



# Improved mixed model for longitudinal data analysis using shrinkage method

M. Rahmani<sup>1</sup> · M. Arashi<sup>1</sup> · N. Mamode Khan<sup>2</sup> · Y. Sunecher<sup>3</sup>

Received: 20 August 2018 / Accepted: 17 October 2018 / Published online: 1 November 2018  
© The Author(s) 2018

## Abstract

The problem of multicollinearity among predictor variables is a frequent issue in longitudinal data analysis. In this context, this paper proposes a mixed ridge regression model via shrinkage methods to analyze such data. Furthermore, in view of obtaining more efficient estimators, we propose preliminary and Stein-type estimators using prior information for fixed-effects parameters. The model parameters are estimated via the EM algorithm. A simulation study is also presented to assess the performance of the estimators under different estimation methods. An application to the HIV data is also illustrated.

**Keywords** EM algorithm · Longitudinal data · Mixed model · Preliminary test · Stein estimation · Ridge regression

**Mathematics Subject Classification** 65C60 · 62J12 · 62H12 · 62J20 · 62J10

## Introduction

In longitudinal data setup, repeated measures of some variables of interest are collected over a specified time period for different independent subjects or individuals. Such types of data are commonly encountered in medical research where the responses are subject to various time-dependent and time-constant effects such as pre- and post-treatment types, gender effect, baseline measures and among others (see Mamode Khan et al. [1], Yuan et al. [2], Verbeke et al. [3], Temesgen and Kebede [4], Seyoum et al. [5] and the references therein). It is quite natural, in the above examples, the repeated measures shall exhibit some forms of dependence

that may be resulted from some serial or random effects as outlined by Zeger and Liang [6], Thall and Vail [7], Laird and Ware [8], Sutradhar [9] and Sutradhar and Jowaher [10]. Thus, the main purpose of the longitudinal studies is to estimate the effects of the various parameters and determine their significance while the dependence estimate is treated as secondary. In this context, FitzMaurice and Laird [11] and Sutradhar et al. [12] have proposed various likelihood-based and pseudo-likelihood-based estimation procedures to estimate the regression effects but the efficiency of the estimators in these approaches may be questionable, in particular, under multi-collinearity among the predictor variables as considered by Eliot et al. [13], Hossain et al. [14] and Saleh et al. [15].

Since longitudinal data mostly arise from clinical studies, the expert knowledge about the parameters has vital impact on the output and is thus an important component in the estimation of model parameters. The preliminary test and shrinkage techniques are mostly used mechanisms in which a prior knowledge can be included in the estimation stage (see papers by Ali and Saleh [16], Ahmed and Fallahpour [17], Roozbeh and Arashi [18] and Yuzbasi and Ahmed [19], Yuzbasi et al. [20] and Asar [21] and the references therein). In this paper, we develop the preliminary test and shrinkage estimation methods for the analysis of longitudinal data in ridge regression context, where some parameters are subject to certain/uncertain restrictions. By this, we improve the

---

✉ M. Arashi  
m\_arashi\_stat@yahoo.com

M. Rahmani  
mrahmani.stat@yahoo.com

N. Mamode Khan  
n.mamodekhan@uom.ac.mu

Y. Sunecher  
ysunecher@umail.utm.ac.mu

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

<sup>2</sup> University of Mauritius, Reduit, Mauritius

<sup>3</sup> University of Technology Mauritius, Pointe-Aux-Sables, Mauritius

estimation technique, in both the mean squared error (MSE) and mean prediction error (MPE) senses.

We begin with the linear mixed effects (LM) model given by

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{b}_i + \varepsilon_i, \tag{1}$$

where  $i = 1, \dots, n$  represents individual,  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$  is a vector of  $n_i$  observations for individual  $i$ , and  $\mathbf{x}_i$  is the corresponding  $n_i \times p$  design matrix of fixed-effect covariates. We further assume  $\mathbf{b}_i \sim N_q(0, \mathbf{D})$  is person-specific random effects,  $\mathbf{z}_i$  the corresponding random effects design matrix and  $\varepsilon_i \sim N_n(0, \sigma^2\mathbf{I}_{n_i \times n_i})$ . Let  $\mathbf{Y}, \mathbf{X}$  and  $\mathbf{Z}$  be appropriately defined matrices representing the concatenation of the corresponding variables over all individuals  $i$ . Then, the LM model in matrix notation has form

