이 학 석 사 학 위 논 문

Multilevel Modeling by Ridge Regression

능형회귀를 통한 다수준 모델링

2013년 2월

서울대학교 자연과학대학원

통계학과

김 서 현

Multilevel Modeling by Ridge Regression

능형회귀를 통한 다수준 모델링

지도교수 오 희 석

이 논문을 이학석사 학위논문으로 제출함
2012년 10월

서울대학교 자연과학대학원
통계학과
김 서 현

김서현의 이학석사 학위논문을 인준함
2012년 12월

위 원 장	전 종 우	(인)
부위원장	오 희 석	(인)
위　　원	장 원 철	(인)

Multilevel Modeling by Ridge Regression

by

Seohyun Kim

A Thesis
submitted in fulfillment of the requirement
for the degree of
Master of Science
in
Statistics

The Department of Statistics
College of Natural Sciences
Seoul National University
February, 2013

Abstract

Multilevel models are used extensively in social and behavioral science research because the models are able to accept hierarchical data structures. However, when multicollinearity among fixed effects of the model exists, multicollinearity may lead to imprecise coefficient estimates. We investigate a new method of estimating fixed effect coefficients in multilevel model when multicollinearity exists. The proposed method of estimating parameters is based on ridge regression. We apply this method to student assessment data and compare the results with an existing method. The proposed method provides coefficient estimates which have smaller variance than the existing method. Furthermore, we present PRESS statistic which is adapted to the proposed method. Results suggest that the proposed method predicts data better than the existing method.

Keywords: Multilevel model, Ridge regression, Multicollinearity, EM algorithm

Student Number: 2011-20239

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In social science research, multilevel models are often used as an analysis tool because multilevel modeling can reflect social science data which have nested structures. Multilevel structures are popular in social science disciplines. For example, suppose that we are interested in students' performances based on their test scores. The students are nested within schools, and they share the same factors of their schools such as their teachers' characteristics or learning environments. Therefore, we cannot assume that individual student test scores are observed independently and have a common variance. In this respect, we should not use ordinary linear regression to analyze data. Multilevel modeling is a more suitable tool for handling the heterogeneity of data.

However, estimation of fixed effect coefficients in a multilevel model could be incorrect when multicollinearity exists (Shieh and Fouladi, 2003; Kreft and de Leeuw, 1998). Standard errors of parameter estimates could be large (Kreft and de Leeuw, 1998), which can cause imprecise parameter estimation and influence analysis of the data. In general, centering of variables is recommended when multicollinearity exists since it can remove unnecessary dependency among variables (Raudenbush and Byrk, 2002; Kreft and de Leeuw, 1998;

Bickel, 2007). In spite of centering variables, the dependency could remain. In the case that dependency remains, we need another method of handling multicollinearity, and it leads us to investigate a new method of estimating fixed effect coefficients when multicollinearity exists.

This paper is organized as follows. In Chapter 2, we review ridge regression and multilevel models. Multilevel modeling by ridge regression, which is for handling multicollinearity, is introduced in Chapter 3. Chapter 4 shows advantages of the proposed method through the Programme for International Student Assessment 2003 (PISA 2003) data.

Chapter 2

Overview

2.1 Multicollinearity

If there is no linear relationship between regressors, we say that the regressors are orthogonal. When regressors are orthogonal, it is relatively easy to interpret a multiple regression model. However, when one tries to analyze data, it is very common that regressors are not orthogonal. If the regressors are nearly linearly related, it can lead to imprecise inference of the regression model.

Consider the multiple regression model

$$Y = X\beta + e, \qquad e \sim N(0, \sigma^2 I),$$

where Y is an $n \times 1$ vector of responses, X is an $n \times p$ matrix of regressor variables, β is a $p \times 1$ vector of unknown coefficients, and e is an $n \times 1$ vector of random errors. Let the jth column of the X be denoted X_j so that $X = [X_1, X_2, \ldots, X_p]$. The column vectors of the X are said to be linearly dependent if there exist constants $a_1, \ldots, a_p$, which are not all zero, such that

$$\sum_{j=1}^{p} a_j X_j = 0,$$

for a subset of column vectors of X. In this case, the rank of the X^TX matrix is less than p and thus, $(X^TX)^{-1}$ does not exist. Suppose that a near-linear relationship among a subset of X columns exists. In such case, we say that multicollinearity, which causes the matrix X^TX to be ill-conditioned, exists among the regressors. Estimating parameter vector $\hat{\beta}$ using the least squares method, $\hat{\beta}^{LSE} = (X^TX)^{-1}X^TY$, and the corresponding standard error could be highly unreliable if X^TX is ill-conditioned. Therefore, if multicollinearity exists, another way of estimating parameters, other than the least squares criterion, is required.

