



**Mini Review**

Volume 7 Issue 5 - July 2018  
 DOI: 10.19080/BBOAJ.2018.07.555725

Biostat Biometrics Open Acc J  
 Copyright © All rights are by Arashi M

# Longitudinal Data Analysis Using Liu Regression



**Rahmani M<sup>1</sup>, Arashi M<sup>1\*</sup> and Salarzadeh Jenatabadi H<sup>2</sup>**

<sup>1</sup>Department of Statistics, Shahrood University of Technology, Shahrood, Iran

<sup>2</sup>Department of Science and Technology Studies, University of Malaya, Malaysia

**Submission:** April 24, 2018; **Published:** July 18, 2018

**\*Corresponding author:** Arashi M, Department of Statistics, Faculty of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran; Email: m\_arashi\_stat@yahoo.com

**Abstract**

For understanding and characterizing disease progression over time, Eliot et al. [1] proposed a mixed ridge regression to account 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,  $Y = (y_1, y_2, \dots, y_n)$  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  $\varepsilon \sim N_n(0, \sigma^2 I_n)$  is an  $n \times 1$  vector of independent errors. The least squares (LS)/maximum likelihood (ML) estimator of the regression coefficients is given by

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

and  $Var(\hat{\beta}) = (X^T X)^{-1} \sigma^2$ . Notably, in the case that the columns of  $X$  are highly correlated,  $X^T X$  will be singular and we replace  $(X^T X)^{-1}$  with  $(X^T X)^{-}$  where ‘-’ denotes the generalized inverse, and a unique solution to equation (1) does not exist. Further, in the case of high correlation where  $X^T X$  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

$$\hat{\beta}_R = \arg \min_{\beta} \{ (Y - X\beta)^T (Y - X\beta) + \lambda^2 \beta^T \beta \} \tag{2}$$

The solution to equation (2) is given by

$$\hat{\beta}_R = (X^T X + \lambda I)^{-1} X^T Y$$

and we have [2],

$$Var(\hat{\beta}_R) = \left[ (X^T X + \lambda I)^{-1} X^T \right] \left[ (X^T X + \lambda I)^{-1} X^T \right]^T$$

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  $EDF = tr \left[ (X^T X + \lambda I)^{-1} X^T X \right]$  [3-5]. However, the ridge regression estimator  $\hat{\beta}_R$  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  $d\hat{\beta}$  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

$$y_i = X_i \beta + z_i^T b_i + \varepsilon_i \tag{3}$$

Where,  $i = 1, \dots, n$  represents individuals,  $y_i = (y_{i1}, y_{i2}, \dots, y_{in})^T$  is a vector of  $n_i$  observations for individual  $i$ , and  $X_i$  is the corresponding  $n_i \times p$  design matrix of fixed effect covariates. We further assume  $b_i \sim N_q(0, D)$  are person-specific random effects,  $z_i$  is the corresponding random effects design matrix, and  $\varepsilon_i \sim N_{n_i}(0, \sigma^2 I_{n_i})$  are independent random errors. Finally, we let  $Y, X$  and  $Z$  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

$$l(Y) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log(V_i) - \frac{1}{2} (Y - X\beta)^T V^{-1} (Y - X\beta) \quad (4)$$

Where  $V = \text{Var}(Y) = ZDZ^T + \sigma^2 I$  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

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \quad (5)$$

**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

$$\hat{\beta}_{MLiu} = \arg \min_{\beta \in \mathbb{R}^p} \left\{ (Y - X\beta)^T V^{-1} (Y - X\beta) + (d\hat{\beta} - \beta)^T (d\hat{\beta} - \beta) \right\} \quad (6)$$

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

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta} \left[ (Y - X\beta)^T V^{-1} (Y - X\beta) + (d\hat{\beta} - \beta)^T (d\hat{\beta} - \beta) \right] \\ &= \frac{\partial}{\partial \beta} \left[ Y^T V^{-1} Y - Y^T V^{-1} X\beta - \beta^T X^T V^{-1} Y + \beta^T X^T V^{-1} X\beta + d^2 \hat{\beta}^T \hat{\beta} - d\hat{\beta}^T \beta + \beta^T \beta \right] \\ &= -2X^T V^{-1} Y + 2X^T V^{-1} X\beta - 2d\hat{\beta}^T + 2\beta \end{aligned}$$

Hence,

$$\begin{aligned} \hat{\beta}_{MLiu} &= (X^T V^{-1} X + 1)^{-1} (X^T V^{-1} Y + d\hat{\beta}) \\ &= (X^T V^{-1} X + 1)^{-1} \left( (X^T V^{-1} X) (X^T V^{-1} X)^{-1} X^T V^{-1} Y + d\hat{\beta} \right) \\ &= (X^T V^{-1} X + 1)^{-1} (X^T V^{-1} X + dI) \hat{\beta} \end{aligned} \quad (7)$$

Additionally, it can be shown that

$$\text{Var}(\hat{\beta}_{MLiu}) = (X^T V^{-1} X + 1)^{-1} (X^T V^{-1} X + dI) (X^T V^{-1} X)^{-1} (X^T V^{-1} X + dI) (X^T V^{-1} X + 1)^{-1}.$$

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

$$\hat{d} = \arg \min_d \left\{ n^{-1} (Y - Y)^T (Y - Y) + 2n^{-2} \text{tr}(S)(Y - Y)^T (Y - Y) \right\} \quad (8)$$

Where

$$S = X(X^T V^{-1} X + I)^{-1} (X^T V^{-1} X + dI) (X^T V^{-1} X)^{-1} X^T V^{-1} + ZDZ^T \\ \left[ I - X(X^T V^{-1} X + I)^{-1} (X^T V^{-1} X + dI) (X^T V^{-1} X)^{-1} X^T V^{-1} \right]$$

And

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

More generally, consider the setting in which the variance parameters  $\theta = (\sigma, D)$  are unknown. Eliot 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  $\hat{\beta}_{MLiu}$  for unknown  $\theta$ . This approach is summarized by the following step-by-step procedure.

