Thesis for the Degree of Master of Science

Confidence Intervals for Regression Coefficient in a Regression model with a Balanced Two-Fold Nested Error Structure

by

Dong Hwi Kim

Department of Statistics

The Graduate School

Pukyong National University

February 2013

Confidence Intervals for Regression Coefficient in a Regression Model with a Balanced Two-Fold Nested Error Structure

(이중중첩오차구조를 갖는 회귀모형의 회귀계수에 대한 신뢰구간)

Advisor: Prof. Dong Joon Park

Dong Hwi Kim

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in Department of Statistics, The Graduate School,
Pukyong National University

February 2013

Contents

List of Tables

이중중첩오차구조를 갖는 회귀모형의 회귀계수에 대한 신뢰구간

김 동 휘

부경대학교 대학원 통계학과

요 약

이 논문은 이중중첩오차구조를 갖는 단순선형회귀모형에 관한 연구를 한다. 이 모형은 3단계 샘플링을 실행을 할 때 나타나는 모형으로서 1단계 샘플링 후, 선택된 샘플링 단위 내에서 2단계 샘플링을 수행하고, 선택된 샘플링 단위 내에서 다시 3단계 샘플링을 하여 선택된 독립변수와 종속변수를 단순회귀모형으로 가정하고 그 모형에서 나타난 회귀계수의 신뢰구간을 유도한다. 1, 2, 3 단계에서 발생하는 오차들로 서로 독립이고 일정한 분산을 갖는 정규확률변수로 가정한다. 이 모형에서 회귀계수의 최소제곱추정량을 사용하여 정확한 신뢰구간과 근사적인 신뢰구간을 유도한다. 유도한 신뢰구간의 시뮬레이션을 수행하여 서로 비교하고 수치예제를 사용하여 실제 활용 방법을 소개한다.

CHAPTER 1
INTRODUCTION

Statistical modeling means to formulate relationships between variables in the form of mathematical equations in order to describe the data that are randomly selected subset of a population of research interest. Practitioners make statistical inferences to quantify the parameters in statistical models that represent characteristics of the population through confidence interval approach. The applications of utilizing confidence intervals are encountered in a variety of fields: optimal maintenance planning, neural networks, quality engineering, behavioral science, economy, biology, etc.

The regression model has been used to describe the relationship between the response and predictor variables. In simple linear regression, the regression model has one predictor variable and the variables are linearly related. Samples from observational studies are often selected with a three-stage sampling structure. This thesis discusses statistical inference concerning the regression coefficient in the simple linear regression model with a balanced two-fold nested error structure. This model is appropriate to use when there is subsampling within secondary sampling units within primary sampling units. The model therefore includes one error term associated with the first-stage sampling unit, a second error term associated with the second-stage sampling unit, and a third error term associated with the last-stage sampling unit. These three error terms are assumed independent and normally distributed with zero means and constant variances. However, this nested error structure yields response variables correlated.

Since confidence intervals are usually more informative than hypothesis testing, confidence intervals are presented to make inferences concerning a regression coefficient. In this thesis we derive exact and approximate confidence intervals for the regression coefficient using large sample theory. A $100(1-\alpha)\%$ confidence interval for a parameter, for example γ, is referred to as a random interval with a lower limit L and an upper limit U that are functions of sample values such that

$$P[L \leq \gamma \leq U] = 1 - \alpha \tag{1.1}$$

where α ranges from 0 to 1. A confidence interval that exactly holds equation (1.1) is called an "exact" two-sided confidence interval. Such exact intervals often do not exist in application. An approximate interval that has a realized confidence coefficient greater than the stated level, i.e.,

$$P[L \leq \gamma \leq U] > 1 - \alpha \tag{1.2}$$

is called a "conservative" two-sided confidence interval. An approximate interval that has a realized confidence coefficient less than the stated level, i.e.,

$$P[L \leq \gamma \leq U] < 1 - \alpha \tag{1.3}$$

is called a "liberal" two-sided confidence interval. Both conservative interval and liberal interval are called "approximate" intervals. In general, conservative intervals are preferred when only approximate intervals are available.

Chapter 2 presents a review of current research concerning the variability of variance components and regression coefficients in a regression model with nested error structure. Chapter 3 considers a simple linear regression model with a balanced two-fold nested error structure and some distributional results are examined. Exact and approximate confidence intervals for the regression coefficient are proposed. Chapter 4 conducts a simulation study to compare the performance of the proposed confidence intervals. Finally, Chapter 5 summarizes the results and a numerical example is presented in order to demonstrate the methodology proposed. Recommendations are provided for selecting an appropriate method. Topics for future research are also presented.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews previous research on confidence intervals on linear combination of variance components and regression coefficients in linear models. A general discussion of statistical inferences for linear models is presented in Section 2.2 and 2.3. The methods and applications of confidence intervals in regression models with nested error structure are reviewed in Section 2.4.

2.2 Confidence Intervals in Linear Models

The earliest work for analyzing a statistical model was performed in 1924 by Fisher and his work covered optimum methods for statistical inferences including point estimation, confidence intervals, and hypothesis tests. In many applications of statistical inferences concerning analysis of variance, interest focuses on variance of the effects rather than linear combinations of the effects. The variances associated with the effects are called variance components. Let $n_q S_q^2/\theta_q$ for $q = 1, \ldots, Q$ be independently distributed chi-squared random variables with n_q degrees of freedom, respectively. It is frequently desired to construct confidence intervals on linear combinations of the θ_q, i.e., $\gamma = \Sigma_q^Q c_q \theta_q$ where $c_q \geq 0$. In practice, θ_q represents variance components. Satterthwaite (1941, 1946) studied the distribution of a linear combination of more than one θ_q. He proposed an approximation based on the point estimator of γ, $\hat{\gamma} = \Sigma_q c_q S_q^2$. By equating the expectation and variance of $\hat{\gamma}$ to that of an exact chi-squared random variable, he was able to derive the approximate distribution of $\hat{\gamma}$. He then used this approximation to construct a confidence interval on γ.

Welch (1956) developed two alternative approximations designed to work well in moderate-sized samples. The confidence limits are based on modification of the large-sample normal approximation using Taylor series and Cornish-Fisher expansions. One of the approximations can be represented as an improved Satterthwaite approximation. Burdick and Sielken (1978) developed exact confidence intervals on γ for one-fold and two-fold nested models. The method can be extended to any K-fold nested model and applied to both balanced and unbalanced designs. Graybill and Wang (1980) developed for constructing confidence intervals on γ with nonnegative linear combinations of variances. Burdick and Graybill (1992) provided the research of confidence intervals on the variance components in linear models.

2.3 Point Estimation in Nested Regression Models

An important model employed in a variety of survey sampling and experimental design applications is a regression model with nested error structure. The model is an example of the mixed models and can be formulated in several ways. Longford (1985) formulated a statistical model for clustered observations where subjects are nested within groups. Observations within clusters tend to be more homogeneous than those between clusters. Longford modified the classical linear regression model by allowing the regression parameters to vary from cluster to cluster. The model is

$$Y_{ij} = \Sigma_{k=0}^{p} X_{ij,k}\beta_{i,k} + E_{ij} \tag{2.3.1}$$
$$i = 1, \ldots, I; \quad j = 1, \ldots, J_i$$

where Y_{ij} is the value of the response variable for the jth observation in the ith cluster, $X_{ij,0} = 1$, $X_{ij,k}$ represents the kth explanatory variable, $\beta_{i,k}$ represents the kth regression parameters, and $\beta_i = [\beta_{i,0}, \ldots, \beta_{i,p}]$ is a random vector from a $(p+1)$-variate normal distribution with mean $\beta = [\beta_0, \ldots, \beta_p]$ and variance matrix Σ where

$$\Sigma = \begin{pmatrix} \theta_0 & \theta_{01} & \theta_{02} & \ldots & \theta_{0p} \\ \theta_{01} & \theta_1 & 0 & \ldots & 0 \\ \theta_{02} & 0 & \theta_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{0p} & 0 & 0 & \ldots & \theta_p \end{pmatrix}. \tag{2.3.2}$$

The term E_{ij} represents a random variable from $N(0, \sigma_E^2)$ and β_i and E_{ij} are independent. Longford applied this variance component model to examine explanatory variables measured on towns (clusters) and on subunits within towns, and studied the relationship of house price with associated explanatory variables.

Aitkin and Longford (1986) studied several statistical models for assessment of school effectiveness in educational research studies. They used a data set of 907 students in 18 schools from one Local Educational Authority (LEA). The LEA score for each student was used as the response variable with a Verbal Reasoning Quotient (VRQ) score and sex as explanatory variables. Although they compared several models, they recommended a variance component model when clustering is inherent in the data structure. The variance component model with one explanatory variable is

$$Y_{ij} = \mu + \beta X_{ij} + A_i + E_{ij} \tag{2.3.3}$$

$$i = 1, \ldots, I; \quad j = 1, \ldots, J_i$$

where A_i is the school effects, E_{ij} is the student errors, and they are independent and random samples from $N(0, \sigma_A^2)$ and $N(0, \sigma_E^2)$, respectively. The observations Y_{ij} within a school are not independent because

$$Cov(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma_A^2 + \sigma_E^2 & \text{if } i = i', j = j'; \\ \sigma_A^2 & \text{if } i = i', j \neq j'; \\ 0 & \text{if } i \neq i'. \end{cases} \tag{2.3.4}$$

Parameter estimates can be determined using the Fisher scoring algorithm of Longford (1987) which provides maximum likelihood estimates and standard errors for the variance components.

Goldstein and McDonald (1988) extended model (2.3.3) into a more general model that includes time series models, longitudinal data models, multiple matrix sampling models, generalizability theory models, multilevel common factor models, and complex sample survey designs. For a wide class of models, he recommends an iterative generalized least square procedure (IGLS) for estimation. This method is relatively straight forward and easily incorporates adjustments for errors in variables not available with standard maximum likelihood (ML) approaches. He also showed that IGLS is equivalent to ML for a wide class of models when the random variables have a multivariate normal distribution.

2.4 Confidence Intervals in Nested Regression Models

Previous research on nested regression models has mostly concentrated on point estimation of the regression coefficients and the variance components. Burdick and Graybill (1992) reviewed the research of confidence intervals in linear models. However, the book rarely dealt with linear models with factors and covariates.

Park and Burdick (1993, 1994) and Park and Hwang (2002) constructed confidence intervals in a simple regression model with one-fold nested error structure. Park and Burdick (2003, 2004) extended their previous works by constructing confidence intervals on the linear functions of variance components in the model with one factor and a covariate that has unbalanced nested error structure. Burdick et al. (2005) applied the methods for constructing confidence intervals on the linear functions of variance components to test a measurement system using gauge R & R(Repeatability and Reproducibility) study which is an application field of quality engineering. When a measurement system in the process is expressed into a statistical model, the confidence intervals for repeatability and reproducibility, i.e. functions of the variances in the model, have been useful tools to determine if the measurement system is under control. Park and Yoon (2009) proposed confidence intervals for functions of variances in a two-factor model with a covariate for gauge R & R study application.

Park (2012) extended further the works by Park and Burdick (1993, 1994) and Park and Hwang (2002) by including two-fold nested error structure in a simple regression model with one factor and a covariate. He derived approximate confidence intervals on the variance components in a simple regression model with balanced two-fold nested error structure. This thesis extends his work and constructs exact and approximate confidence intervals on the regression coefficient in the model.