Detecting multicollinearity

There are several ways to detect multicollinearity (Montgomery *et al.*, 2006). First, we can detect multicollinearity by examining the off-diagonal term of X^TX in correlation form. If regressors x_i and x_j are nearly linearly related, the absolute value of off-diagonal term $|r_{ij}|$ will be close to unity. However, this diagnostic method is not appropriate for detecting multicollinearity in anything more complex than two regressors.

Second, we can use variance inflation factors for detecting multicollinearity. The diagonal elements C_{jj} of the inverse matrix of X^TX in correlation form are called variance inflation factors (VIFs), which are given by

$$\mathrm{VIF}_j = C_{jj} = \frac{1}{1 - R_j^2},$$

where R_j^2 is the coefficient of determination obtained when x_j is regressed on the remaining $p - 1$ regressors. Typically, if one or more VIFs exceed 10 or larger, it indicates that multicollinearity among regressors exists and that the corresponding regression coefficients are poorly estimated.

Third, eigenvalues of X^TX in correlation form, for instance, $k_1, \ldots, k_p$, can be used to detect multicollinearity. If some regressors are nearly linearly related, one or more eigenvalues of the matrix X^TX will be small. In this respect, the condition number of X^TX is frequently used to diagnose multicollinearity, which is defined by

$$l = \frac{k_{max}}{k_{min}}.$$

Generally, if l is 1000 or larger, severe multicollinearity is indicated.

2.2 Ridge regression

Hoerl and Kennard (1970a) proposed a biased estimator called ridge regression as an alternative to the least squares estimator. The ridge estimator $\hat{\beta}^R$ is defined by

$$\hat{\beta}^R = (X^TX + \lambda I)X^TY.$$

The $\hat{\beta}^R$ can be expressed as

$$\hat{\beta}^R = \frac{1}{\lambda + 1}\hat{\beta}^{LSE}.$$

As shown in Figure 2.1, if $\lambda > 0$, applying ridge regression has the effect of shrinking $\hat{\beta}^{LSE}$ toward the origin. As λ close to 0, we obtain the least squares estimator, $\hat{\beta}^R = \hat{\beta}^{LSE}$, and as λ goes to infinity, we have $\hat{\beta}^R = 0$. Although $\hat{\beta}^R$ is a biased estimator of β, Hoerl and Kennard (1970a) proved that there always exists a λ such that

$$\text{MSE}(\hat{\beta}^R) \leq \text{MSE}(\hat{\beta}^{LSE}).$$

Choice of ridge parameter λ

The elements of $\hat{\beta}^R$ are determined by the ridge parameter λ. We will briefly review two ways of choosing a ridge parameter. The first way is to use a ridge trace, which is the plot of the components of $\hat{\beta}^R$ against λ. Hoerl and

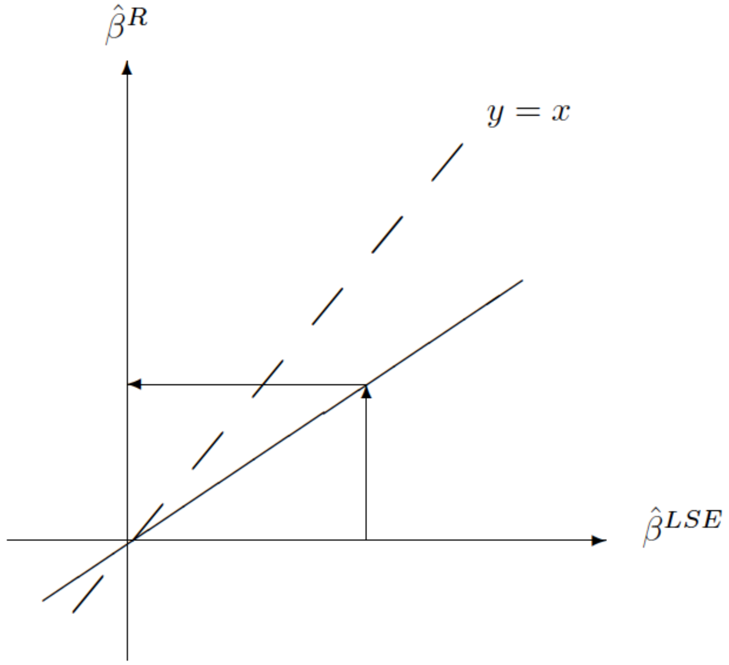

Figure 2.1: The effect of shrinking estimates toward zero

Kennard (1970b) suggested that an appropriate value of λ could be determined by choosing a sufficiently small λ at which the ridge estimates $\hat{\beta}^R$ are stable in the ridge trace.

Second, Golub, Heath, and Wahba (1979) suggested choosing a ridge parameter by using a generalized cross-validation (GCV) error, which is given by

$$\text{GCV} = \frac{1}{n} \sum_i \left(\frac{y_i - \hat{y}_i}{1 - tr(H)/n} \right)^2,$$

where $H = X(X^T X + \lambda I)^{-1} X^T$. We can choose λ so that it minimizes the GCV.