$$\begin{aligned} \boldsymbol{\mu} &= E(\mathbf{Y}|\mathbf{b}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} \\ \begin{pmatrix} E(\mathbf{y}_1|\mathbf{b}) \\ E(\mathbf{y}_2|\mathbf{b}) \\ \vdots \\ E(\mathbf{y}_n|\mathbf{b}) \end{pmatrix} &= \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \mathbf{z}_1^T & 0 & \dots & 0 \\ 0 & \mathbf{z}_2^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{z}_n^T \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{pmatrix}, \end{aligned} \tag{2}$$

where  $\boldsymbol{\mu}: N \times 1, \mathbf{X}: N \times p, \boldsymbol{\beta}: p \times 1, \mathbf{Z}: N \times nq$  and  $\mathbf{b}: nq \times 1$ . Here,  $N = \sum_{i=1}^n n_i$ .

The log-likelihood function of  $\mathbf{Y}$  based on this model is given by

$$l(\mathbf{Y}) \propto -\frac{1}{2} \sum_{i=1}^n \log|\mathbf{V}_i| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{3}$$

where  $\mathbf{V} = \text{Var}(\mathbf{Y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \sigma^2\mathbf{I}$  and  $\mathbf{V}_i$  is the variance component corresponding to individual  $i$ . Maximizing (3) with respect to the fixed-effects parameter vector,  $\boldsymbol{\beta}$ , in the non-penalized setting is equivalent to minimizing the least squares objective function that gives the estimate of  $\boldsymbol{\beta}$  as

$$\hat{\boldsymbol{\beta}}_M = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

Consider a situation in which the multi-collinearity problem is present. The ridge regression approach designed specifically to handle correlated predictors involves introducing a shrinkage penalty  $k$  to the least squares equation, and subsequently solving for the value of  $\boldsymbol{\beta}$  such that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MR} &= \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\text{argmin}} \{ (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^T\boldsymbol{\beta} \} \\ &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \end{aligned}$$

The structure of the paper is therefore organized as follows: In “[Shrinkage approach](#)” section, the novel estimation approach based on shrinkage method is proposed which also includes some computational aspects and the EM algorithm. “[Simulation study](#)” section provides the Monte Carlo simulation experiments followed by the section on the application

of the new techniques to the HIV data, and a comparative study is also presented. The conclusion is presented in final section.

## Shrinkage approach

In this section, we consider the mixed ridge (MR) model (2) and develop preliminary test and Stein-type estimation of the fixed-effects parameter  $\boldsymbol{\beta}$  when it is a priori suspected that  $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ . Usually,  $\boldsymbol{\beta}$  comes from different sources: (1) a fact known from theoretical or experimental considerations, (2) hypothesis that may have to be tested or (3) an artificially imposed condition to reduce redundancy in the description of the model. Here, we interpret as expert knowledge.

## Incorporating expert knowledge

In the statistical literature, preliminary test estimation of parameters was introduced by Bancroft [22] to estimate the parameters of a model when it is suspected that some “uncertain prior information” (UPI) on the parameter of interest is available. The method involves a statistical test of the UPI based on an appropriate statistic and a decision on whether the model-based sample estimate or the prior information-based estimate of the model parameters should be taken.

In our case, if we suspect  $\boldsymbol{\beta}$  to be  $\boldsymbol{\beta}_0$ , then  $\boldsymbol{\beta}_0$  is termed as the restricted estimator (RE) that must be incorporated in the analysis. Otherwise,  $\hat{\boldsymbol{\beta}}_{MR}$  is used. In practice, the prior information that  $\boldsymbol{\beta} = \boldsymbol{\beta}_0$  is *uncertain*. The doubt on this prior information can be removed by testing the following hypothesis:

$$\mathbf{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{versus} \quad \mathbf{H}_A : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$$