æ

CHAPTER 3

CONFIDENCE INTERVALS ON THE REGRESSION COEFFICIENT IN REGRESSION MODEL WITH A BALANCED TWO-FOLD NESTED ERROR STRUCTURE

3.1 Introduction

The parameter of interest in this chapter is the regression coefficient, β, in regression model with a balanced two-fold nested error structure. The model is defined in Section 3.2. A possible partitioning for source of variability of the model is shown in Section 3.3. Distributional properties of several ordinary least squares (OLS) estimators for β are derived in Section 3.4. The independence of estimators of β and sums of squares that are appeared in the partition for source of variability are proved in Section 3.5. Exact and approximate intervals for β are proposed in Section 3.6 using Theorems in Sections 3.4 and 3.5.

3.2 Regression Model with a Balanced Two-fold Nested Error Structure

The regression model with a balanced two-fold nested error structure is defined as

$$Y_{ijk} = \mu + \beta X_{ijk} + P_i + O_{ij} + E_{ijk} \tag{3.2.1}$$
$$i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, r$$

where Y_{ijk} is the kth random observation within the jth secondary level within the ith primary level, μ and β are unknown constants, X_{ijk} is a fixed predictor variable, P_i, O_{ij}, and E_{ijk} are respectively error terms associated with the first-stage, second-stage, and last-stage sampling unit, and P_i, O_{ij}, and E_{ijk} are jointly independent normal random variables with zero means and variances σ_P^2, σ_O^2, and σ_E^2, respectively. Since β and X_{ijk} are fixed, model (3.1) is a mixed model.

Model (3.2.1) is written in matrix notation as

$$\mathbf{y} = \mathbf{X}\underline{\alpha} + \mathbf{B}_1\mathbf{p} + \mathbf{B}_2\mathbf{o} + \mathbf{B}_3\mathbf{e} \tag{3.2.2}$$

where $\mathbf{y}$ is an $abr \times 1$ random vector of observations, $\mathbf{X}$ is an $abr \times 2$ matrix with a column of 1's in the first column and a column of known X_{ijk}'s in the second column, $\underline{\alpha}$ is a 2×1 vector with elements μ and β, $\mathbf{B}_1 = \overset{a}{\underset{i=1}{\oplus}} \mathbf{1}_{br}$, $\mathbf{B}_2 = \overset{a}{\underset{i=1}{\oplus}} \overset{b}{\underset{j=1}{\oplus}} \mathbf{1}_r$, and $\mathbf{B}_3 = \overset{a}{\underset{i=1}{\oplus}} \overset{b}{\underset{j=1}{\oplus}} \overset{r}{\underset{k=1}{\oplus}} 1 = \mathbf{I}_{abr}$ are design matrices, $\oplus$ is the direct sum operator, $\mathbf{1}_{br}$ and $\mathbf{1}_r$ are respectively $br \times 1$ and $r \times 1$ column vectors of 1's, $\mathbf{I}_{abr}$ is an $abr \times abr$ identity matrix, $\mathbf{p}$ is an $a \times 1$ vector of random P_i effects, $\mathbf{o}$ is an $ab \times 1$ vector of random O_{ij} effects, and $\mathbf{e}$ is an $abr \times 1$ vector of random error terms, E_{ijk}. Under the distributional assumptions of (3.2.1), $\mathbf{y}$ has a multivariate normal distribution with mean $\mathbf{X}\underline{\alpha}$ and covariance matrix $\sigma_P^2\mathbf{B}_1\mathbf{B}_1' + \sigma_O^2\mathbf{B}_2\mathbf{B}_2' + \sigma_E^2\mathbf{I}_{abr}$.

3.3 ANOVA for the Regression Model

In order to form confidence intervals on the regression coefficient, an appropriate set of sums of squares is needed. One possible partitioning for source of variability of model (3.2.1) that is useful for subsampling is shown in Table 3.3.1.

Table 3.3.1
A Partition for Source of Variability of Model (3.2.1)

SV	DF	SS
Among Primaries	n_1+1	S_{yy1}
Among Primaries Regression	1	$\hat{\beta}_1^2 S_{xx1}$
Among Primaries Residual	n_1	R_1
Among Secondaries	n_2+1	S_{yy2}
Among Secondaries Regression	1	$\hat{\beta}_2^2 S_{xx2}$
Among Secondaries Residual	n_2	R_2
Within Secondaries	n_3+1	S_{yy3}
Within Secondaries Regression	1	$\hat{\beta}_3^2 S_{xx3}$
Within Secondaries Residual	n_3	R_3
Adjusted Total	$n_.+3$	S_{yy123}

The notation for the sums of squares and the estimators of β in Table 3.1 is defined as follows:

$$
\begin{aligned}
n_1 =& a-2, \\
n_2 =& a(b-1)-1, \\
n_3 =& ab(r-1)-1, \\
n_. =& n_1+n_2+n_3, \\
\bar{Y}_{ij.} =& \Sigma_k Y_{ijk}/r, \qquad (3.3.1a) \\
\bar{Y}_{i..} =& \Sigma_j \Sigma_k Y_{ijk}/br, \\
\bar{Y}_{...} =& \Sigma_i \Sigma_j \Sigma_k Y_{ijk}/abr, \\
\bar{X}_{ij.} =& \Sigma_k X_{ijk}/r, \\
\bar{X}_{i..} =& \Sigma_j \Sigma_k X_{ijk}/br, \\
\bar{X}_{...} =& \Sigma_i \Sigma_j \Sigma_k X_{ijk}/abr,
\end{aligned}
$$

$$
\begin{aligned}
S_{yy1} &= br\Sigma_i(\bar{Y}_{i..} - \bar{Y}_{...})^2, \\
S_{yy2} &= r\Sigma_i\Sigma_j(\bar{Y}_{ij.} - \bar{Y}_{i..})^2, \\
S_{yy3} &= \Sigma_i\Sigma_j\Sigma_k(Y_{ijk} - \bar{Y}_{ij.})^2, \\
S_{xx1} &= br\Sigma_i(\bar{X}_{i..} - \bar{X}_{...})^2, \\
S_{xx2} &= r\Sigma_i\Sigma_j(\bar{X}_{ij.} - \bar{X}_{i..})^2, \\
S_{xx3} &= \Sigma_i\Sigma_j\Sigma_k(X_{ijk} - \bar{X}_{ij.})^2, \\
S_{xy1} &= br\Sigma_i(\bar{X}_{i..} - \bar{X}_{...})(\bar{Y}_{i..} - \bar{Y}_{...}), \\
S_{xy2} &= r\Sigma_i\Sigma_j(\bar{X}_{ij.} - \bar{X}_{i..})(\bar{Y}_{ij.} - \bar{Y}_{i..}), \\
S_{xy3} &= \Sigma_i\Sigma_j\Sigma_k(X_{ijk} - \bar{X}_{ij.})(Y_{ijk} - \bar{Y}_{ij.}), \\
S_{yy12} &= S_{yy1} + S_{yy2}, \\
S_{yy123} &= S_{yy12} + S_{yy3}, \\
S_{xx12} &= S_{xx1} + S_{xx2}, \\
S_{xx123} &= S_{xx12} + S_{xx3}, \\
S_{xy12} &= S_{xy1} + S_{xy2}, \\
S_{xy123} &= S_{xy12} + S_{xy3}, \\
\hat{\beta}_1 &= S_{xy1}/S_{xx1}, \\
\hat{\beta}_2 &= S_{xy2}/S_{xx2}, \\
\hat{\beta}_3 &= S_{xy3}/S_{xx3}, \\
R_1 &= S_{yy1} - \hat{\beta}_1^2 S_{xx1}, \\
R_2 &= S_{yy2} - \hat{\beta}_2^2 S_{xx2}, \quad \text{and} \\
R_3 &= S_{yy3} - \hat{\beta}_3^2 S_{xx3}.
\end{aligned}
\tag{3.3.1b}
$$

The estimators of β and the sums of squares in Table 3.3.1 are now described in the context of a standard linear regression model. The estimator $\hat{\beta}_1$ is obtained from the least squares regression of $\bar{Y}_{i..}$ on $\bar{X}_{i..}$. The estimator $\hat{\beta}_2$ is obtained from the least squares regression of $\bar{Y}_{ij.}$ on $\bar{X}_{ij.}$ and grouping variables that represent i primary levels. The estimator $\hat{\beta}_3$ is obtained from the least squares regression of Y_{ijk} on X_{ijk} and grouping variables that represent i primary and j secondary levels.

Two more possible estimators of β in model (3.2.1) and the sums of squares associated with the estimators can be obtained. The estimator $\hat{\beta}_S = S_{xy12}/S_{xx12}$ is obtained from the least squares regression of $\bar{Y}_{ij.}$ on $\bar{X}_{ij.}$. The sum of squares R_S is written as $R_S = R_{12} - R_1 - R_2$ where $R_{12} = S_{yy12} - \hat{\beta}_S^2 S_{xx12}$. The estimator $\hat{\beta}_T = S_{xy123}/S_{xx123}$ is obtained from the least squares regression of Y_{ijk} on X_{ijk}. The sum of squares R_T is written as $R_T = R_{123} - R_{12} - R_3$ where $R_{123} = S_{yy123} - \hat{\beta}_T^2 S_{xx123}$.

3.4 Distributional Results of the Estimators of Regression Coefficient

In order to construct confidence intervals on the regression coefficient, the ordinary least square(OLS) estimators of β are examined.

Theorem 3.4.1 *Under the assumptions in (3.2.1), an OLS estimator* $\hat{\beta}_1 \sim N\left(\beta, \quad (br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2)/S_{xx1}\right)$.

Proof. From Section 3.3 define

$$\hat{\beta}_1 = \frac{S_{xy1}}{S_{xx1}} = \frac{br\Sigma_i(\bar{X}_{i..} - \bar{X}_{...})(\bar{Y}_{i..} - \bar{Y}_{...})}{br\Sigma_i(\bar{X}_{i..} - \bar{X}_{...})^2}$$

$$
\begin{aligned}
&= \frac{\Sigma_i(\bar{X}_{i..} - \bar{X}_{...})\bar{Y}_{i..}}{\Sigma_i(\bar{X}_{i..} - \bar{X}_{...})^2} \\
&= \Sigma_i k_i \bar{Y}_{i..}
\end{aligned}
$$

where

$$
k_i = \frac{\bar{X}_{i..} - \bar{X}_{...}}{\Sigma_i(\bar{X}_{i..} - \bar{X}_{...})^2}.
$$

It can be shown by model (3.2.1) that

$$
\begin{aligned}
\bar{Y}_{i..} =& \mu + \beta\bar{X}_{i..} + P_i + \bar{O}_{i.} + \bar{E}_{i..}, \\
E(\bar{Y}_{i..}) =& E(\mu + \beta\bar{X}_{i..} + P_i + \bar{O}_{i.} + \bar{E}_{i..}) \\
=& \mu + \beta\bar{X}_{i..}, \qquad \text{and} \\
V(\bar{Y}_{i..}) =& V(\mu + \beta\bar{X}_{i..} + P_i + \bar{O}_{i.} + \bar{E}_{i..}) \\
=& \sigma_P^2 + \frac{\sigma_O^2}{b} + \frac{\sigma_E^2}{br}.
\end{aligned}
$$

Using the fact $\Sigma_i k_i = 0$, $\Sigma_i k_i \bar{X}_{i..} = 1$, $\Sigma_i k_i^2 = br/S_{xx1}$, one obtains that

$$
\begin{aligned}
E(\hat{\beta}_1) &= E(\Sigma_i k_i \bar{Y}_{i..}) \\
&= \Sigma_i k_i(\mu + \beta\bar{X}_{i..}) \\
&= \beta \qquad \text{and} \\
V(\hat{\beta}_1) &= V(\Sigma_i k_i \bar{Y}_{i..}) \\
&= \Sigma_i k_i^2\left(\sigma_P^2 + \frac{\sigma_O^2}{b} + \frac{\sigma_E^2}{br}\right) \\
&= \frac{br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2}{S_{xx1}}.
\end{aligned}
$$

Since $\hat{\beta}_1$ is a linear combination of the $\bar{Y}_{i..}$ and $\bar{Y}_{i..}$ is normally distributed,

$$\hat{\beta}_1 \sim N\left(\beta, \quad \frac{br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2}{S_{xx1}}\right). \tag{3.4.1}$$

Theorem 3.4.2 *Under the assumptions in (3.2.1), an OLS estimators* $\hat{\beta}_2 \sim N\left(\beta, \quad (r\sigma_O^2 + \sigma_E^2)/S_{xx2}\right)$.