2.3 Multilevel models

Multilevel models are discussed in various literatures under a variety of titles such as *mixed effect models* and *random effects models* (Laird and Ware, 1982), *variance component models* (Longford, 1993), *random coefficient models* (de Leeuw and Kreft, 1986), and *hierarchical linear models* (Bryk and Raudenbush, 2002). Also, multilevel models can be considered a part of *generalized linear models* (McCullagh and Nelder, 1989). In this paper, we adopt the term multilevel model because it reflects data with a nested structure. As indicated by the word, multilevel, the model requires at least two levels of data structure.

Following the notation of Bryk and Raudenbush (2002), consider a simple model of the form,

$$y_{ij} = \beta_{0i} + \beta_{1i} X_{ij} + e_{ij} \tag{2.3.1}$$

$$\beta_{0i} = \beta_{00} + \beta_{01} W_i + u_{0i}, \quad \beta_{1i} = \beta_{10} + \beta_{11} W_i + u_{1i}. \tag{2.3.2}$$

We call (2.3.1) the level-1 model, and (2.3.2) the level-2 model. The index j represents level-1 units, i represents level-2 units, X_{ij} represents a level-1 predictor, and W_i represents a level-2 predictor. The combined form of the level-1 model and the level-2 model is

$$Y_{ij} = \underbrace{\beta_{00} + \beta_{10} X_{ij} + \beta_{01} W_i + \beta_{11} X_{ij} W_i}_{fixed\ effects} + \underbrace{u_{0i} + u_{1i} X_{ij} + e_{ij}}_{random\ effects}, \tag{2.3.3}$$

where

$$E(e_{ij}) = 0, \qquad Var(e_{ij}) = \sigma^2,$$

$$E \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad Var \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = \mathbf{D},$$

and

$$Cov(u_{0j}, e_{ij}) = Cov(u_{1j}, e_{ij}) = 0.$$

As shown in (2.3.3), the model has two subparts: fixed effects and random effects. The model (2.3.3) has more than one residual to obtain, and thus, we cannot simply use the ordinary least squares method to obtain the fixed effect coefficient estimates. In order to obtain estimates of the parameters, there are several methods in which we can apply. We can use a maximum likelihood procedure, a restricted maximum likelihood procedure via an EM algorithm (Laird and Ware, 1982) or iterative generalized least squares (Goldstein, 1986). Without loss of generality, two-level-multilevel models can be extended to multilevel models of three or more levels.

Chapter 3

Multilevel Modeling by Ridge Regression

3.1 Model

We consider a two-level-multilevel model to deal with the problem caused by multicollinearity among fixed effects. Ridge regression is conducted in the process of estimating fixed effect coefficients.

For each i ($i = 1, \ldots, m$), consider the level-1 model

$$Y_i = X_i^1 \beta_i + e_i,$$

and the level-2 model

$$\beta_i = W_i \gamma + u_i,$$

where Y_i is an $n_i \times 1$ vector, X_i^1 is a known $n_i \times q$ matrix, β_i is a $q \times 1$ vector of level-1 coefficients, and e_i is an $n_i \times 1$ vector. Here e_i is normally distributed with a mean of 0 and an $n_i \times n_i$ positive-definite covariance matrix $\sigma^2 I_{n_i}$ (I_{n_i} is an identity matrix). The e_i depends on the ith dimension n_i but is independent of i. Also, W_i is a $q \times p$ matrix, γ is a $p \times 1$ vector, and u_i is a $q \times 1$ vector.

Here, u_i has normal distribution with a mean of 0 and a $q \times q$ positive-definite covariance matrix D. The u_is are also independent of each other and of e_i.

The combination of the level-1 model and the level-2 model becomes

$$Y_i = X_i^1 W_i \gamma + X_i^1 u_i + e_i. \tag{3.1.1}$$

If we define new variables by letting

$$X_i = X_i^1 W_i, \quad \alpha = \gamma, \quad b_i = u_i, \quad Z_i = X_i^1,$$

then the combined model (3.1.1) becomes

$$Y_i = X_i \alpha + Z_i b_i + e_i, \tag{3.1.2}$$

which is the same as a linear mixed effect model (Laird and Ware, 1982). $X_i \alpha$ and $Z_i b_i + e_i$ contain fixed effects and random effects of the model, respectively. For each i, we have

$$Var(Y_i) = Z_i D Z_i^T + \sigma^2 I_{n_i} = V_i. \tag{3.1.3}$$

We can express all of the data in a matrix form by letting

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix},$$

$$\mathbf{\Sigma} = diag(\sigma^2 I_{n_1}, \sigma^2 I_{n_2}, \ldots, \sigma^2 I_{n_m}), \quad \mathbf{D} = diag(D, D, \ldots, D),$$

and

$$\mathbf{Z} = diag(Z_1, Z_2, \ldots, Z_m).$$

Thus, we write the combined model (3.1.2) for all of the data as follows:

$$\mathbf{Y} = \mathbf{X}\alpha + \mathbf{Z}\mathbf{b} + \mathbf{e}. \tag{3.1.4}$$

