Longitudinal Data Analysis Using Liu Regression

Rahmani M¹, Arashi M*² and Salarzadeh Jenatabadi H²

¹Department of Statistics, Shahrood University of Technology, Shahrood, Iran
²Department of Science and Technology Studies, University of Malaya, Malaysia

Abstract

For understanding and characterizing disease progression over time, Eliot et al. [1] proposed a mixed ridge regression to account for correlated outcomes and potentially high degree of correlated predictors for Biomarker data. However, the ridge estimator is non-linear in nature w.r.t. the ridge parameter and hence it is hard to estimate. In this paper, we propose a linear unified approach to combat this difficulty. Numerical studies illustrate the usefulness of our approach compared to the mixed model.

Keywords: EM-algorithm; Liu regression; Longitudinal; Mixed model; Predictor variables; Design matrix; Regression coefficients; Predictive precision; Correlated predictors; Square root; T-distribution; Non-linear; Mixed model; Log-likelihood; Maximizing; Minimizing; Expectation-maximization; Algorithm; Real data analysis; Longitudinal data; Mean prediction error

Abbreviations: EDF: Effective Degrees of Freedom; EM: Expectation-Maximization; SD: Standard Deviation; MPE: Mean Prediction Error

Introduction

We begin with the simple linear regression model given by

\[ Y = X\beta + \varepsilon \tag{1} \]

Where, \( Y = (y_1, y_2, \ldots, y_n) \) is an \( n \times 1 \) vector of responses, \( X \) is an \( n \times p \) design matrix comprised of \( p < n \) columns representing each of the potential predictor variables, \( n \) is the number of individuals in our sample and \( \varepsilon \sim N(0, \sigma^2I) \) is an \( n \times 1 \) vector of independent errors. The least squares (LS)/maximum likelihood (ML) estimator of the regression coefficients is given by

\[ \hat{\beta} = (X'X)^{-1}X'Y \]

and \( \text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1} \). Notably, in the case that the columns of \( X \) are highly correlated, \( X'X \) will be singular and we replace \((X'X)^{-1}\) with \((X'X)\) where \((X'X)^{-1}\) denotes the generalized inverse, and a unique solution to equation (1) does not exist. Further, in the case of high correlation where \( X'X \) is still invertible, the resulting coefficient estimates will have largely inflated variances, which in turn, results in low predictive precision.

Ridge regression, designed specifically to handle correlated predictors, involves introducing a shrinkage penalty \( \lambda \) to the least squares equation, and subsequently solving for the value such that

\[ \hat{\beta}_s = \arg\min_{\beta} \|Y - X\beta\|^2 + \lambda \|\beta\|^2 \tag{2} \]

The solution to equation (2) is given by

\[ \hat{\beta}_s = (X'X + \lambda \mathbf{I})^{-1}X'Y \]

and we have [2].

\[ \text{Var}(\hat{\beta}_s) = \left[ (X'X + \lambda \mathbf{I})^{-1}X' \right] \left[ (X'X + \lambda \mathbf{I})^{-1}X' \right]' \]

Further, dividing \( \hat{\beta} \) by root \( n \) times the square root of its variance has a Student’s t-distribution with effective degrees of freedom (EDF) given by \( \text{EDF} = n\lambda/(\lambda + \text{tr}(XX')) \) [3-5]. However, the ridge regression estimator \( \hat{\beta}_s \) is non-linear with respect to \( \lambda \) and its estimation is challenging. An alternative approach is proposed by Mayer & Willke [6]. The key idea is that \( \hat{\beta}_s \) is closer to the true \( \beta \) for \( 0 < d < 1 \). In section 2, we will develop their idea for longitudinal mixed model.