Proof. From Section 3.3 define

$$\hat{\beta}_2 = \frac{S_{xy2}}{S_{xx2}}$$

$$= \frac{r\Sigma_i\Sigma_j(\bar{X}_{ij.} - \bar{X}_{i..})(\bar{Y}_{ij.} - \bar{Y}_{i..})}{r\Sigma_i\Sigma_j(\bar{X}_{ij.} - \bar{X}_{i..})^2}$$

$$= \frac{\Sigma_i\Sigma_j(\bar{X}_{ij.} - \bar{X}_{i..})\bar{Y}_{ij.}}{\Sigma_i(\bar{X}_{ij.} - \bar{X}_{i..})^2}$$

$$= \Sigma_i\Sigma_j l_{ij}\bar{Y}_{ij.}$$

where

$$l_{ij} = \frac{\bar{X}_{ij.} - \bar{X}_{i..}}{\Sigma_i(\bar{X}_{ij.} - \bar{X}_{i..})^2}.$$

It can be shown by model (3.2.1) that

$$\bar{Y}_{ij.} = \mu + \beta\bar{X}_{ij.} + P_i + O_{ij} + \bar{E}_{ij.},$$

$$E(\bar{Y}_{ij.}) = E(\mu + \beta\bar{X}_{ij.} + P_i + O_{ij} + \bar{E}_{ij.})$$

$$= \mu + \beta\bar{X}_{ij.} \quad \text{and}$$

$$V(\bar{Y}_{ij.}) = V(\mu + \beta\bar{X}_{ij.} + P_i + O_{ij} + \bar{E}_{ij.})$$

$$= \sigma_P^2 + \sigma_O^2 + \frac{\sigma_E^2}{r}.$$

Using the fact $\Sigma_j l_{ij} = 0$, $\Sigma_i\Sigma_j l_{ij}\bar{X}_{ij.} = 1$, $\Sigma_i\Sigma_j l_{ij}^2 = r/S_{xx2}$, $2\Sigma_i\Sigma_{j<j'} l_{ij}l_{ij'} = -r/S_{xx2}$, and $Cov(\bar{Y}_{ij.}, \bar{Y}_{ij'.}) = \sigma_P^2$ one obtains that

$$
\begin{aligned}
E(\hat{\beta}_2) &= E(\Sigma_i\Sigma_j l_{ij}\bar{Y}_{ij.}) \\
&= \Sigma_i\Sigma_j l_{ij}(\mu + \beta\bar{X}_{ij.}) \\
&= \beta \quad \text{and} \\
V(\hat{\beta}_2) &= V(\Sigma_i\Sigma_j l_{ij}\bar{Y}_{ij.}) \\
&= \Sigma_i\Sigma_j l_{ij}^2\bar{V}(Y_{ij.}) + 2\Sigma_i\Sigma_{j<j'} l_{ij}l_{ij'}Cov(\bar{Y}_{ij.}, \bar{Y}_{ij'.}) \\
&= \Sigma_i\Sigma_j l_{ij}^2\left(\sigma_P^2 + \sigma_O^2 + \frac{\sigma_E^2}{r}\right) + 2\Sigma_i\Sigma_{j<j'} l_{ij}l_{ij'}\sigma_P^2 \\
&= \frac{r\sigma_O^2 + \sigma_E^2}{S_{xx2}}.
\end{aligned}
$$

Therefore,

$$\hat{\beta}_2 \sim N\left(\beta, \quad \frac{r\sigma_O^2 + \sigma_E^2}{S_{xx2}}\right). \tag{3.4.2}$$

Theorem 3.4.3 *Under the assumptions in (3.2.1), an OLS estimators* $\hat{\beta}_3 \sim N\left(\beta, \quad \sigma_E^2/S_{xx3}\right)$.

Proof. From Section 3.3 define

$$
\begin{aligned}
\hat{\beta}_3 &= \frac{S_{xy3}}{S_{xx3}} \\
&= \frac{\Sigma_i\Sigma_j\Sigma_k(X_{ijk} - \bar{X}_{ij.})(Y_{ijk} - \bar{Y}_{ij.})}{\Sigma_i\Sigma_j\Sigma_k(X_{i..} - \bar{X}_{i..})^2} \\
&= \frac{\Sigma_i\Sigma_j\Sigma_k(X_{ijk} - \bar{X}_{ij.})Y_{ij.}}{\Sigma_i\Sigma_j\Sigma_k(X_{ijk} - \bar{X}_{ij.})^2} \\
&= \Sigma_i\Sigma_j\Sigma_k m_{ijk}Y_{ijk}
\end{aligned}
$$

where

$$m_{ijk} = \frac{X_{ijk} - \bar{X}_{ij.}}{\Sigma_i\Sigma_j\Sigma_k(X_{ijk} - \bar{X}_{ij.})^2}.$$

It can be shown by model (3.2.1) that

$$\begin{aligned} E(Y_{ijk}) &= E(\mu + \beta X_{ijk} + P_i + O_{ij} + E_{ijk}) \\ &= \mu + \beta X_{ijk} \quad \text{and} \\ V(Y_{ijk}) &= V(\mu + \beta X_{ijk} + P_i + O_{ij} + E_{ijk}) \\ &= \sigma_P^2 + \sigma_O^2 + \sigma_E^2. \end{aligned}$$

Using the fact $\Sigma_i\Sigma_j\Sigma_k m_{ijk} = 0$, $\Sigma_i\Sigma_j\Sigma_k m_{ijk}X_{ijk} = 1$, $\Sigma_i\Sigma_j\Sigma_k m_{ijk}^2 = 1/S_{xx3}$, $2\Sigma_i\Sigma_j\Sigma_{k<k'} m_{ijk}m_{ijk'} = -1/S_{xx3}$, $\Sigma_{j<j'}\Sigma_{k,k'} m_{ijk}m_{ij'k'} = 0$, $Cov(Y_{ijk}, Y_{ijk'}) = \sigma_P^2 + \sigma_O^2$, and $Cov(Y_{ijk}, Y_{ij'k'}) = \sigma_P^2$, one obtains that

$$\begin{aligned} E(\hat{\beta}_3) &= E(\Sigma_i\Sigma_j\Sigma_k m_{ijk}Y_{ijk}) \\ &= \Sigma_i\Sigma_j\Sigma_k m_{ijk}(\mu + \beta X_{ijk}) \\ &= \beta \quad \text{and} \\ V(\hat{\beta}_3) &= V(\Sigma_i\Sigma_j\Sigma_k m_{ijk}Y_{ijk}) \\ &= \Sigma_i\Sigma_j\Sigma_k m_{ijk}^2 V(Y_{ijk}) \\ &\quad + 2\Sigma_i\Sigma_j\Sigma_{k<k'} m_{ijk}m_{ijk'}Cov(Y_{ijk}, Y_{ijk'}) \\ &\quad + 2\Sigma_i\Sigma_{j<j'}\Sigma_{k,k'} m_{ijk}m_{ij'k'}Cov(Y_{ijk}, Y_{ij'k'}) \\ &= \frac{\sigma_P^2 + \sigma_O^2 + \sigma_E^2}{S_{xx3}} - \frac{\sigma_P^2 + \sigma_O^2}{S_{xx3}} \\ &= \frac{\sigma_E^2}{S_{xx3}}. \end{aligned}$$

Therefore,

$$\hat{\beta}_3 \sim N\left(\beta, \quad \frac{\sigma_E^2}{S_{xx3}}\right). \tag{3.4.3}$$

Theorem 3.4.4 *Under the assumptions in (3.2.1), an OLS estimators* $\hat{\beta}_S \sim N\left(\beta, \quad (k_{12}br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2)/S_{xx12}\right)$.

Proof. From Section 3.3 define

$$\hat{\beta}_S = \frac{S_{xy12}}{S_{xx12}} = \frac{S_{xx1}\hat{\beta}_1 + S_{xx2}\hat{\beta}_2}{S_{xx12}}.$$

The estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are independent by

$$\begin{aligned} Cov(\hat{\beta}_1, \hat{\beta}_2) &= Cov(\Sigma_i k_i \bar{Y}_{i..}, \Sigma_i \Sigma_j l_{ij} \bar{Y}_{ij.}) \\ &= \Sigma_i k_i \Sigma_j l_{ij} Cov(\bar{Y}_{i..}, \bar{Y}_{ij.}) \\ &= 0. \end{aligned}$$

By use of (3.4.1) and (3.4.2), one obtains that

$$\begin{aligned} E(\hat{\beta}_S) &= \frac{S_{xx1}E(\hat{\beta}_1) + S_{xx2}E(\hat{\beta}_2)}{S_{xx1} + S_{xx2}} \\ &= \beta \quad \text{and} \\ V(\hat{\beta}_S) &= \frac{S_{xx1}^2 V(\hat{\beta}_1) + S_{xx2}^2 V(\hat{\beta}_2) + 2S_{xx1}S_{xx2}Cov(\hat{\beta}_1, \hat{\beta}_2)}{S_{xx12}^2} \\ &= \frac{S_{xx1}(br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2) + S_{xx2}(r\sigma_O^2 + \sigma_E^2)}{S_{xx12}^2} \\ &= \frac{S_{xx12}(\frac{S_{xx1}}{S_{xx12}}br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2)}{S_{xx12}^2} \\ &= \frac{k_{12}br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2}{S_{xx12}} \end{aligned}$$

where

$$k_{12} = \frac{S_{xx1}}{S_{xx12}}.$$

Since $\hat{\beta}_S$ is a linear combination of the $\hat{\beta}_1$ and $\hat{\beta}_2$,

$$\hat{\beta}_S \sim N\left(\beta, \quad \frac{k_{12}br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2}{S_{xx12}}\right). \tag{3.4.4}$$

Theorem 3.4.5 *Under the assumptions in (3.2.1), an OLS estimators* $\hat{\beta}_T \sim N\left(\beta, \quad (k_{13}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2)/S_{xx123}\right)$.