3.2 Ridge regression with the generalized least squares method

Consider a linear regression model

$$Y = X\alpha + \varepsilon, \tag{3.2.1}$$

with $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2 V$, and $K^T K = KK^T = V$ for some K. If we define new variables by $H = K^{-1}Y$, $B = K^{-1}X$, and $g = K^{-1}\varepsilon$, the model (3.2.1) follows:

$$H = B\alpha + g, \tag{3.2.2}$$

where $Var(g) = E\{[g-E(g)][g-E(g)]^T\} = \sigma^2 I$. The ridge estimator of (3.2.2) is

$$\begin{aligned}\hat{\alpha}^R &= [B^T B + \lambda I]^{-1} B^T H \\ &= [X^T (K^{-1})^T K^{-1} X + \lambda I]^{-1} X^T (K^{-1})^T K^{-1} Y \\ &= [X^T V^{-1} X + \lambda I]^{-1} X^T V^{-1} Y. \end{aligned} \tag{3.2.3}$$

3.3 Parameter estimation with ridge regression using the EM algorithm

In model (3.1.4), if covariance matrices of **e** and **b** are known, we can obtain estimate of α by using ridge regression. By (3.1.3) and (3.2.3),

$$\hat{\alpha}^R = \left[\sum_{i=1}^{m} X_i^T W_i X_i + \frac{\lambda}{\sigma^2} I_p\right]^{-1} \sum_{i=1}^{m} X_i^T W_i Y_i,$$

where $W_i = V_i^{-1}$ and λ is a ridge parameter. If $\lambda = 0$, $\hat{\alpha}^R$ is the same as the ordinary fixed effect coefficient estimates, which is applied the generalized least squares criterion (Laird and Ware, 1982). b_i can be estimated by

$$\hat{b}_i = DZ_i^T W_i (Y_i - X_i \hat{\alpha}),$$

which is empirical Bayes. Since both $\hat{\alpha}^R$ and $\hat{b}_i$ are linear functions of Y, expressions for their standard errors are

$$Var(\hat{\alpha}) = \Big[\sum_{i=1}^{m} X_i^T W_i X_i + \frac{\lambda}{\sigma^2} I_p\Big]^{-1} \sum_{i=1}^{m} X_i^T W_i X_i \Big[\sum_{i=1}^{m} X_i^T W_i X_i + \frac{\lambda}{\sigma^2} I_p\Big]^{-1}$$

and

$$Var(\hat{b}_i) = D Z_i^T \Big\{ W_i - W_i X_i \Big(\sum_{i=1}^{m} X_i^T W_i X_i\Big)^{-1} X_i^T W_i \Big\} Z_i D.$$

When covariance matrices of $\mathbf{e}$ and $\mathbf{b}$ are unknown, we can estimate parameters by using the EM algorithm (Laird and Ware, 1982). Consider θ be a k–vector of variance and covariance parameters which is found in $\sigma^2 I_{n_i}$ $(i = 1, \ldots, m)$ and D. The complete data ML estimators for D and σ^2 are following

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{m} e_i^T e_i}{N} = \frac{d_1}{N}, \tag{3.3.1}$$

where $N = \sum_{i=1}^{m} n_i$, and

$$\hat{D} = \frac{\sum_{i=1}^{m} b_i b_i^T}{m} = \frac{d_2}{m}. \tag{3.3.2}$$

The E-step gives the updated d_1 and d_2:

$$\hat{d}_1 = E[\sum_{i=1}^{m} e_i^T e_i \mid Y_i, \hat{\alpha}^R(\hat{\theta}_{old}), \hat{\theta}_{old}\,] = \sum_{i=1}^{m} \hat{e}_i^T \hat{e}_i + \hat{\sigma}_{old}^2 tr(\sum_{i=1}^{m} C_i^{-1} Z_i^T Z_i) \tag{3.3.3}$$

and

$$\hat{d}_2 = E[\sum_{i=1}^{m} b_i b_i^T \mid Y_i, \hat{\alpha}^R(\hat{\theta}_{old}), \hat{\theta}_{old}\,] = \sum_{i=1}^{m} \hat{b}_i \hat{b}_i^T + \hat{\sigma}_{old}^2 \sum_{i=1}^{m} C_i^{-1}, \tag{3.3.4}$$

where $\hat{b}_i = C_i Z_i^T (Y_i - X_i \hat{\alpha}^R(\hat{\theta}_{old}))$, $C_i = Z_i^T Z_i + \hat{\sigma}_{old}^2 D_{old}^{-1}$, and $\hat{e}_i = Y_i - X_i \hat{\alpha}^R(\hat{\theta}_{old}) - Z_i \hat{b}_i$. Substituting $\hat{d}_1$ and $\hat{d}_2$ into (3.3.1) and (3.3.2) respectively gives **the M-step** formulas