As a result of this test, we choose  $\boldsymbol{\beta}_0$  or not based on the rejection or acceptance of  $\mathbf{H}_0$ . Accordingly, we develop the estimator

$$\hat{\boldsymbol{\beta}}_{MR}^{PT} = \begin{cases} \hat{\boldsymbol{\beta}}_{MR} & \text{if } R\mathbf{H}_0 \\ \boldsymbol{\beta}_0 & \text{if } A\mathbf{H}_0 \end{cases}$$

called the preliminary test MR (PTMR) estimator. Using indicator function, this estimator can be rewritten as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MR}^{PT} &= \hat{\boldsymbol{\beta}}_{MR}I_{\{RH_0\}} + \boldsymbol{\beta}_0I_{\{AH_0\}} \\ &= \hat{\boldsymbol{\beta}}_{MR} - \hat{\boldsymbol{\beta}}_{MR}I_{\{AH_0\}} + \boldsymbol{\beta}_0I_{\{AH_0\}} \\ &= \hat{\boldsymbol{\beta}}_{MR} - (\hat{\boldsymbol{\beta}}_{MR} - \boldsymbol{\beta}_0)I_{\{AH_0\}} \\ &= \hat{\boldsymbol{\beta}}_{MR} - (\hat{\boldsymbol{\beta}}_{MR} - \boldsymbol{\beta}_0)I(L_n < L_n(\alpha)) \end{aligned} \tag{4}$$

where  $I_A$  is the indicator function of the set  $A$ ,  $L$  is the Wald statistic given by  $L = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$  and  $L_n(\alpha)$

is the upper  $\alpha$  level critical value of the  $\chi$ -distribution with  $p$  d.f.

The PTMR estimator is highly dependent on the level of significance  $\alpha$  and has discrete nature which is simplified to one of the extremes  $\hat{\beta}_{MR}$  and  $\beta_0$ , according to the output of test. In this respect, making use of a continuous and  $\alpha$ -free estimator may make more sense.

Stein-type estimation was introduced by Stein [23] and James and Stein [24] in the statistical literature. It combines UPI on the parameters of interest and the sample observation from the statistical model. In the context of LM model, using the same approach as in Saleh [15], the Stein-type estimator of  $\beta$  has form

$$\hat{\beta}_{MR}^S = \hat{\beta}_{MR} - c(\hat{\beta}_{MR} - \beta_0)L_n^{-1}, \quad c = \frac{(p - 2)(N - p)}{p(N - p + 2)} \quad (5)$$

Unlike the PTMR estimator, the Stein-type estimator is a smooth function and independent of the level of significance  $\alpha$ .

Notice that the forms of  $\hat{\beta}_{MR}^S$  and  $\hat{\beta}_{MR}^{PT}$  are similar where Eq. (5) obtained from Eq. (4) by replacing  $I(L_n < L_n(\alpha))$  by  $cL_n^{-1}$  to make it independent of  $\alpha$ .

### Computational aspects

Consider the setting in which the variance parameters  $\theta = (\mathbf{D}, \sigma^2)$  are unknown. Eliot et al. [13] proposed an extension of the expectation–maximization (EM) algorithm described by Laird and Ware [8] that includes an additional step for estimation of the ridge component. We refine their procedure to evaluate shrinkage estimators as well.

Further, for the ease of computations, we use the estimate of Hoerl and Kenard [25] for the ridge parameter as the initial value. It is given by

$$\hat{k}_0 = \frac{\hat{\sigma}^2}{\hat{\alpha}_{max}^2},$$

where  $\hat{\alpha}_{max} = \max(\hat{\alpha}_1, \dots, \hat{\alpha}_p)$  in which  $\hat{\alpha} = \mathbf{\Gamma}^T \hat{\beta}$  and  $\mathbf{\Gamma}$  is the orthogonal eigenvector is spectral decomposition of  $\mathbf{X}^T \mathbf{X}$ , i.e.,  $\mathbf{X}^T \mathbf{X} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T$ ,  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ , where  $\lambda_i$  is the  $i$ th eigenvalue of  $\mathbf{X}^T \mathbf{X}$ . In what follows, we propose the refined EM algorithm.

---

#### Algorithm 1 EM Algorithm

---

i) Initialize  $\hat{\theta}^{(t)} = \theta_0$  and  $\hat{k}^{(t)} = k_0$ . Solve for  $\hat{\beta}_{MR}^{(t)}$  and the sufficient statistics  $\hat{t}_1^{(t)}$  and  $\hat{t}_2^{(t)}$  given by

$$\begin{aligned} \hat{t}_1^{(t)} &= E \left( \sum_{i=1}^n \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i \mid \mathbf{Y}_i, \hat{\beta}_{MR}^{(t)}, \hat{\theta}^{(t)} \right) \\ &= \sum_{i=1}^n \left[ \boldsymbol{\varepsilon}_i^T(\hat{\theta}^{(t)}) \boldsymbol{\varepsilon}_i(\hat{\theta}^{(t)}) + \text{trVar} \left( \boldsymbol{\varepsilon}_i \mid \mathbf{Y}_i, \hat{\beta}_{MR}^{(t)}, \hat{\theta}^{(t)} \right) \right] \end{aligned}$$