Proof. From Section 3.3 define

$$\begin{aligned}\hat{\beta}_T &= \frac{S_{xy123}}{S_{xx123}} \\ &= \frac{S_{xx1}\hat{\beta}_1 + S_{xx2}\hat{\beta}_2 + S_{xx3}\hat{\beta}_3}{S_{xx123}}.\end{aligned}$$

The estimators $\hat{\beta}_1$ and $\hat{\beta}_3$ are independent by

$$\begin{aligned}Cov(\hat{\beta}_1, \hat{\beta}_3) &= Cov(\Sigma_i k_i \bar{Y}_{i..}, \Sigma_i \Sigma_j \Sigma_k m_{ijk} Y_{ijk}) \\ &= \Sigma_i k_i \Sigma_j \Sigma_k m_{ijk} Cov(\bar{Y}_{i..}, Y_{ijk}) \\ &= 0 \quad \text{and}\end{aligned}$$

$\hat{\beta}_2$ and $\hat{\beta}_3$ are independent by

$$\begin{aligned}Cov(\hat{\beta}_2, \hat{\beta}_3) &= Cov(\Sigma_i \Sigma_j l_{ij} \bar{Y}_{ij.}, \Sigma_i \Sigma_j \Sigma_k m_{ijk} Y_{ijk}) \\ &= \Sigma_i \Sigma_j l_{ij} \Sigma_k m_{ijk} Cov(\bar{Y}_{ij.}, Y_{ijk}) \\ &= 0.\end{aligned}$$

By use of (3.4.1), (3.4.2), and (3.4.3), one obtains that

$$\begin{aligned}E(\hat{\beta}_T) &= \frac{S_{xx1}E(\hat{\beta}_1) + S_{xx2}E(\hat{\beta}_2) + S_{xx3}E(\hat{\beta}_3)}{S_{xx123}} \\ &= \beta \quad \text{and} \\ V(\hat{\beta}_T) &= \frac{1}{S_{xx123}^2}\big[S_{xx1}^2 V(\hat{\beta}_1) + S_{xx2}^2 V(\hat{\beta}_2) + S_{xx3}^2 V(\hat{\beta}_3)\end{aligned}$$

$$
\begin{aligned}
&+ 2S_{xx1}S_{xx2}Cov(\hat{\beta}_1, \hat{\beta}_2) \\
&+ 2S_{xx1}S_{xx3}Cov(\hat{\beta}_1, \hat{\beta}_3) \\
&+ 2S_{xx2}S_{xx3}Cov(\hat{\beta}_2, \hat{\beta}_3)\big]
\end{aligned}
$$

$$
=\frac{1}{S_{xx123}^2}\left[S_{xx1}(br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2) + S_{xx2}(r\sigma_O^2 + \sigma_E^2) + S_{xx3}\sigma_E^2\right]
$$

$$
=\frac{\frac{S_{xx1}}{S_{xx123}}br\sigma_P^2 + \frac{S_{xx12}}{S_{xx123}}r\sigma_O^2 + \sigma_E^2}{S_{xx123}}
$$

$$
=\frac{k_{13}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2}{S_{xx123}}
$$

where

$$
k_{13} = \frac{S_{xx1}}{S_{xx123}} \quad \text{and}
$$
$$
k_{23} = \frac{S_{xx12}}{S_{xx123}}.
$$

Therefore,

$$
\hat{\beta}_T \sim N\left(\beta, \quad \frac{k_{13}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2}{S_{xx123}}\right). \tag{3.4.5}
$$

3.5 Independence of the Estimators and Sums of Squares

It is necessary to show that the OLS estimators in Section 3.4 and the sums of squares in Table 3.3.1 are independent.

Theorem 3.5.1 *Under the distributional assumptions in (3.2.1), $\hat{\beta}_1$ and R_1 are independent.*

Proof. The estimator $\hat{\beta}_1$ is the second element of the vector $(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1$ where $\mathbf{X}_1 = \mathbf{W}_1\mathbf{X}$, $\mathbf{y}_1 = \mathbf{W}_1\mathbf{y}$, and $\mathbf{W}_1 = \frac{1}{br}\mathbf{B}_1'$. The sum of squares R_1 is written in a quadratic form as $R_1 = br\mathbf{y}_1'\mathbf{F}_1\mathbf{y}_1$ where $\mathbf{F}_1 = \mathbf{I}_a - \mathbf{H}_1$ and $\mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$. It can be shown that $\mathbf{y}_1 \sim N(\mathbf{X}_1\underline{\alpha}, \quad \frac{1}{br}(br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2)\mathbf{I}_a)$.

Theorem 7.5 in Searle (1987, p. 233) is applied to show independence. Noting $\mathbf{X}_1'\mathbf{H}_1 = \mathbf{X}_1'$, one obtains that

$$\begin{aligned}
&(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\frac{1}{br}(br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2)\mathbf{I}_a br\mathbf{F}_1 \\
=&(br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2)(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{I}_a - \mathbf{H}_1) \\
=&\mathbf{0}.
\end{aligned}$$

Thus, $\hat{\beta}_1$ and R_1 are independent.

Theorem 3.5.2 *Under the distributional assumptions in (3.2.1), $\hat{\beta}_2$ and R_2 are independent.*

Proof. The estimator $\hat{\beta}_2$ is the second element of the vector $(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_1'\mathbf{y}_2$ where $\mathbf{X}_2 = \mathbf{W}_2[\mathbf{X} \quad \mathbf{B_1}]$, $\mathbf{y}_2 = \mathbf{W}_2\mathbf{y}$, and $\mathbf{W}_2 = \frac{1}{r}\mathbf{B}_1'$. The sum of squares R_2 is written in a quadratic form as $R_2 = r\mathbf{y}_2'\mathbf{S}_1\mathbf{y}_2$ where $\mathbf{S}_1 = \mathbf{I}_{ab} - \mathbf{H}_2$ and $\mathbf{H}_2 = \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_2'$. It can be shown that $\mathbf{y}_2 \sim N(\mathbf{X}_S\underline{\alpha}, \quad \sigma_P^2\mathbf{W}_2\mathbf{B}_1\mathbf{B}_1'\mathbf{W}_2' + \sigma_O^2\mathbf{I}_{ab} + \frac{1}{r}\sigma_E^2\mathbf{I}_{ab})$.

Noting $\mathbf{B}_1'\mathbf{W}_2\mathbf{S}_1 = \mathbf{0}$ and $(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_2'\mathbf{S}_1 = \mathbf{0}$, one obtains that

$$\begin{aligned}
&(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_2'(\sigma_P^2\mathbf{W}_2\mathbf{B}_1\mathbf{B}_1'\mathbf{W}_2' + \sigma_O^2\mathbf{I}_{ab} + \frac{1}{r}\sigma_E^2\mathbf{I}_{ab})r\mathbf{S}_1 \\
=&r\sigma_P^2(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_2'\mathbf{W}_2\mathbf{B}_1\mathbf{B}_1'\mathbf{W}_2'\mathbf{S}_1 + r\sigma_O^2(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_2'\mathbf{S}_1 \\
&+ \sigma_E^2(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_2'\mathbf{S}_1 \\
=&\mathbf{0}.
\end{aligned}$$

Thus, $\hat{\beta}_2$ and R_2 are independent.

Theorem 3.5.3 *Under the distributional assumptions in (3.2.1), $\hat{\beta}_3$ and R_3 are independent.*

Proof. The estimator $\hat{\beta}_3$ is the second element of the vector $(\mathbf{X}_3'\mathbf{X}_3)^{-}\mathbf{X}_3'\mathbf{y}$ where $\mathbf{X}_3 = \mathbf{W}_2[\mathbf{X} \quad \mathbf{B_1} \quad \mathbf{B_2}]$. The sum of squares R_3 is written in a quadratic form as $R_3 = \mathbf{y}'\mathbf{T}_1\mathbf{y}$ where $\mathbf{T}_1 = \mathbf{I}_{abr} - \mathbf{H}_3$ and $\mathbf{H}_3 = \mathbf{X}_3(\mathbf{X}_3'\mathbf{X}_3)^{-}\mathbf{X}_3'$. It can be shown that $\mathbf{y} \sim N(\mathbf{X}\underline{\alpha}, \quad \sigma_P^2\mathbf{B}_1\mathbf{B}_1' + \sigma_O^2\mathbf{B}_2\mathbf{B}_2' + \sigma_E^2\mathbf{I}_{abr})$.

Noting $\mathbf{B}_1\mathbf{B}_1'\mathbf{T}_1 = 0$, $\mathbf{B}_2\mathbf{B}_2'\mathbf{T}_1 = 0$, and $\mathbf{X}_3'\mathbf{H}_3 = \mathbf{X}_3'$, one obtains that

$$\begin{aligned}
&(\mathbf{X}_3'\mathbf{X}_3)^{-}\mathbf{X}_3'(\sigma_P^2\mathbf{B}_1\mathbf{B}_1' + \sigma_O^2\mathbf{B}_2\mathbf{B}_2' + \sigma_E^2\mathbf{I}_{abr})\mathbf{T}_1 \\
=&\sigma_P^2(\mathbf{X}_3'\mathbf{X}_3)^{-}\mathbf{X}_3'\mathbf{B}_1\mathbf{B}_1'\mathbf{T}_1 + \sigma_O^2(\mathbf{X}_3'\mathbf{X}_3)^{-}\mathbf{X}_3'\mathbf{B}_2\mathbf{B}_2'\mathbf{T}_1 \\
&+ \sigma_E^2(\mathbf{X}_3'\mathbf{X}_3)^{-}\mathbf{X}_3'\mathbf{T}_1 \\
=&\mathbf{0}.
\end{aligned}$$

Thus, $\hat{\beta}_3$ and R_3 are independent.

Theorem 3.5.4 *Under the distributional assumptions in (3.2.1), $\hat{\beta}_S$ and R_{12} are independent where $R_{12} = \kappa_1 R_1 + \kappa_2 R_2$, $\kappa_1 = \frac{k_{12}br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2}{br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2}$, and $\kappa_2 = \frac{k_{12}br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2}{r\sigma_O^2 + \sigma_E^2}$.*

Proof. Park (2012) proved that $R_1/(br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2) \sim \chi^2_{n_1}$, $R_2/(r\sigma_O^2 + \sigma_E^2) \sim \chi^2_{n_2}$, and they are independent. Noting that R_{12} is linear combination of R_1 and R_2, one obtains that $R_{12} \sim \chi^2_{ab-3}$. The estimator $\hat{\beta}_S$ is the second element of the vector $(\mathbf{X}_S'\mathbf{X}_S)^{-1}\mathbf{X}_S'\mathbf{W}_2\mathbf{y}$ where $\mathbf{X}_S = \mathbf{W}_2\mathbf{X}$. It can be shown that $\mathbf{y} \sim N(\mathbf{X}\underline{\alpha}, \quad \sigma_P^2\mathbf{B}_1\mathbf{B}_1' + \sigma_O^2\mathbf{B}_2\mathbf{B}_2' + \sigma_E^2\mathbf{I}_{abr})$.

Noting that $\mathbf{X}_S'\mathbf{W}_2\mathbf{B}_1\mathbf{B}_1'\mathbf{W}_1'\mathbf{F}_1 = \mathbf{0}$, $\mathbf{B}_1'\mathbf{W}_2'\mathbf{S}_1 = \mathbf{0}$, $\mathbf{X}_S'\mathbf{W}_2\mathbf{B}_2\mathbf{B}_2'\mathbf{W}_1'\mathbf{F}_1 = \mathbf{0}$, $\mathbf{X}_S'\mathbf{W}_2\mathbf{B}_2\mathbf{B}_2'\mathbf{W}_2'\mathbf{S}_1 = \mathbf{0}$, $\mathbf{X}_S'\mathbf{W}_2\mathbf{W}_1\mathbf{F}_1 = \mathbf{0}$, and $\mathbf{X}_S'\mathbf{W}_2\mathbf{W}_2'\mathbf{S}_1 = \mathbf{0}$, one obtains that

$$(\mathbf{X}_S'\mathbf{X}_S)^{-}\mathbf{X}_S'\mathbf{W}_2(\sigma_P^2\mathbf{B}_1\mathbf{B}_1' + \sigma_O^2\mathbf{B}_2\mathbf{B}_2' + \sigma_E^2\mathbf{I}_{abr})[\kappa_1 br\mathbf{W}_1'\mathbf{F}_1\mathbf{W}_1 + \kappa_2 r\mathbf{W}_2'\mathbf{S}_1\mathbf{W}_2]$$
$$=\mathbf{0}.$$

Thus, $\hat{\beta}_S$ and R_{12} are independent.