$$\hat{\sigma}_{new}^2 = \frac{1}{N}\{\sum_{i=1}^{m} \hat{e}_i^T \hat{e}_i + \hat{\sigma}_{old}^2 tr(\sum_{i=1}^{m} C_i^{-1} Z_i^T Z_i)\} \tag{3.3.5}$$

and

$$\hat{D}_{new} = \frac{1}{m}\{\sum_{i=1}^{m} \hat{b}_i\hat{b}_i^T + \hat{\sigma}^2_{old}\sum_{i=1}^{m} C_i^{-1}\}. \tag{3.3.6}$$

To obtain ML estimates, $\hat{\alpha}^R$, $\hat{\sigma}^2$, and $\hat{D}^2$, the EM algorithm begins from obtaining appropriate starting values. We iterate between (3.3.3) and (3.3.4), which is for the E-step, and (3.3.5) and (3.3.6), which is for the M-step until $\hat{\alpha}^R, \hat{\sigma}^2$, and $\hat{D}^2$ converge.

3.4 The PRESS statistic

Ridge regression is usually more effective for predicting future observations than the least squares method when multicollinearity exists (Montgomery *et al.*, 2006). To measure how the model predicts new data well, we use the PRESS statistic of the ith group, which is

$$\begin{aligned} PRESS_i &= \sum_{j=1}^{n_i}[y_{ij} - \hat{y}_{(ij)}]^2 \\ &= \sum_{j=1}^{n_i}\left(\frac{e_{ij}}{1-h_{jj}}\right)^2, \end{aligned} \tag{3.4.1}$$

where y_{ij} is the jth response of the ith group (the ijth response), $\hat{y}_{(ij)}$ is the fitted value of the ijth response based on all observations except the ijth one, and h_{jj} is the jth diagonal element of the matrix $H_i = X_i[\sum_{i=1}^m X_i^T W_i X_i + \frac{\lambda}{\sigma^2}I_p]^{-1}\sum_{i=1}^m X_i^T W_i$, as $\hat{Y}_i = X_i\hat{\alpha}^R = H_iY_i$.

Chapter 4

Application to the Korea PISA 2003 Data

In this section, we apply our method to the Korea PISA 2003 data. The PISA 2003 was organized by the OECD; 41 countries participated in its assessment. In Korea, 149 schools were engaged in the program.

4.1 Data analysis

To identify relationships between self-related cognitions in mathematics and students' math scores, we choose the regressors, that indicate interest in and enjoyment of mathematics ($INTMAT$), instrumental motivation in mathematics ($INSTMOT$), mathematics self-concept ($SCMAT$), mathematics self-efficacy ($MATHEFF$), and artificial regressors—$NEW1$ and $NEW2$. We define $NEW1$ by $x_1 + INSTMOT$, where x_1 is a random sample of $N(0, 0.3^2)$ and $NEW2$ by $x_2 + INSTMOT + SCMAT$, where x_2 is a random sample of $N(0, 0.03^2)$. We generated artificial regressors so that we could obtain large VIFs, which indicate the presence of multicollinearity. Regressors were standardized before analysis because centering removes nonessential ill-conditioning. Missing data for the six regressors and the dependent variable were deleted so that we could use a complete data set. We conducted two level multilevel modeling since students are nested within schools. The level-1 model is as follows:

$$\begin{aligned} PV1MATH_{ij} &= \beta_{0i} + \beta_{1i}INTMAT_{ij} + \beta_{2i}INSTMOT_{ij} + \beta_{3i}SCMAT_{ij} \\ &\quad + \beta_{4i}MATHEFF_{ij} + \beta_{5i}NEW1_{ij} + \beta_{6i}NEW2_{ij} + e_{ij}, \end{aligned}$$

where $PV1MATH_{ij}$ is a plausible value of mathematics score for a student j in school i, and e_{ij} is a residual that follows normal distribution with a mean of 0 and a variance of σ^2. The level-2 model becomes

$$\begin{aligned} \beta_{0i} = \gamma_0 + u_i \ &, \ \beta_{1i} = \gamma_1 \\ \beta_{2i} = \gamma_2 \ &, \ \beta_{3i} = \gamma_3 \\ \beta_{4i} = \gamma_4 \ &, \ \beta_{5i} = \gamma_5 \\ \beta_{6i} = \gamma_6, & \end{aligned}$$

where u_i is a random variable that is normally distributed with a mean of 0 and a variance of τ_{00}. The combined model is specified as

$$\begin{aligned} PV1MATH_{ij} &= \gamma_0 + \gamma_1 INTMAT_{ij} + \gamma_2 INSTMOT_{ij} + \gamma_3 SCMAT_{ij} \\ &\quad + \gamma_4 MATHEFF_{ij} + \gamma_5 NEW1_{ij} + \gamma_6 NEW2_{ij} + u_i \\ &\quad + e_{ij}. \end{aligned} \tag{4.1.1}$$

The model (4.1.1) has seven fixed effects along with two random effects. The two random effects are independent of each other.