Linear mixed effects model

Now, consider the setting in which multiple measurements are observed for each individual over time. The mixed effects model for this setting is given by

\[ y_i = X_\beta + z_i'h + \varepsilon \tag{3} \]

Where, \( i = 1, \ldots, n \) represents individuals, \( X_\beta = (y_{1i}, y_{2i}, \ldots, y_{ni})' \) is a vector of \( n_i \) observations for individual \( i \), and \( X_\beta \) is the corresponding \( n_i \times p \) design matrix of fixed effect covariates. We further assume \( h_i \sim N(0, \Sigma) \) are person-specific random effects, \( z_i \) is the corresponding random effects design matrix, and \( \varepsilon_i \sim N(0, \sigma^2I_{n_{im}}) \) are independent random errors. Finally, we let \( y_i', X_\beta \) and \( z_i \) be appropriately defined matrices representing the concatenation of the corresponding variables over all individuals \( i \).
The log-likelihood function of $Y$ based on this model is given by

$$l(Y) = \frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{N} \log(\hat{Y}_i - Y_i)^2 + \frac{1}{2} (Y_i - X_i \hat{\beta})' Y_i - \frac{1}{2} (Y_i - X_i \hat{\beta})' Y_i - (Y_i - X_i \hat{\beta})$$  (4)

Where $V - \text{Var}(Y) = \sigma^2 I + \sigma^2 I$ and $V_i$ is component corresponding to individual $i$. Maximizing this function with respect to the fixed effects parameter vector, $\beta$ in the non-penalized setting is equivalent to minimizing the least squares objective function that gives the estimate of $\beta$ as

$$\hat{\beta}(X'V_i^{-1}X)^{-1} X'V_i Y$$

Mixed-liu regression

In this section, we introduce a penalized regression approach to estimation for the mixed model given in equation (3). To begin, we assume the variance parameters $\theta = (\sigma, D)$ are known and add a penalization term to objective function of mixed model, which yields

$$\hat{\beta}_{\text{MLiu}} = \arg\min_{\beta \in \mathbb{R}^p} \left\{ (Y_i - X_i \beta)' (Y_i - X_i \beta) + d(\hat{\beta} - \beta)' (\hat{\beta} - \beta) \right\}$$  (6)

Differentiating the objective function in equation (6), setting the resultant equal to 0 and solving, we have:

$$\theta = -\frac{1}{\sigma^2} \frac{\partial}{\partial \theta} \left[ (Y_i' Y_i - Y_i' X_i \beta + \beta' X_i' Y_i + d \hat{\beta}' - \hat{\beta} + \beta)' \beta \right]
= -2X_i'Y_i + 2X_i'X_i \beta - 2d \hat{\beta} + 2\beta
$$

Hence,

$$\hat{\beta}_{\text{MLiu}} = \left( X_i'V_i^{-1}X_i + 1 \right)^{-1} \left( X_i'V_i Y_i + d \hat{\beta} \right)
= \left( X_i'V_i^{-1}X_i + 1 \right)^{-1} \left( X_i'V_i Y_i + d \hat{\beta} \right)
= \left( X_i'V_i^{-1}X_i + 1 \right)^{-1} \left( X_i'V_i Y_i + d \hat{\beta} \right)
$$

Additionally, it can be shown that

$$\text{Var}(\hat{\beta}_{\text{MLiu}}) = \left( X_i'V_i^{-1}X_i + 1 \right)^{-1} \left( X_i'V_i X_i + d \hat{\beta} \right) \text{Var}(Y_i' X_i) \left( X_i'V_i X_i + d \hat{\beta} \right) \left( X_i'V_i^{-1}X_i + 1 \right)^{-1}$$

We suggest to estimate $d$ in equation (7) by

$$\hat{d} = \arg\min_{d \in \mathbb{R}} \left\{ \text{Var}(\hat{\beta}_{\text{MLiu}}) \right\}$$

Where

$$\text{Var}(\hat{\beta}_{\text{MLiu}}) = \left( X_i'V_i^{-1}X_i + 1 \right)^{-1} \left( X_i'V_i X_i + d \hat{\beta} \right) \text{Var}(Y_i' X_i) \left( X_i'V_i X_i + d \hat{\beta} \right) \left( X_i'V_i^{-1}X_i + 1 \right)^{-1}$$

And

$$Y = X \hat{\beta}_{\text{MLiu}}$$

More generally, consider the setting in which the variance parameters $\theta = (\sigma, D)$ are unknown. Elliot et al. [1] proposed an extension of the expectation-maximization (EM) algorithm described by Laird & Ware [4], that includes an additional step for estimation of the ridge component. Here, we exhibit an EM algorithm to solve $\hat{\beta}_{\text{MLiu}}$ for unknown $\theta$. This approach is summarized by the following step-by-step procedure.