Theorem 3.5.5 *Under the distributional assumptions in (3.2.1), $\hat{\beta}_T$ and R_{123} are independent where $R_{123} = \tau_1 R_1 + \tau_2 R_2 + \tau_3 R_3$, $\tau_1 = \frac{k_{13}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2}{br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2}$, $\tau_2 = \frac{k_{13}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2}{r\sigma_O^2 + \sigma_E^2}$, and $\tau_3 = \frac{k_{13}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2}{\sigma_E^2}$.*

Proof. Park (2012) proved that $R_1/(br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2) \sim \chi^2_{n_1}$, $R_2/(r\sigma_O^2 + \sigma_E^2) \sim \chi^2_{n_2}$, $R_3/\sigma_E^2 \sim \chi^2_{n_3}$, and they are independent. Noting that R_{123} is linear combination of R_1, R_2, and R_3, one obtains that $R_{123} \sim \chi^2_{abr-4}$. The estimator $\hat{\beta}_T$ is the second element of the vector $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

Noting that $\mathbf{X}'\mathbf{B}_1\mathbf{B}_1'\mathbf{W}_1'\mathbf{F}_1 = \mathbf{0}$, $\mathbf{X}'\mathbf{B}_2\mathbf{B}_2'\mathbf{W}_1'\mathbf{F}_1 = \mathbf{0}$, $\mathbf{X}_1'\mathbf{W}_1'\mathbf{F}_1\mathbf{W}_1 = \mathbf{0}$, $\mathbf{B}_1'\mathbf{W}_2'\mathbf{S}_1 = \mathbf{0}$, $\mathbf{X}'\mathbf{B}_2\mathbf{B}_2'\mathbf{W}_2'\mathbf{S}_1 = \mathbf{0}$, $\mathbf{X}'\mathbf{W}_2'\mathbf{S}_1 = \mathbf{0}$, $\mathbf{B}_1'\mathbf{T}_1 = \mathbf{0}$, $\mathbf{B}_2'\mathbf{T}_1 = \mathbf{0}$, and $\mathbf{X}_1'\mathbf{T}_1 = \mathbf{0}$, one obtains that

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma_P^2\mathbf{B}_1\mathbf{B}_1' + \sigma_O^2\mathbf{B}_2\mathbf{B}_2' + \sigma_E^2\mathbf{I}_{abr})$$
$$\times [\tau_1 br\mathbf{W}_1'\mathbf{F}_1\mathbf{W}_1 + \tau_2 r\mathbf{W}_2'\mathbf{S}_1\mathbf{W}_2 + \tau_3\mathbf{T}_1]$$
$$=\mathbf{0}.$$

Thus, $\hat{\beta}_T$ and R_{123} are independent.

3.6 Confidence Intervals for Regression Coefficient

The independence shown in Section 3.5 is used to make t distributed random variables with proper degrees of freedom. The confidence intervals for the regression coefficient in the model are constructed using this property. The modified large sample property is also used to derive approximate confidence intervals.

Five chi-squared random variables, R_1, R_1, R_3, R_{12}, and R_{123}, in Section 3.5 are summarized as follows:

$$\frac{R_1}{br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2} \sim \chi^2_{a-2} \tag{3.6.1a}$$

$$\frac{R_2}{r\sigma_O^2 + \sigma_E^2} \sim \chi^2_{a(b-1)-1} \tag{3.6.1b}$$

$$\frac{R_3}{\sigma_E^2} \sim \chi^2_{ab(r-1)-1} \tag{3.6.1c}$$

$$\frac{R_{12}}{k_{12}br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2} \sim \chi_{ab-3} \tag{3.6.1d}$$

$$\frac{R_{123}}{k_{13}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2} \sim \chi^2_{abr-4}. \tag{3.6.1e}$$

In order to construct confidence intervals for regression coefficient it is convenient to summarize expected mean squares using sums of squares defined in Table 3.3.1. In particular,

$$E(S_1^2) = \sigma_E^2 + r\sigma_O^2 + br\sigma_P^2 = \theta_1 \tag{3.6.2a}$$

$$E(S_2^2) = \sigma_E^2 + r\sigma_O^2 = \theta_2 \tag{3.6.2b}$$

$$E(S_3^2) = \sigma_E^2 = \theta_3 \tag{3.6.2c}$$

where $S_1^2 = R_1/n_1$, $S_2^2 = R_2/n_2$, and $S_3^2 = R_3/n_3$.

The confidence intervals for β are now constructed using five Theorems in Section 3.5. Since $\hat{\beta}_1$ and R_1 are independent, an exact $100(1-\alpha)\%$ two-sided confidence interval on β using Theorem 3.5.1 is

$$\hat{\beta}_1 \ \pm \ t_{(\alpha/2 \ : \ n_1)} \sqrt{\frac{S_1^2}{S_{xx1}}}. \tag{3.6.3}$$

where $t_{(\delta:\nu)}$ is the t-value for ν degrees of freedom with δ area to the right. This method is referred to as EX1 method.

Using independence of $\hat{\beta}_2$ and R_2, an exact $100(1-\alpha)\%$ two-sided confidence interval on β using Theorem 3.5.2 is

$$\hat{\beta}_2 \ \pm \ t_{(\alpha/2 \ : \ n_2)} \sqrt{\frac{S_2^2}{S_{xx2}}}. \tag{3.6.4}$$

This method is referred to as EX2 method.

Using independence of $\hat{\beta}_3$ and R_3, an exact $100(1-\alpha)\%$ two-sided confidence interval on β using Theorem 3.5.3 is

$$\hat{\beta}_3 \ \pm \ t_{(\alpha/2 \ : \ n_3)} \sqrt{\frac{S_3^2}{S_{xx3}}}. \tag{3.6.5}$$

This method is referred to as EX3 method.

Using independence of $\hat{\beta}_S$ and R_{12}, an exact $100(1-\alpha)\%$ two-sided confidence interval on β using Theorem 3.5.4 is

$$\hat{\beta}_S \ \pm \ t_{(\alpha/2 \ : \ n_4)} \sqrt{\frac{S_{12}^2}{S_{xx12}}}. \tag{3.6.6}$$

where $S_{12}^2 = R_{12}/n_4$ and $n_4 = ab-3$. This method is referred to as EXS method.

Using independence of $\hat{\beta}_T$ and R_{123}, an exact $100(1-\alpha)\%$ two-sided confidence interval on β using Theorem 3.5.5 is

$$\hat{\beta}_T \ \pm \ t_{(\alpha/2\ :\ n_5)}\sqrt{\frac{S_{123}^2}{S_{xx123}}}. \tag{3.6.7}$$

where $S_{123}^2 = R_{123}/n_5$ and $n_5 = abr - 4$. This method is referred to as EXT method.

From expected mean squares in (3.6.2) the unbiased estimators of variance components, σ_P^2, σ_O^2, and σ_E^2, can be obtained as $(S_1^2 - S_2^2)/br$, $(S_2^2 - S_3^2)/r$, and S_3^2, respectively. By modifying (3.6.6) and using unbiased estimators of the variance components, an approximate $100(1-\alpha)\%$ two-sided confidence interval on β is

$$\hat{\beta}_S \ \pm \ Z_{\alpha/2}\sqrt{\frac{k_{12}S_1^2 + (1-k_{12})S_2^2}{S_{xx12}}}. \tag{3.6.8}$$

where $Z_{\alpha/2}$ is the Z-value with $\alpha/2$ area to the right. This large sample interval is referred to as LSS method.

By modifying (3.6.7) and using unbiased estimators of the variance components, an approximate $100(1-\alpha)\%$ two-sided confidence interval on β is

$$\hat{\beta}_T \ \pm \ Z_{\alpha/2}\sqrt{\frac{k_{13}S_1^2 + (k_{23}-k_{13})S_2^2 + (1-k_{23})S_3^2}{S_{xx123}}}. \tag{3.6.9}$$

This large sample interval is referred to as LST method.

æ

CHAPTER 4

SIMULATION STUDY

The performance of confidence intervals proposed in Chapter 3 is examined using a simulation study. Sixty four designs are formed by taking all combinations of $a = 3, 5, 10, 15$, $b = 2, 5, 10, 15$, and $r = 2, 5, 10, 15$. The values of σ_P^2 are selected from the set of values (0.01, 0.2, 0.4, 0.6, 0.8, 0.98) and the values of σ_O^2 and σ_E^2 are determined to set $\sigma_P^2 + \sigma_O^2 + \sigma_E^2 = 1$. The eighteen sets of specific values for the variance components used in the simulation study are shown in Table 4.1.

Table 4.1

Values for the Variance Components used in Simulation

σ_P^2	σ_O^2	σ_E^2	σ_P^2	σ_O^2	σ_E^2	σ_P^2	σ_O^2	σ_E^2
0.01	0.01	0.98	0.40	0.01	0.59	0.80	0.01	0.19
0.01	0.49	0.50	0.40	0.30	0.30	0.80	0.10	0.10
0.01	0.98	0.01	0.40	0.59	0.01	0.80	0.19	0.01
0.20	0.01	0.79	0.60	0.01	0.39	0.98	0.001	0.019
0.20	0.40	0.40	0.60	0.20	0.20	0.98	0.010	0.010
0.20	0.79	0.01	0.60	0.39	0.01	0.98	0.019	0.001

Recall that the mean squares in Chapter 3 are chi-squared random variables. In particular, $S_1^2 \sim [(br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2)/n_1]\chi^2_{n_1}$, $S_2^2 \sim [(r\sigma_O^2 + \sigma_E^2)/n_2]\chi^2_{n_2}$, $S_3^2 \sim [\sigma_E^2/n_3]\chi^2_{n_3}$, $S_{12}^2 \sim [(k_{12}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2)/n_4]\chi^2_{n_4}$, and $S_{123}^2 \sim [(k_{13}br\sigma_P^2 + k_{23}r\sigma_O^2 + \sigma_E^2)/n_5]\chi^2_{n_5}$. These mean squares are generated by the RANGAM function of the SAS by substituting the specific values in Table 4.1 for each design.

The values of S_{xx1} are selected from the set of values (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8) and the values of S_{xx2} and S_{xx3} are determined to set $S_{xx1}+S_{xx2}+S_{xx3}=1$. Twenty one sets of specific values for the sums of squares used in the simulation study are shown in Table 4.2. The values of the sums of squares are used to calculate the constants k_{12}, k_{13}, and k_{23}.

The OLS estimators of β are generated by the distributional properties. In particular, $\hat{\beta}_1 \sim N(\beta, (br\sigma_P^2 + r\sigma_O^2 + \sigma_E^2)/S_{xx1})$, $\hat{\beta}_2 \sim N(\beta, (r\sigma_O^2 + \sigma_E^2)/S_{xx2})$, $\hat{\beta}_3 \sim N(\beta, \sigma_E^2/S_{xx3})$, $\hat{\beta}_S = k_{12}\hat{\beta}_1 + (1-k_{12})\hat{\beta}_2$, and $\hat{\beta}_T = k_{13}\hat{\beta}_1 + (k_{23}-k_{13})\hat{\beta}_2 + (1-k_{23})\hat{\beta}_3$. These estimators are generated by using RANNOR function of SAS and by substituting the specific values in Table 4.1 and 4.2. Simulated values for S_1^2, S_2^2, S_3^2, S_{12}^2, S_{123}^2, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_S$, and $\hat{\beta}_T$ are substituted into the appropriate formula. The confidence intervals are then computed.

Table 4.2

Values for the Sums of Squares used in Simulation

S_{xx1}	S_{xx2}	S_{xx3}	S_{xx1}	S_{xx2}	S_{xx3}	S_{xx1}	S_{xx2}	S_{xx3}
0.1	0.1	0.8	0.2	0.7	0.1	0.4	0.5	0.1
0.1	0.3	0.6	0.3	0.1	0.6	0.5	0.1	0.4
0.1	0.5	0.4	0.3	0.2	0.5	0.5	0.2	0.3
0.1	0.7	0.2	0.3	0.4	0.3	0.5	0.4	0.1
0.2	0.1	0.7	0.3	0.5	0.2	0.6	0.1	0.3
0.2	0.3	0.5	0.4	0.1	0.5	0.7	0.1	0.2
0.2	0.5	0.3	0.4	0.3	0.3	0.8	0.1	0.1

For each design 2000 iterations are simulated and two-sided confidence intervals on regression coefficient are computed for each proposed method. Confidence coefficients are determined by counting the number of the intervals that contain regression coefficient β. The average lengths of the two-sided confidence intervals are computed.