4.2 Results

As shown in Table 4.1, some pairwise correlations are very large. This means that near-linear dependencies of the regressors exist. Furthermore, as listed in Table 4.2, VIFs of $INSTMOT$, $SCMAT$, and $NEW2$ are far larger than 10. This implies that severe multicollinearity exists (Montgomery *et al.*, 2006), and therefore, using the generalized least squares method would not provide precise parameter estimates.

Table 4.1: Correlation coefficient for the Korea PISA 2003 data

Variable	INTMAT	INSTMOT	SCMAT	MATHEFF	NEW1	NEW2
INTMAT	1	0.651	0.750	0.512	0.622	0.799
INSTMOT		1	0.524	0.445	0.956	0.883
SCMAT			1	0.606	0.500	0.863
MATHEFF				1	0.422	0.598
NEW1					1	0.843
NEW2						1

Table 4.2: Variance inflation factors of the fixed effects

Variable	INTMAT	INSTMOT	SCMAT	MATHEFF	NEW1	NEW2
VIF	2.890	1029.082	883.707	1.642	11.523	2889.486

We calculated parameter estimates of the model (4.1.1) by using the ordinary fixed effect coefficient estimation for a multilevel model (hereafter, the existing method), which is the same as the proposed method of $\lambda = 0$ (Table 4.3). It shows that $INSTMOT$, $SCMAT$, and $NEW2$'s variances of fixed effect coefficient estimates are very large. This is not surprising because these variables are highly correlated by the definition of the variable $NEW2$. Because of these large variances, fixed effect coefficients could be very sensitive to data in the particular sample collected.

Table 4.3: Parameter estimates of the model (4.1.1) with λ=0

Fixed effect	Coefficient	Standard error	Variance
intercept	539.393	3.664	13.425
INTMAT	0.170	1.472	2.167
INSTMOT	42.691	27.687	766.581
SCMAT	56.165	25.665	658.697
MATHEFF	25.260	1.175	1.381
NEW1	5.278	2.941	8.648
NEW2	-76.109	46.408	2153.658
Random effect		**Variance**	
level-1 variance ($\hat{\sigma}^2$)		3927.044	
level-2 variance ($\hat{\tau}_{00}$)		1860.794	

Before applying the proposed method, it is necessary to choose the ridge parameter. We considered the modified model of (4.1.1), from which u_i was excluded, to make a simple variance structure of the model. The ridge trace is shown in Figure 4.1, where it is apparent that the coefficient estimates of the model are stabilized near $\lambda = 0.4 \sim 0.45$. We choose the ridge parameter $\lambda = 0.406$ to minimizes the GCV of the model.

We then implemented the proposed method to estimate the fixed effect coefficients of the model (4.1.1). We used the starting value from Laird *et al.* (1987). Estimated parameters converged within 30 iterations. As shown in Table 4.4, the parameter estimates, which had large variances in Table 4.3, become smaller. Variances of the corresponding parameter estimates also decreased. To be specific, the variances of coefficient estimates of $INSTMOT$, $SCMAT$, and $NEW2$ were reduced by about 46%, 46.5%, and 46.6%, respectively. This indicates that the proposed method provides more stable fixed effect coefficient estimates in comparison to the existing method. On the other hand, when $\lambda = 0.406$, variance estimates of the random effects are slightly more increased than those of $\lambda = 0$.