$$\begin{aligned} \hat{t}_2^{(t)} &= E \left( \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^T \mid \mathbf{Y}_i, \hat{\beta}_{MR}^{(t)}, \hat{\theta}^{(t)} \right) \\ &= \sum_{i=1}^n \left[ \mathbf{b}_i(\hat{\theta}^{(t)}) \mathbf{b}_i^T(\hat{\theta}^{(t)}) + \text{Var} \left( \mathbf{b}_i \mid \mathbf{Y}_i, \hat{\beta}_{MR}^{(t)}, \hat{\theta}^{(t)} \right) \right] \end{aligned}$$

ii) Solve for  $\hat{\theta}^{(t+1)}$  where  $\hat{\sigma}^{2(t+1)} = \hat{t}_1^{(t)}/N$ ,  $\hat{D}^{(t+1)} = \hat{t}_2^{(t)}/n$ ,  $N = \sum_{i=1}^n n_i$  and  $n$  is the number of individuals in our sample, and let  $\hat{\mathbf{V}}^{(t+1)} = \mathbf{Z} \hat{D}^{(t+1)} \mathbf{Z}^T + \hat{\sigma}^{2(t+1)} \mathbf{I}$

iii) Update  $\hat{k}^{(t+1)}$  and let

$$\hat{\beta}_{MR}^{(t+1)} = \left( \mathbf{X}^T \hat{\mathbf{V}}^{-1(t+1)} \mathbf{X} + \hat{k}^{(t+1)} \mathbf{I} \right)^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1(t+1)} \mathbf{Y}$$

iv) Repeat Steps (i)-(iv) a large number of times and until a convergence criterion is met and obtain  $\hat{\beta}_M$  and  $\hat{\beta}_{MR}$

v) Compute  $L_n$  using  $\hat{\beta}_M$  and

$$\begin{aligned} \hat{\beta}_{MR}^{PT} &= \hat{\beta}_{MR} - (\hat{\beta}_{MR} - \hat{\beta}_0) I(L_n < L_n(\alpha)) \\ \hat{\beta}_{MR}^S &= \hat{\beta}_{MR} - c(\hat{\beta}_{MR} - \hat{\beta}_0)L_n^{-1} \end{aligned}$$


---

**Table 1** Stein-type estimation for mixed and MR

$\rho$	$\beta$	$M$		MR		Shrinkage $M$		Shrinkage MR	
		Estimate	sd	Estimate	sd	Estimate	sd	Estimate	sd
0.0	0.0	-0.01651	0.01897	0.03828	0.01708	-0.01585	0.01833	0.03707	0.01651
	0.1	0.09016	0.01864	0.11907	0.01593	0.08725	0.01802	0.11513	0.01530
	0.2	0.19841	0.01832	0.21519	0.01559	0.19189	0.01777	0.20805	0.01498
	0.4	0.38544	0.02153	0.37015	0.01870	0.37298	0.02110	0.35803	0.01861
	0.8	0.83942	0.01785	0.75063	0.01868	0.81200	0.01640	0.72591	0.01979
MSE		0.294153		0.243336		0.271858		0.243966	
0.2	0.0	-0.00962	0.02271	0.06112	0.02014	-0.00914	0.02193	0.05919	0.01947
	0.1	0.09251	0.02057	0.12655	0.01728	0.08954	0.01989	0.12235	0.01659
	0.2	0.19622	0.02042	0.21971	0.01648	0.18980	0.01980	0.21244	0.01581
	0.4	0.37928	0.02446	0.35959	0.02064	0.36707	0.02396	0.34783	0.02057
	0.8	0.83806	0.01875	0.72513	0.02178	0.81081	0.01734	0.70131	0.02312
MSE		0.370953		0.312257		0.344432		0.307289	
0.5	0.0	-0.00016	0.02968	0.07565	0.02530	0.00004	0.02866	0.07331	0.02445
	0.1	0.09685	0.02508	0.14022	0.02006	0.09381	0.02424	0.13562	0.01926
	0.2	0.19709	0.02523	0.22438	0.01961	0.19065	0.02445	0.21697	0.01883
	0.4	0.35966	0.03151	0.34586	0.02530	0.34827	0.03089	0.33473	0.02511
	0.8	0.84249	0.02363	0.70636	0.02399	0.81521	0.02205	0.68334	0.02541
MSE		0.595086		0.442459		0.55458		0.440721	
0.7	0.0	0.01952	0.03965	0.11723	0.03073	0.01919	0.03831	0.11360	0.02973
	0.1	0.09350	0.03343	0.16157	0.02493	0.09062	0.03233	0.15625	0.02394
	0.2	0.18086	0.03260	0.22593	0.02314	0.17494	0.03164	0.21845	0.02226
	0.4	0.35057	0.03940	0.32776	0.02927	0.33946	0.03852	0.31720	0.02904
	0.8	0.85111	0.03275	0.65915	0.03311	0.82374	0.03098	0.63787	0.03439
MSE		1.026224		0.691952		0.958517		0.686111	
0.9	0.0	0.06336	0.06138	0.21356	0.03809	0.06172	0.05929	0.21040	0.03704
	0.1	0.05429	0.05000	0.21684	0.02302	0.05289	0.04841	0.21351	0.02192
	0.2	0.15672	0.05137	0.20109	0.02179	0.15134	0.04986	0.19811	0.02068
	0.4	0.31727	0.05493	0.29482	0.02799	0.30729	0.05360	0.28893	0.02728
	0.8	0.90350	0.05326	0.52754	0.05286	0.87430	0.05059	0.51431	0.05294
MSE		2.387184		1.109869		2.226858		1.083424	