Tables 4.3 - 4.10 present the results of the simulation for stated 90% confidence intervals on β. The EX1, EX2, EX3, EXS, EXT, LSS, and LST methods refer to the intervals in (3.6.3)-(3.6.9), respectively. Using the normal approximation to the binomial, if the true confidence coefficient is 0.90, there is less than a 2.5% chance that a simulated confidence coefficient based on 2000 replications will be less than 0.88685. The comparison criteria are: i) the ability to maintain the stated confidence coefficient and ii) the average length of two-sided confidence intervals. Although shorter average lengths are preferable, it is necessary that an interval first maintain the stated confidence level.

The EX1, EX2, EX3, EXS, and EXT methods generally maintain the stated confidence level in Table 4.3 - 4.10. The LSS method is too liberal when $(a = 3)$, $(a = 5)$, $(a = 10,\ b = 2 \text{ to } 10)$, and $(a = 15,\ b = 2)$ since the simulated confidence coefficients of the method fall below the 0.88685. The LST method is too liberal when $(a = 3,\ b = 2,\ r = 2)$, $(a = 5,\ b = 2,\ r = 2)$, and $(a = 10,\ b = 2,\ r = 2)$. LST method is preferred to LSS method because LST method utilizes more degrees of freedom n_1, n_2, and n_3 whereas LSS method uses n_1 and n_2 only.

In general, EX3, EXT, and LST methods generate shorter average interval lengths than EX1, EX2, and EXS methods. EX3, EXT, and LST methods are comparable because they utilize all X and Y information from sampling units to subsampling units, which makes their degrees of freedom large. EXT and LST methods are very similar because LST method is a modified large sample interval of EXT method. The more degrees of freedom the method uses, the shorter interval the method generates. Therefore we recommend EX3, EXT, and LST methods in order to derive confidence intervals for regression coefficient β

because they generally keep stated confidence coefficient and yield shorter interval lengths.

Table 4.3

Simulated Confidence Coefficients for 90 % Two-sided Intervals on β

a	b	r		EX1	EX2	EX3	EXS	EXT	LSS	LST
3	2	2	max	0.9175	0.9190	0.9180	0.9175	0.9210	0.8380	0.8825
			min	0.8825	0.8810	0.8780	0.8830	0.8790	0.7770	0.8395
3	2	5	max	0.9210	0.9205	0.9160	0.9180	0.9150	0.8350	0.9040
			min	0.8770	0.8810	0.8750	0.8830	0.8770	0.7735	0.8685
3	2	10	max	0.9170	0.9165	0.9255	0.9200	0.9235	0.8320	0.9165
			min	0.8795	0.8780	0.8740	0.8790	0.8770	0.7735	0.8735
3	2	15	max	0.9180	0.9195	0.9170	0.9175	0.9200	0.8280	0.9160
			min	0.8795	0.8790	0.8765	0.8740	0.8775	0.7750	0.8720
3	5	2	max	0.9190	0.9230	0.9170	0.9200	0.9205	0.8940	0.9070
			min	0.8765	0.8835	0.8805	0.8800	0.8825	0.8525	0.8685
3	5	5	max	0.9190	0.9200	0.9210	0.9225	0.9230	0.8975	0.9200
			min	0.8800	0.8805	0.8765	0.8815	0.8800	0.8540	0.8760
3	5	10	max	0.9195	0.9225	0.9155	0.9215	0.9175	0.8915	0.9155
			min	0.8825	0.8810	0.8780	0.8830	0.8790	0.8540	0.8395
3	5	15	max	0.9175	0.9200	0.9200	0.9195	0.9195	0.8955	0.9170
			min	0.8795	0.8785	0.8810	0.8820	0.8730	0.8535	0.8730
3	10	2	max	0.9165	0.9155	0.9235	0.9200	0.9185	0.9090	0.9120
			min	0.8805	0.8795	0.8795	0.8815	0.8775	0.8650	0.8700
3	10	5	max	0.9205	0.9180	0.9230	0.9185	0.9215	0.9070	0.9195
			min	0.8820	0.8785	0.8820	0.8815	0.8820	0.8690	0.8790
3	10	10	max	0.9170	0.9215	0.9230	0.9185	0.9165	0.9055	0.9160
			min	0.8775	0.8790	0.8785	0.8815	0.8775	0.8680	0.8775
3	10	15	max	0.9160	0.9160	0.9230	0.9180	0.9185	0.9060	0.9185
			min	0.8695	0.8790	0.8805	0.8805	0.8795	0.8665	0.8790
3	15	2	max	0.9180	0.9185	0.9220	0.9195	0.9195	0.9160	0.9165
			min	0.8800	0.8830	0.8835	0.8795	0.8810	0.8715	0.8775
3	15	5	max	0.9210	0.9180	0.9190	0.9160	0.9150	0.9100	0.9135
			min	0.8825	0.8740	0.8780	0.8825	0.8810	0.8750	0.8810
3	15	10	max	0.9225	0.9205	0.9195	0.9185	0.9195	0.9085	0.9195
			min	0.8810	0.8820	0.8785	0.8715	0.8805	0.8650	0.8795
3	15	15	max	0.9175	0.9210	0.9200	0.9150	0.9165	0.9090	0.9165
			min	0.8820	0.8805	0.8800	0.8775	0.8780	0.8710	0.8780

Table 4.4

Simulated Confidence Coefficients for 90 % Two-sided Intervals on β

a	b	r		EX1	EX2	EX3	EXS	EXT	LSS	LST
5	2	2	max	0.9175	0.9190	0.9190	0.9160	0.9210	0.8780	0.9055
			min	0.8770	0.8775	0.8810	0.8815	0.8810	0.8350	0.8560
5	2	5	max	0.9175	0.9240	0.9165	0.9165	0.9195	0.8755	0.9110
			min	0.8740	0.8760	0.8805	0.8790	0.8790	0.8315	0.8730
5	2	10	max	0.9210	0.9190	0.9200	0.9200	0.9195	0.8815	0.9155
			min	0.8825	0.8785	0.8790	0.8825	0.8820	0.8360	0.8790
5	2	15	max	0.9225	0.9200	0.9200	0.9180	0.9215	0.8840	0.9180
			min	0.8785	0.8805	0.8775	0.8800	0.8780	0.8355	0.8765
5	5	2	max	0.9165	0.9155	0.9210	0.9205	0.9185	0.9065	0.9155
			min	0.8735	0.8750	0.8810	0.8845	0.8850	0.8665	0.8775
5	5	5	max	0.9155	0.9180	0.9150	0.9180	0.9250	0.9050	0.9215
			min	0.8720	0.8715	0.8805	0.8790	0.8830	0.8670	0.8810
5	5	10	max	0.9180	0.9160	0.9120	0.9220	0.9220	0.9135	0.9200
			min	0.8775	0.8725	0.8810	0.8805	0.8845	0.8665	0.8825
5	5	15	max	0.9175	0.9170	0.9160	0.9220	0.9200	0.9095	0.9190
			min	0.8815	0.8760	0.8805	0.8785	0.8855	0.8690	0.8850
5	10	2	max	0.9170	0.9200	0.9175	0.9210	0.9245	0.9175	0.9210
			min	0.8775	0.8780	0.8830	0.8790	0.8835	0.8705	0.8810
5	10	5	max	0.9140	0.9165	0.9195	0.9195	0.9815	0.9150	0.9170
			min	0.8835	0.8815	0.8820	0.8820	0.8840	0.8755	0.8825
5	10	10	max	0.9170	0.9170	0.9165	0.9235	0.9220	0.9180	0.9220
			min	0.8820	0.8805	0.8805	0.8815	0.8830	0.8750	0.8815
5	10	15	max	0.9195	0.9190	0.9170	0.9235	0.9235	0.9190	0.9225
			min	0.8795	0.8790	0.8825	0.8795	0.8805	0.8740	0.8790
5	15	2	max	0.9160	0.9185	0.9165	0.9210	0.9170	0.9160	0.9155
			min	0.8835	0.8795	0.8815	0.8770	0.8830	0.8755	0.8810
5	15	5	max	0.9215	0.9190	0.9190	0.9220	0.9175	0.9180	0.9175
			min	0.8835	0.8780	0.8770	0.8785	0.8800	0.8745	0.8770
5	15	10	max	0.9175	0.9200	0.9195	0.9215	0.9195	0.9185	0.9195
			min	0.8825	0.8770	0.8760	0.8830	0.8805	0.8785	0.8805
5	15	15	max	0.9210	0.9205	0.9170	0.9175	0.9180	0.9145	0.9175
			min	0.8825	0.8770	0.8705	0.8815	0.8780	0.8775	0.8760

Table 4.5

Simulated Confidence Coefficients for 90 % Two-sided Intervals on β

a	b	r		EX1	EX2	EX3	EXS	EXT	LSS	LST
10	2	2	max	0.9215	0.9170	0.9210	0.9250	0.9195	0.9045	0.9110
			min	0.8800	0.8820	0.8810	0.8800	0.8800	0.8560	0.8665
10	2	5	max	0.9180	0.9240	0.9170	0.9190	0.9205	0.8995	0.9170
			min	0.8780	0.8785	0.8800	0.8780	0.8795	0.8590	0.8760
10	2	10	max	0.9205	0.9190	0.9210	0.9160	0.9195	0.9005	0.9185
			min	0.8830	0.8760	0.8735	0.8750	0.8715	0.8500	0.8705
10	2	15	max	0.9190	0.9210	0.9225	0.9175	0.9715	0.9005	0.9160
			min	0.8840	0.8795	0.8735	0.8775	0.8790	0.8570	0.8785
10	5	2	max	0.9165	0.9155	0.9210	0.9205	0.9185	0.9065	0.9155
			min	0.8735	0.8750	0.8810	0.8845	0.8850	0.8665	0.8775
10	5	5	max	0.9185	0.9215	0.9200	0.9210	0.9200	0.9140	0.9160
			min	0.8790	0.8770	0.8765	0.8770	0.8780	0.8700	0.8725
10	5	10	max	0.9220	0.9165	0.9160	0.9170	0.9190	0.9115	0.9190
			min	0.8815	0.8830	0.8800	0.8830	0.8795	0.8745	0.8785
10	5	15	max	0.9200	0.9170	0.9200	0.9165	0.9170	0.9100	0.9170
			min	0.8800	0.8800	0.8800	0.8820	0.8835	0.8720	0.8830
10	10	2	max	0.9205	0.9170	0.9185	0.9230	0.9195	0.9190	0.9170
			min	0.8825	0.8695	0.8780	0.8770	0.8825	0.8735	0.8805
10	10	5	max	0.9180	0.9155	0.9165	0.9175	0.9215	0.9150	0.9205
			min	0.8810	0.8810	0.8835	0.8770	0.8805	0.8720	0.8805
10	10	10	max	0.9190	0.9170	0.9170	0.9215	0.9210	0.9200	0.9210
			min	0.8785	0.8835	0.8795	0.8785	0.8810	0.8745	0.8805
10	10	15	max	0.9220	0.9180	0.9185	0.9200	0.9195	0.9185	0.9195
			min	0.8815	0.8830	0.8810	0.8790	0.8805	0.8750	0.8805
10	15	2	max	0.9165	0.9195	0.9195	0.9175	0.9195	0.9160	0.9165
			min	0.8910	0.8930	0.8920	0.8915	0.8920	0.8895	0.8895
10	15	5	max	0.9210	0.9175	0.9185	0.9180	0.9195	0.9160	0.9195
			min	0.8935	0.8925	0.8930	0.8965	0.8975	0.8920	0.8970
10	15	10	max	0.9185	0.9210	0.9190	0.9205	0.9220	0.9175	0.9220
			min	0.8790	0.8805	0.8785	0.8785	0.8780	0.8775	0.8775
10	15	15	max	0.9185	0.9220	0.9215	0.9220	0.9210	0.9180	0.9210
			min	0.8790	0.8795	0.8785	0.8800	0.8795	0.8775	0.8795