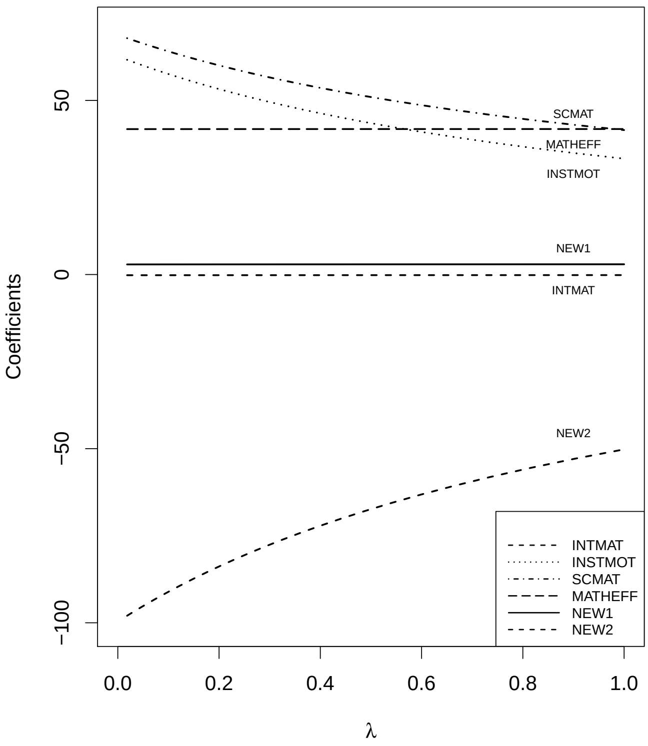

Figure 4.1: Ridge trace for the Korea PISA 2003 data using six regressors

Table 4.4: Parameter estimates of the model (4.1.1) with λ=0.406

Fixed effect	Coefficient	Standard error	Variance
intercept	538.642	3.660	13.395
INTMAT	0.187	1.472	2.166
INSTMOT	29.889	20.290	411.673
SCMAT	44.259	18.768	352.230
MATHEFF	25.274	1.175	1.380
NEW1	5.285	2.936	8.621
NEW2	-54.561	33.907	1149.679
Random effect		Variance	
level-1 variance($\hat{\sigma}^2$)		3927.181	
level-2 variance($\hat{\tau}_{00}$)		1861.784	

To measure how the proposed method predicts new data well, we calculated the PRESS statistic, which is defined in (3.4.1), for each school i ($i = 1, \ldots, 149$) and for each case $\lambda = 0$ and $\lambda = 0.406$, respectively. We also numerated which case had smaller PRESS statistic. Consequently, for all 149 schools, the PRESS statistic of $\lambda = 0.406$ had a smaller value than the PRESS statistic of $\lambda = 0$. This means that when multicollinearity exists, the proposed method has a better capacity for prediction than the existing method.

Chapter 5

Conclusion

The main goal of our study is to investigate the fixed effect coefficient estimation in multilevel model when multicollinearity exists. Our method consists of ridge regression to obtain better fixed effect coefficient estimates, rather than the existing method. We have implemented our method through a simple multilevel model for the Korea PISA 2003 data. As a result, the proposed method has provided coefficient estimates with considerably decreased standard error in comparison to the existing method, even though variances of the random effects increased slightly. Furthermore, we have presented how the proposed method better predicts new data. We have applied PRESS statistic, which we have modified appropriately for our method. We have found that for all of the schools in the Korea PISA 2003 data, the proposed method had smaller PRESS statistic in comparison to the existing method. These findings suggest that our method provides stabilized coefficient estimates and better predicts data when multicollinearity exists.

References

Bickel, R. (2007). *Multilevel analysis for applied research*, Guilford Press, New York.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 78, 45-51.

Golub, G. H., Heath, M., and Wahba, G. (1979). Generalized Cross Validation as a Method for Choosing a Good Ridge Parameter, *Technometrics*, 21, 215-223.

Hoerl, A. E., and Kennard, R. W. (1970a). Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 12, 55-67.

Hoerl, A. E., and Kennard, R. W. (1970b). Ridge regression: Applications to nonorthogonal problems, *Technometrics*, 12, 69-82.

Kreft, I., and de Leeuw, J. (1998). *Introducing multilevel modeling.* Thousand Oaks, CA: Sage.

Laird, N., Lange, N., and Stram, D. (1987). Maximum likelihood computations with repeated measures: Application of the EM algorithm. *Journal of the American Statistical Association*, 82, 97-105.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.

Longford. N. (1993). *Random coefficient models*. Oxford: Clarendon.

McCullagh, P., and Nelder, J. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2006). *Introduction to linear regression analysis* (4th ed.). John Wiley and Sons, Inc.

Organisation for Economic Co-operation and Development (2005). *PISA 2003 Technical Report*. Paris: OECD.

Raudenbush, S., and Bryk, A. (2002). *Hierarchical Linear Models* (2nd ed.). Thousand Oaks, CA: Sage.

Shieh, Y.-Y., and Fouladi, R. T. (2003). The effect of multicollinearity on multilevel modeling parameter estimates and standard errors. *Educational and Psychological Measurement*, 63(6), 951-985.

국문초록

다수준 모델은 데이터의 위계적인 구조를 고려할 수 있기 때문에 사회과학 분야 연구에서 광범위하게 쓰인다. 하지만 다수준 모델에서 고정효과 사이에 다중공선성이 존재하게 되면 계수 추정을 정확하게 하지 못할 수 있다. 이 논문에서는 다수준 모델에서 다중공선성이 존재할 때 고정효과계수를 추정하는 능형회귀에 기반 한 방법에 대해 연구한다. 또한 본 논문에서 제안한 방법을 학생들의 평가 자료를 통해서 기존 방법과 비교하고 제안한 방법이 기존 방법에 비해 얼마나 새로운 데이터를 잘 예측 하는지 PRESS 통계량을 통해 확인한다.

