* Springer, New York.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, **14**, 1261–1295.

Yates, F. and Cochran W. G. (1938). The analysis of groups of experiments, *Journal of Agricultural Sciences*, **28**, 556–580.

END

18-07-94

FIN