Table 4.6

Simulated Confidence Coefficients for 90 % Two-sided Intervals on β

a	b	r		EX1	EX2	EX3	EXS	EXT	LSS	LST
15	2	2	max	0.9175	0.9220	0.9250	0.9195	0.9190	0.9090	0.9115
			min	0.8820	0.8820	0.8785	0.8785	0.8805	0.8670	0.8755
15	2	5	max	0.9255	0.9195	0.9220	0.9230	0.9230	0.9110	0.9205
			min	0.8825	0.8810	0.8775	0.8755	0.8775	0.8615	0.8735
15	2	10	max	0.9190	0.9205	0.9185	0.9205	0.9160	0.9145	0.9160
			min	0.8815	0.8775	0.8805	0.8810	0.8770	0.8675	0.8755
15	2	15	max	0.9190	0.9200	0.9160	0.9165	0.9155	0.9070	0.9145
			min	0.8800	0.8810	0.8805	0.8800	0.8790	0.8665	0.8790
15	5	2	max	0.9160	0.9195	0.9215	0.9190	0.9180	0.9155	0.9170
			min	0.8810	0.8820	0.8800	0.8815	0.8815	0.8780	0.8780
15	5	5	max	0.9175	0.9150	0.9220	0.9210	0.9190	0.9155	0.9175
			min	0.8780	0.8760	0.8780	0.8855	0.8780	0.8800	0.8775
15	5	10	max	0.9175	0.9170	0.9230	0.9245	0.9235	0.9195	0.9230
			min	0.8780	0.8740	0.8775	0.8805	0.8790	0.8770	0.8785
15	5	15	max	0.9205	0.9175	0.9200	0.9220	0.9220	0.9165	0.9220
			min	0.8750	0.8715	0.8775	0.8805	0.8815	0.8750	0.8815
15	10	2	max	0.9200	0.9200	0.9180	0.9165	0.9155	0.9145	0.9145
			min	0.8835	0.8820	0.8795	0.8810	0.8795	0.8785	0.8785
15	10	5	max	0.9175	0.9180	0.9215	0.9180	0.9185	0.9160	0.9185
			min	0.8810	0.8815	0.8805	0.8835	0.8800	0.8810	0.8800
15	10	10	max	0.9155	0.9175	0.9195	0.9165	0.9185	0.9145	0.9185
			min	0.8795	0.8840	0.8810	0.8820	0.8830	0.8810	0.8830
15	10	15	max	0.9160	0.9185	0.9185	0.9195	0.9190	0.9180	0.9190
			min	0.8805	0.8810	0.8790	0.8825	0.8825	0.8815	0.8825
15	15	2	max	0.9185	0.9245	0.9215	0.9165	0.9190	0.9150	0.9190
			min	0.8790	0.8835	0.8835	0.8805	0.8820	0.8790	0.8810
15	15	5	max	0.9190	0.9205	0.9200	0.9145	0.9165	0.9135	0.9165
			min	0.8815	0.8785	0.8820	0.8845	0.8825	0.8830	0.8825
15	15	10	max	0.9195	0.9210	0.9205	0.9160	0.9175	0.9145	0.9170
			min	0.8800	0.8810	0.8805	0.8825	0.8820	0.8820	0.8820
15	15	15	max	0.9205	0.9190	0.9205	0.9155	0.9170	0.9150	0.9170
			min	0.8790	0.8805	0.8805	0.8810	0.8825	0.8805	0.8825

Table 4.7
Average Interval Lengths for 90 % Two-sided Intervals on β

a	b	r		EX1	EX2	EX3	EXS	EXT	LSS	LST
3	2	2	max	64.185	23.018	12.070	13.919	6.4700	9.7290	5.7230
			min	11.470	0.894	0.134	3.439	2.2637	2.4041	2.0023
3	2	5	max	104.14	38.785	10.634	21.825	9.5092	15.254	9.1705
			min	11.66	0.948	0.120	4.706	3.3644	3.289	3.2445
3	2	10	max	144.78	51.428	10.454	30.822	13.286	21.543	13.066
			min	12.76	1.053	0.118	4.890	3.370	3.418	3.314
3	2	15	max	177.80	62.862	10.409	37.723	16.197	26.366	16.022
			min	13.25	1.143	0.117	5.017	3.387	3.506	3.350
3	5	2	max	102.40	15.753	10.822	17.341	9.4695	16.004	9.1322
			min	11.86	0.605	0.123	3.686	2.1603	3.4026	2.0834
3	5	5	max	164.39	24.797	10.447	27.405	14.797	25.291	14.604
			min	12.69	0.646	0.117	3.839	3.351	3.543	3.308
3	5	10	max	230.33	34.941	10.387	38.880	20.777	35.882	20.646
			min	14.13	0.710	0.116	3.997	3.393	3.689	3.371
3	5	15	max	282.89	42.794	10.356	47.604	25.373	43.933	25.267
			min	15.33	0.767	0.116	4.161	3.440	3.840	3.426
3	10	2	max	143.43	15.044	10.553	23.776	13.221	22.961	13.003
			min	11.79	0.584	0.119	3.609	2.171	3.485	2.135
3	10	5	max	227.05	23.778	10.365	37.570	20.743	36.281	20.611
			min	13.53	0.621	0.117	3.791	3.365	3.661	3.343
3	10	10	max	321.53	33.655	10.351	53.159	29.251	51.335	29.160
			min	16.19	0.684	0.116	4.046	3.459	3.907	3.449
3	10	15	max	393.67	41.121	10.339	65.045	35.783	62.813	35.708
			min	17.88	0.741	0.116	4.294	3.551	4.147	3.544
3	15	2	max	176.91	14.883	10.491	28.717	16.119	28.083	15.946
			min	12.46	0.578	0.117	3.619	2.189	3.539	2.166
3	15	5	max	282.20	23.465	10.342	45.652	25.325	44.645	25.219
			min	14.47	0.617	0.116	3.846	3.403	3.761	3.389
3	15	10	max	398.75	33.198	10.332	64.406	35.740	62.985	35.666
			min	17.61	0.681	0.116	4.181	3.538	4.089	3.530
3	15	15	max	487.27	40.679	10.323	78.776	43.761	77.039	43.700
			min	20.36	0.737	0.116	4.430	3.663	4.332	3.658

Table 4.8

Average Interval Lengths for 90 % Two-sided Intervals on β

a	b	r		EX1	EX2	EX3	EXS	EXT	LSS	LST
5	2	2	max	27.788	18.004	11.250	11.561	6.1748	10.037	5.8174
			min	4.897	0.691	0.125	2.891	2.1646	2.510	2.0393
5	2	5	max	44.077	28.317	10.507	18.379	9.4207	15.956	9.2310
			min	5.147	0.741	0.118	3.924	3.3209	3.406	3.2540
5	2	10	max	62.820	39.860	10.398	25.897	13.171	22.483	13.044
			min	5.533	0.812	0.117	4.068	3.339	3.353	3.307
5	2	15	max	77.371	48.990	10.362	31.730	16.121	27.548	16.018
			min	5.830	0.882	0.116	4.193	3.363	3.641	3.342
5	5	2	max	43.490	15.266	10.663	16.852	9.3581	16.143	9.1696
			min	5.149	0.587	0.119	3.605	2.1376	3.454	2.0946
5	5	5	max	69.109	24.038	10.399	26.738	14.690	25.612	14.577
			min	5.562	0.631	0.116	3.719	3.332	3.562	3.297
5	5	10	max	97.000	33.986	10.334	37.670	20.712	36.084	20.636
			min	6.178	0.696	0.116	3.892	3.380	3.728	3.367
5	5	15	max	118.87	41.593	10.319	46.165	25.340	44.222	25.277
			min	6.728	0.753	0.116	4.066	3.432	3.895	3.424
5	10	2	max	61.514	14.896	10.473	23.403	13.164	22.941	13.037
			min	5.318	0.577	0.118	3.592	2.148	3.522	2.127
5	10	5	max	97.147	23.492	10.342	36.974	20.699	36.246	20.621
			min	6.076	0.618	0.116	3.758	3.360	3.684	3.347
5	10	10	max	136.91	33.217	10.317	52.405	29.229	51.372	29.174
			min	7.09	0.679	0.116	4.016	3.454	3.936	3.447
5	10	15	max	167.92	40.708	10.313	64.178	35.773	62.913	35.729
			min	7.95	0.736	0.116	4.255	3.545	4.171	3.540
5	15	2	max	74.957	14.774	10.426	28.504	16.057	28.138	15.955
			min	5.570	0.574	0.117	3.618	2.168	3.571	2.155
5	15	5	max	118.92	23.344	10.327	45.139	25.291	44.558	25.228
			min	6.52	0.613	0.116	3.834	3.398	3.785	3.390
5	15	10	max	168.50	32.998	10.310	63.890	35.737	63.068	35.688
			min	7.82	0.675	0.116	4.166	3.529	4.113	3.524
5	15	15	max	206.50	40.386	10.307	78.326	43.755	77.318	43.719
			min	8.90	0.731	0.116	4.440	3.655	4.349	3.652

Table 4.9

Average Interval Lengths for 90 % Two-sided Intervals on β

a	b	r		EX1	EX2	EX3	EXS	EXT	LSS	LST
10	2	2	max	22.747	15.952	10.770	10.903	6.0032	10.309	5.8487
			min	4.141	0.614	0.121	2.733	2.1282	2.584	2.0734
10	2	5	max	36.100	25.153	10.418	17.281	9.3294	16.339	9.2394
			min	4.307	0.658	0.117	3.697	3.2993	3.495	3.2670
10	2	10	max	50.934	35.625	10.368	24.379	13.142	23.051	13.080
			min	4.587	0.723	0.116	3.819	3.334	3.611	3.318
10	2	15	max	62.351	43.683	10.343	29.841	16.067	28.215	16.016
			min	4.854	0.780	0.116	3.936	3.359	3.722	3.349
10	5	2	max	36.115	14.969	10.479	16.588	9.2996	16.261	9.2098
			min	4.237	0.578	0.117	3.557	2.1294	3.487	2.1089
10	5	5	max	56.695	23.528	10.345	26.225	14.627	25.708	14.572
			min	4.582	0.620	0.116	3.671	3.317	3.598	3.305
10	5	10	max	80.246	33.368	10.320	37.043	20.661	36.312	20.622
			min	5.096	0.681	0.116	3.849	3.339	3.773	3.366
10	5	15	max	98.258	40.825	10.316	45.307	25.285	44.414	25.254
			min	5.572	0.737	0.116	4.012	3.427	3.393	3.423
10	10	2	max	50.814	14.774	10.417	23.241	13.081	23.019	13.019
			min	4.449	0.573	0.117	3.570	2.150	3.536	2.140
10	10	5	max	80.159	23.344	10.327	36.782	20.657	36.431	20.619
			min	5.016	0.613	0.116	3.736	3.358	3.701	3.352
10	10	10	max	113.66	33.034	10.312	52.083	29.182	51.586	29.154
			min	5.84	0.674	0.116	3.987	3.450	3.948	3.447
10	10	15	max	139.54	40.405	10.309	63.789	35.730	63.179	35.708
			min	6.57	0.730	0.116	4.225	3.542	4.185	3.540
10	15	2	max	62.017	14.728	10.378	28.389	15.993	28.210	15.943
			min	4.618	0.572	0.116	3.595	2.171	3.572	2.164
10	15	5	max	98.705	23.276	10.321	44.917	25.277	44.634	25.246
			min	5.38	0.610	0.116	3.809	3.397	3.785	3.393
10	15	10	max	139.59	32.857	10.312	63.550	35.721	63.149	35.699
			min	6.48	0.671	0.116	4.138	3.528	4.112	3.526
10	15	15	max	171.09	40.217	10.309	77.806	43.736	77.315	43.718
			min	7.43	0.727	0.116	4.386	3.656	4.358	3.654