주 요 어 : 다수준 모델, 능형회귀, 다중공선성, EM 알고리즘
학 번 : 2011-20239

감사의 글

졸업을 앞둔 지금, 저는 대학원 오길 참 잘했다는 생각이 듭니다. 지난 2년간의 대학원 생활은 저의 시야를 넓혀 주었고 새로운 도전을 할 수 있는 기회를 주었습니다. 덕분에 저는 지금 새로운 꿈을 꾸고 있습니다. 만약에 제가 대학원을 입학하지 않았다면 저는 아직도 가야할 길을 찾으며 방황하고 있을지도 모릅니다. 하지만 그 무엇보다도 제가 대학원을 와서 얻게 된 것은 사람들이라고 생각 합니다. 저를 믿고 도와준 모든 사람들 덕분에 제가 포기하지 않고 잘 마무리 할 수 있었습니다.

먼저 제가 사랑하는 가족에게 제일 먼저 감사의 마음을 전하고 싶습니다. 아버지, 어머니의 지지와 사랑이 저에게 얼마나 큰 힘이 되었는지 모릅니다. 제가 어떤 일을 할 때마다 항상 저를 믿어주시며 제 편이 되어주시는 부모님 정말 감사하고 사랑합니다. 나의 동생 예원이 에게도 너무 고맙다고 말하고 싶습니다. 예원아, 네가 있어서 언니가 2년간의 대학원 생활을 잘 이겨낸 것 같다. 언니가 늘 고마워하고 있단다.

제 대학원 생활 중 절반을 함께 한 웨이블릿 연구실 식구들에게도 감사의 말씀을 드립니다. 먼저 늘 같이 점심을 먹었던 동익오빠, 민수오빠, 수진이에게 그동안 고마웠다고 말하고 싶습니다. 동익오빠, 오빠의 연구에 대한 열정을 닮고 싶습니다. 우리 연구실의 든든한 방장 민수오빠, 오빠에게 여러가지로 도움 많이 받고 갑니다. 그리고 수진이, 수진이가 있었기에 연구실 생활이 너무 즐거웠어. 사회에 나가서도 잘 할 수 있으리라 믿는다. 또한 제가 헤매고 있을 때 중요한 도움을 줬던 정란언니, 늘 밝은 웃음을 주며 재미있게 이야기 나누었던 예지언니, 공부할 때 많은 도움을 받았던 용민이 그리고 우리 연구실 막내 승민이 까지 모두 고맙습니다. 졸업하면 연구실이 많이 그리울 것 같습니다.

제가 즐겁고 원만한 대학원 생활을 할 수 있게 도와준 소중한 동생들 효영, 선영, 은정, 우석이를 비롯하여 대학원 11학번 동기들 모두에게 감사를 표하고 싶습니다. 훌륭한 동기들과 함께 공부할 수 있어서 지난 2년을 뜻깊게 보낼 수 있었습니다. 신선희 선생님, 박찬호 선생님께도 감사를 드리고 싶습니다. 제가 어떤 선택을 해야 할지 방황하고 있을 때 저에게 용기를 주시고 힘을 주셔서 감사합니다. 선생님들께 받은 도움들 잊지 않고 저도 베푸는 사람이 되도록 하겠습니다. 그리고 제가 대학원 생활중 많이 의지하며 이야기를 나누었던 나의 친구 은희, 서울대에서 만나 너무 반가웠고 많은 고민을 나누었던 기현 언니, 그리고 저를 항상 응원해주고 챙겨주었던 빛나언니 까지 모두 고맙고 성공적으로 대학원 마치길 바랄 께요. 또한 서울에서 열심히 살아가고 있는 재윤, 선화, 영경이 모두 고맙습니다.

광주에서 저에게 조언을 아끼지 않았던 많은 분들에게도 감사의 말씀을 전합니다. 제가 나아가야 할 방향에 대하여 늘 격려를 아끼지 않고 도움을 주시는 강순자 교수님 감사합니다. 앞으로도 열심히 해서 후배들에게 귀감이 되는 선배가 되고 싶습니다. 그리고 저에게 많은 도움을 주고 늘 제 편이 되어 지지해 주었던 요셉, 언제나 그 자리에서 반겨주는 지현언니, 지수, 연순언니, 내 소중한 친구들 소연, 민정, 그리고 이제 새로운 가정을 꾸린 경해까지 모두 고맙습니다.

마지막으로 저의 지도 교수님, 오희석 교수님께 감사하다는 말씀을 드리고 싶습니다. 저의 선택에 대하여 적극적으로 지지해주시고 제가 헤매고 있을 때 아낌없이 도움을 주셨던 것들 모두 제 가슴속에 남아 있습니다. 저는 이제 다시 학생들을 지도하는 교사로 돌아갑니다. 교수님을 생각하며 학생들에게 감화를 줄 수 있는 교사가 될 수 있도록 노력하겠습니다.

2013년 1월

김 서 현