I. (E-Step) Initialize $\theta^{(0)} = \theta$ and $d^{(0)} = d$. Solve for $\hat{\beta}_{\text{MLiu}}^{(0)}$ and the sufficient statistics $\hat{t}_1^{(0)}$ and $\hat{t}_2^{(0)}$ given by:

$$\left( \frac{\partial}{\partial \theta} \text{Var}(\hat{\beta}_{\text{MLiu}}) \right) = \text{Var}(\hat{\beta}_{\text{MLiu}}^{(0)})$$

II. (M-Step) Solve for $\theta^{(t+1)}$ where $N = \sum_i n_i$, and $n_i$ is the number of individuals in our sample, and let

$$V^{(t+1)} = Z \hat{D}^{(t+1)} Z' + \hat{\sigma}^{(t+1)} I$$

III. Update $\hat{d}^{(t+1)}$ and let

$$\hat{\beta}_{\text{MLiu}}^{(t+1)} = \left( X_i'V_i^{-1}X_i + 1 \right)^{-1} \left( X_i'V_i X_i + d^{(t+1)}I \right) \hat{\beta}_{\text{MLiu}}^{(t)}$$

IV. Repeat Steps (1)-(3) a large number of times and until a convergence criterion is met.

In the forthcoming section we evaluate the performance of the mixed Liu estimator by a Monte Carlo simulation study.

**Simulation study**

A simple simulation study is conducted to characterize the relative performances of mixed Liu regression and the usual mixed effects modeling model in the context of multiple, correlated predictors. For simplicity of presentation, the simulation study assumes repeatedly measured outcomes, while the predictor variables are measured at a single, baseline time point, as in Elliot et al. [1]. We further assume

**Table 1: Results of the Monte Carlo simulation study.**

| $\rho$ | $\beta$ | sd | Estimate | sd |
|-------|---------|----|----------|----|
| 0     | -0.0036 | 0.107 | -0.0032 | 0.107 |
| 0.4   | 0.3982  | 0.12 | 0.3984  | 0.1199 |
| 1     | 1.01052 | 0.1239 | 1.0104 | 0.1238 |
| 0.6   | 1.5992  | 0.1364 | 1.5986 | 0.1364 |
| 0.3   | 1.6002  | 0.1054 | 1.6     | 0.1054 |
| 0.4   | 1.9816  | 0.172 | 1.9808  | 0.1719 |
0.9

| 0   | 0.0344 | 0.3271 | 0.0373 | 0.3264 |
|-----|--------|--------|--------|--------|
| 0.4 | 0.395  | 0.3755 | 0.3968 | 0.3743 |
| 1   | 1.0264 | 0.3833 | 1.0261 | 0.3821 |
| 1.6 | 1.5959 | 0.2588 | 1.5942 | 0.2581 |
| 2   | 1.9474 | 0.3645 | 1.9445 | 0.3638 |

Each predictor variable is assumed to arise from a normal distribution with mean equal to 5 and variance equal to 1. The correlation between predictor variables, given by \( \rho \) in Table 1, is assumed to take on values between 0 and 0.9. Starting values for the variance components are derived based on fitting a mixed model with no Liu component. In total, \( M = 100 \) simulations are conducted for each condition based on sample sizes of \( n = 40 \) individuals. According to the results of Table 1, the mixed Liu estimates often have lesser bias than the mixed ones. Also the mixed Liu is superior, in standard deviation (sd) sense.