Table 4.10

Average Interval Lengths for 90 % Two-sided Intervals on β

a	b	r		EX1	EX2	EX3	EXS	EXT	LSS	LST
15	2	2	max	21.993	15.474	10.590	10.726	5.9321	10.358	5.8340
			min	3.946	0.597	0.119	2.695	2.1108	2.602	2.0759
15	2	5	max	34.801	24.417	10.380	16.958	9.3025	16.376	9.2435
			min	4.136	0.636	0.116	3.628	3.2985	3.503	3.2776
15	2	10	max	49.063	34.587	10.340	23.916	13.122	23.096	13.081
			min	4.423	0.703	0.116	3.439	3.331	3.615	3.213
15	2	15	max	60.033	42.423	10.326	29.252	16.055	28.248	16.021
			min	4.682	0.760	0.116	3.864	3.360	3.731	3.353
15	5	2	max	34.840	14.847	10.404	16.509	9.2657	16.297	9.2069
			min	4.110	0.573	0.117	3.598	2.1281	3.494	2.1146
15	5	5	max	54.921	23.441	10.323	26.056	14.607	25.720	14.571
			min	4.436	0.612	0.116	3.653	3.332	3.606	3.313
15	5	10	max	77.826	33.129	10.313	36.811	20.653	36.337	20.628
			min	4.931	0.674	0.116	3.825	3.370	3.775	3.367
15	5	15	max	95.163	40.523	10.312	45.104	25.292	44.524	25.271
			min	5.380	0.730	0.116	3.988	3.428	3.936	3.425
15	10	2	max	49.150	14.714	10.376	23.215	13.073	23.069	13.032
			min	4.272	0.572	0.116	3.556	2.145	3.534	2.139
15	10	5	max	77.440	23.227	10.323	36.601	20.642	36.371	20.616
			min	4.826	0.611	0.116	3.719	3.359	3.696	3.354
15	10	10	max	109.87	32.796	10.314	51.801	29.176	51.474	29.158
			min	5.63	0.672	0.116	3.976	3.453	3.951	3.450
15	10	15	max	134.32	40.165	10.309	64.472	35.727	63.072	35.712
			min	6.33	0.727	0.116	4.213	3.545	4.187	3.544
15	15	2	max	59.945	14.678	10.352	28.319	15.995	28.201	15.962
			min	4.461	0.571	0.116	3.590	2.168	3.575	2.164
15	15	5	max	95.131	23.168	10.311	44.769	25.254	44.582	25.233
			min	5.210	0.610	0.116	3.804	3.398	3.788	3.395
15	15	10	max	134.48	32.741	10.304	63.315	35.712	63.051	35.697
			min	6.27	0.671	0.116	4.135	3.530	4.118	3.528
15	15	15	max	164.67	40.086	10.303	77.551	43.735	77.227	43.723
			min	7.18	0.726	0.116	4.382	3.657	4.364	3.656

æ

CHAPTER 5
NUMERICAL EXAMPLE AND CONCLUSIONS

Belsley et al. (1980) analyzed the relationship of house prices on quality of the environment using 506 observations on census tracts belonging to 92 towns in the Boston Standard Metropolitan Statistical Area(SMSA) in 1970. The response variable used in the study is logarithm of the median value of owner-occupied homes(Y) and one of the predictor variables is the per capita crime rate(X).

A subset of these data was selected to conform to the design with $a = 3$, $b = 2$, and $r = 5$ of our simulation. We selected six towns(Salem, Woburn, Natick, Winchester, Belmont, and Arlington) and 5 observations from each town. We assume that Salem and Woburn, Natick and Winchester, and Belmont and Arlington belong to the same district, respectively, and that observations are nested within towns(secondary units) that are nested within districts(primary units) in order to correspond with a balanced two-fold nested error structure.

The data chosen are in Table 5.1. Using the data in Table 5.1 we computed necessary statistics as follows: $\hat{\beta}_1 = -3.484$, $\hat{\beta}_2 = -7.103$, $\hat{\beta}_3 = -3.072$, $\hat{\beta}_S = -5.666$, $\hat{\beta}_T = -4.493$, $S_1^2 = 0.0163$, $S_2^2 = 0.3068$, $S_3^2 = 0.0199$, $S_S^2 = 0.0820$, $S_T^2 = 0.0796$, $k_{12} = 0.3970$, $k_{13} = 0.2174$, and $k_{23} = 0.5477$. The resulting 90% two-sided confidence intervals for β are shown in Table 5.2. LSS method is not recommended because it did not maintain stated confidence level when $a = 3$, $b = 2$, and $r = 5$ in the simulation study. EX3, EXT, and LST methods are recommended to construct 90% two sided confidence interval for β. EX3 and EXT methods use t-values with $n_3 = 23$ and $n_5 = 26$ degrees of freedom, respectively, and LST uses Z-value. The confidence intervals of the three methods contain negative values for lower and upper limits. Therefore the null hypothesis $H_0 : \beta = 0$ is not rejected at $\alpha = 10\%$

Table 5.1

Selected Data Set from Boston SMSA in 1970

Salem		Woburn		Natick	
Y	X	Y	X	Y	X
10.0389	0.08829	10.3735	0.07875	10.0732	0.08244
10.2073	0.14455	10.3023	0.12579	10.0562	0.09252
9.71112	0.21124	10.4602	0.08370	9.99880	0.11329
9.84692	0.17004	10.5187	0.09068	9.90848	0.10612
9.61581	0.22489	10.3255	0.06911	10.0078	0.10290
Winchester		**Belmont**		**Arlington**	
Y	X	Y	X	Y	X
10.2541	0.12204	10.5241	0.05780	10.0732	0.13914
9.97115	0.11504	10.5916	0.06588	10.0690	0.09178
10.5636	0.12083	10.4968	0.06888	10.0257	0.08447
10.6874	0.08187	10.5427	0.09103	10.2888	0.06664
10.4103	0.06860	10.3890	0.10008	10.0519	0.07022

Table 5.2

90% Confidence Intervals on β for Example Data

Method	Estimates	Lower bound	Upper bound	Interval length
EX1	$\hat{\beta}_1 = -3.484$	-11.40	4.43	15.83
EX2	$\hat{\beta}_2 = -7.103$	-19.98	5.78	25.76
EX3	$\hat{\beta}_3 = -3.072$	-4.72	-1.43	3.29
EXS	$\hat{\beta}_S = -5.666$	-12.03	0.70	12.73
EXT	$\hat{\beta}_T = -4.493$	-7.13	-1.86	5.27
LSS	$\hat{\beta}_S = -5.666$	-10.12	-1.22	8.90
LST	$\hat{\beta}_T = -4.493$	-7.03	-1.95	5.08

In summary this thesis presents an approach for constructing confidence intervals for regression coefficient in a simple linear regression model with a balanced two-fold nested error structure. We used OLS estimators to construct confidence intervals for β and performed simulations to compare the confidence intervals derived for primary sampling units a from 3 to 15, secondary sampling units b from 2 to 15, and the last sampling units r from 2 to 15. The simulation study was therefore performed 64 combinations for different values of a, b, and r.

When $a = 3,\ 5,$ and 10, LSS method is not recommended to compute confidence interval. For $a = 3,\ 5,$ and 10 with $b = 2$ and $r = 2$, LST method is not recommended. Except other cases, LSS and LST methods can be applied to construct confidence interval. Except for LSS and LST methods, other five methods can be used across all combinations of a, b, and r. We therefore recommend to compute all seven confidence intervals for β and choose the shortest confidence interval for regression coefficient to apply for real examples.

REFERENCES

Aitkin, M. and Longford, N.T. (1986), "Statistical modelling issues in school effectiveness studies (with discussion)," *Journal of the Royal Statistical Society Series A*, 14A, 1-43.

Belsley, D. A., Kuh, E., and Welsch, E. R. (1980), *"Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*," John Wiley & Sons, NY.

Burdick, R.K. and Sielken Jr., R.L. (1978), "Exact confidence intervals for linear combinations of variance components in nested classifications," *Journal of the American Statistical Association*, 73, 632-635.

Burdick, R. K., Borror, C. M., and Montgomery, D. C. (2005), *"Design and Analysis of Gauge R & R Studies,"* SIAM, Inc.

Burdick, R. K. and Graybill, F. A. (1992), *"Confidence Intervals on Variance Components,"* Marcel Dekker, Inc.

Goldstein, H. and McDonald, R.P. (1988), "A general model for the analysis of multilevel data," *Psychometrika*, 54, 4, 455-467.

Graybill, F.A. and Wang, C.M. (1980), "Confidence intervals on nonnegative linear combinations of variances," *Journal of the American Statistical Association*, 75, 869-873.

Longford, N.T. (1985), "Statistical Modelling of Data from Hierarchical Structures Using Variance Component Analysis," in *Generalized Linear Models,* Proceedings, Lancaster, Lecture Notes in Statistics 32, eds. R. Gilchrist, B. Francis and J. Whittaker, 112-9. Berlin: Springer-Verlag.

Longford, N.T. (1987), "A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects," *Biometrika*, 74, 4, 817-827.

Park, D. J. and Burdick, R. K.(1993), "Confidence intervals on the among group variance component in a simple linear regression model with nested error structure." *Communications in Statistics - Theory and Methods*, 22, 12, 3435-3452.

Park, D. J. and Burdick, R. K.(1994), "Confidence intervals on the regression coefficient in a simple linear regression model with nested error structure." *Communications in Statistics - Simulation and Computation*, 23, 1, 43-58.

Park, D. J. and Hwang, H. M.(2002), "Confidence intervals for the mean response in the simple linear regression model with balanced nested error structure." *Communications in Statistics - Theory and Methods*, 31, 1, 107-117.

Park, D. J. and Burdick, R. K.(2003), "Performance of confidence intervals in regression model with unbalanced one-fold nested error structures." *Communications in Statistics - Simulation and Computation*, 32, 3, 717-732.

Park, D. J. and Burdick, R. K.(2004), "Confidence intervals on total variance in a regression model with an unbalanced one-fold nested error structure." *Communications in Statistics - Theory and Methods*, 33, 1, 2735-2744.

Park, D. J. and Yoon, M.(2009), "Optimal confidence intervals on the variation for operators and gauge in manufacturing process with repeated measurements." *Optimization Engineering*, 10, 273-287.

Park, D. J. (2013), "Alternative Confidence Intervals on Variance Components in a Simple Regression Model with a Balanced Two-fold Nested Error Structure," *Communications in Statistics - Theory and Methods* (will be published).

Satterthwaite, F.E. (1941), "Synthesis of variance," *Psychometrika*, 6, 309-316.

Satterthwaite, F.E. (1946), "An approximate distribution of estimates of variance components," *Biometrics Bulletin*, 2, 110-114.

Searle, S. R. (1987), "*Linear Models for Unbalanced Data*", John Wiley & Sons, Inc.

Welch, B.L. (1956), "On linear combinations of several variances," *Journal of the American Statistical Association*, 51, 132-148.