I. (E-Step) Initialize  $\hat{\theta}^{(i)} = \theta_0$  and  $\hat{d}^{(i)} = d_0$ . Solve for  $\hat{\beta}_{MLiu}^{(i)}$  and the sufficient statistics  $\hat{t}_1^{(i)}$  and  $\hat{t}_2^{(i)}$  given by:

$$\left( {}_{(i)}\theta^{\epsilon, n_i} \mid {}_{(i)}\mathcal{G}^{\epsilon, 1} \mid X \mid {}_{(i)}\mathcal{Z}, {}_{(i)}\mathcal{Z}^{1=i} \mid \Sigma \right) \mathcal{I} = {}_{(i)}\mathcal{I}$$

$$\hat{t}_2^{(i)} = E \left( \sum_{i=1}^n b_i^T b_i \mid Y_i, \hat{\beta}_{MLiu}^{(i)}, \hat{\theta}^{(i)} \right)$$

II. (M-Step) Solve for  $\hat{\theta}^{(t+1)}$  where  $\hat{\sigma}^{2(t+1)} = \frac{\hat{t}_1^{(i)}}{N}$ ,  $D^{(t+1)} = \frac{\hat{t}_2^{(i)}}{N}$ ,  $N = \sum_{i=1}^n n_i$  and  $n$  is the number of individuals in our sample, and let

$$V^{(t+1)} = ZD^{(t+1)}Z^T + \hat{\sigma}^{2(t+1)}I$$

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

$$\hat{\beta}_{MLiu}^{(t+1)} = (X^T V^{-(t+1)} X + I)^{-1} (X^T V^{-(t+1)} X + d^{(t+1)} I) \hat{\beta}^{(t+1)}$$

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 approach 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 Eliot et al. [1]. We further assume

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

$\rho$	$\beta$	Mixed		Mixed-Liu	
		Estimate	sd	Estimate	sd
0	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
	1.6	1.6031	0.0901	1.6029	0.0901
	2	1.9917	0.1061	1.9914	0.1061
0.3	0	0.001	0.1302	0.0015	0.1301
	0.4	0.3978	0.1439	0.3981	0.1438
	1	1.0113	0.1437	1.01125	0.1436
	1.6	1.6002	0.1054	1.6	0.1054
	2	1.9893	0.128	1.9888	0.128
0.6	0	0.0097	0.1746	0.0105	0.1745
	0.4	0.3958	0.1859	0.3963	0.1857
	1	1.0133	0.1912	1.0132	0.191
	1.6	1.599	0.1365	1.5986	0.1364
	2	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

$n_i = 4$  measurements for each subject  $i$  and generate data according to the model of equation(3) where  $\beta = (0, 0.4, 1, 1.6, 2)$ ,  $b_{ijk} \sim N(0, 0.6)$  and  $\varepsilon_{ijk} \sim N(0, 1)$ . 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-penicil.	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  $\hat{\beta}^{train}$ .

**Real data analysis**

The data set we are analyzing here is the Mayo Clinic Primary Biliary Cirrhosis data, from the package “JMbayes” in R software. It consists of 312 randomized patients with primary biliary cirrhosis, a rare autoimmune liver disease, at Mayo Clinic. In this study we have 1945 observations on the 20 variables, listed in Table 2. The response variable is the number of years, indicated as “years” in Table 2 and the variables considered as fixed and random effects are marked as “F” and “R”, respectively in this table. Since the variables have been measured number of times for each individuals, so we have a longitudinal data set. On the other hand, some of the variables like sex in this data set will put the subjects in special groups, so we can consider these variables as random effects, as marked as “R” in Table 2, so we should use mixed model for analyzing this data set. The estimate of coefficients are obtained using the EM algorithm as outlined in Section 2. To compare the performance of the mixed Liu estimator, we evaluate the mean prediction error (MPE); the lesser, the better. In what follows, we describe the scheme we used to derive the MPE.

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.

$$PE^k = \left\| X_k^{test} \beta - y_k^{test} \right\|^2$$

Where  $y_k^{test} = X_k^{test} \hat{\beta}_k^{train}$ . 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

$$MPE = median \left\{ \frac{1}{k} \sum_{k=1}^k PE_1^k, \dots, \frac{1}{k} \sum_{k=1}^k PE_N^k \right\}$$

Where  $PE_N^k$  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**

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



This work is licensed under Creative Commons Attribution 4.0 License  
 DOI: [10.19080/BBOAJ.2018.07.555725](https://doi.org/10.19080/BBOAJ.2018.07.555725)

**Your next submission with Juniper Publishers will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
- ( Pdf, E-pub, Full Text, Audio )**
- Unceasing customer service

**Track the below URL for one-step submission**  
<https://juniperpublishers.com/online-submission.php>