### Table 2: Introduction to data and variables format.

| Variables | Description | Effect |
|-----------|-------------|--------|
| id        | patients identifier; in total there are 312 patients. | F |
| Years     | number of years between registration and the earlier of death, transplantation, or study analysis time. | F |
| Status    | a factor with levels alive, transplanted and dead. | R |
| Drug      | a factor with levels placebo and D-penicill. | R |
| Age       | at registration in years. | F |
| Sex       | a factor with levels male and female. | R |
| Years     | number of years between enrollment and this visit date, remaining values on the line of data refer to this visit. | F |
| Ascites   | a factor with levels No and Yes. | R |
| Hepatomegaly | a factor with levels No and Yes. | R |
| Spiders   | a factor with levels No and Yes. | R |
| edema     | a factor with levels No edema, edema no diuretics and edema despite diuretics. | R |
| SerBilir  | serum bilirubin in mg/dl. | F |
| SerChol   | serum cholesterol in mg/dl. | F |
| Albumin   | albumin in gm/dl. | F |
| Alkaline  | alkaline phosphatase in U/liter. | F |
| SGOT      | SGOT in U/ml. | F |
| Platelets | platelets per cubic ml / 1000. | F |
| Prothrombin | prothrombin time in seconds. | F |
| Histologic| histologic stage of disease. | F |
| Status2   | a numeric vector with the value 1 denoting if the patient was dead, and 0 if the patient was alive or transplanted. | R |

For our purpose, a \( K \)-fold cross validation is used to obtain an estimate of the prediction errors of the model. In a \( K \)-fold cross validation, the dataset is randomly divided into \( K \) subsets of roughly equal size. One subset is left aside, \( \{ X^{test}, Y^{test}\} \), termed as test set, while the remaining \( K - 1 \) subsets, called the training set, are used to fit model. The resultant estimator is called \( \hat{\beta}_{train} \).

The fitted model is then used to predict the responses of test data set. Finally, prediction errors are obtained by taking the squared deviation of the observed and predicted values in the test set, i.e.

\[
PE^k = \left\| X^{test} \hat{\beta} - Y^{test} \right\|^2
\]
Where \( y_{i}^{test} = X_{i}^{test} \hat{\beta}_{mix}^{est} \). The process is repeated for all \( K \) subsets and the prediction errors are combined. To account for the random variation of the cross validation, the process is reiterated \( N \) times and is estimated the average prediction error is given by

\[
MPE = \text{median} \left\{ \frac{1}{k} \sum_{i=1}^{k} PE_{i}^{k}, \ldots, \frac{1}{k} \sum_{i=1}^{k} PE_{N}^{k} \right\}
\]

Where \( PE_{i}^{k} \) is the prediction error of considering \( k \)th test set in \( i \)th iteration. Our result are based on \( N = 200 \) case resampled bootstrap sample. In Table 3, we report the estimates and MPE values. Based on the results, the proposed mixed-Liu estimator performs better than the mixed one, in MPE sense. Further, the absolute value of estimates in the mixed Liu estimates are lesser than the mixed.

Table 3: Estimates of read data.

| Covariate | Mixed | Mixed-Liu |
|-----------|-------|-----------|
| Age       | -0.0056 | -0.0053 |
| Year      | 0.7095  | 0.7085 |
| Serbilir  | -0.1711 | -0.1724 |
| Serchol   | -0.0004 | -0.0004 |
| Albumin   | 2.3504  | 2.3236 |
| Alkaline  | 0.0001  | 0.0001 |
| SGOT      | 0.0023  | 0.0023 |
| Platelets | 0.0009  | 0.001 |
| Prothrombin | -0.008  | -0.0037 |
| Histologic| -0.5086 | -0.5046 |
| MPE       | 0.01559 | 0.01535 |

Conclusion

In this paper, we developed a linear unified procedure called Liu in the linear mixed model for longitudinal data analysis. Hence, we considered a penalized likelihood approach and propose the Liu-mixed regression estimator for the vector of regression coefficients. An EM algorithm also exhibited to solve the penalized likelihood for the unknown parameters. Numerical studies demonstrated the good performance of the proposed mixed Liu estimator for the multicollinear situation.

References

1. Elliot M, Ferguson J, Reilly MP, Foulkes AS (2011) Ridge Regression for Longitudinal Biomarker Data. Int J Biostat 7(1): 37.
2. Hoerl A, Kennard R (1970) Ridge regression: Applications to nonorthogonal problems. Technometrics 12: 69-82.
3. Hoerl A, Kennard R (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12: 55-67.
4. Laird N, Ware J (1982) Random-effects models for longitudinal data. Biometrics 38(4): 963-974.
5. Malo N, Libiger O, Schork N (1982) Random-effects models for longitudinal data. Biometrics 82: 375-385.
6. Mayer LS, Willke IA (1973) On biased estimation in linear model. Technometrics 15: 497-508.