### Simulation study

A Monte Carlo simulation study is conducted to evaluate the performance of the proposed PTMR and shrinkage estimators compared to the MR estimator of Eliot et al. [13]. In our simulation scheme, we fix  $n_i = 4$  measurements for each subject  $i$  and generate data according to the model of (2) where  $\beta = (0.0, 0.1, 0.2, 0.4, 0.8)$ ,  $b_{ijk} \sim N(0, 0.6)$  and  $\epsilon_{ijk} \sim N(0, 1)$ . To generate multicollinear data, using the “EnvStats” package in R, each predictor variable is assumed to arise from  $N(5, 1)$ , and the correlation  $\rho$  between predictor variables is taken from the set  $\{0.0, 0.2, 0.5, 0.7, 0.9\}$ . Initial values of the variance components are set to be the estimates obtained from fitting a mixed model with no ridge component. In total,  $B = 100$  simulations are conducted for each of the  $\rho$  values based on  $n = 40$  individuals. For further assumed

the expert prior knowledge is  $\beta = \mathbf{0}$ , for the computation of the PTMR and Stein-type estimators. Simulation results are summarized in Tables 1 and 2. From Table 2, it is clear the estimation is dependent on the level of significance  $\alpha$ .

From Table 1, it is apparent the shrinkage mixed ridge (shrinkage MR) estimator has smaller MSE and standard error (sd). Hence, the shrinkage MR is the best among all other competitors; i.e., the shrinkage MR performs better than the mixed, MR and shrinkage mixed (shrinkage  $M$ ). Knowing this, the preliminary test approach is only applied to the mixed ridge estimator, giving rise to the PTMR estimator. According to the results of Table 2, as the level of significance increases, the MSE increases. The graphs of the MSE against the different values of  $\rho$  are also shown in Fig. 1.

**Table 2** Preliminary test estimation for MR

$\rho$	$\beta$	PTMR, $\alpha = 0.001$		PTMR, $\alpha = 0.01$		PTMR, $\alpha = 0.05$		PTMR, $\alpha = 0.1$	
		Estimate	sd	Estimate	sd	Estimate	sd	Estimate	sd
0.0	0.0	0.01246	0.00856	0.03605	0.01578	0.03828	0.01708	0.03828	0.01708
	0.1	0.02731	0.01393	0.10728	0.01466	0.11907	0.01593	0.11907	0.01593
	0.2	0.05582	0.02647	0.18226	0.01764	0.21519	0.01559	0.21519	0.01559
	0.4	0.09815	0.04943	0.32858	0.02632	0.37015	0.01870	0.37015	0.01870
	0.8	0.19518	0.09782	0.66143	0.04317	0.75063	0.01868	0.75063	0.01868
MSE		2.55996		0.55901		0.24334		0.24334	
0.2	0.0	0.01804	0.01010	0.05750	0.01853	0.06112	0.02014	0.06112	0.02014
	0.1	0.03039	0.01426	0.11355	0.01621	0.12655	0.01728	0.12655	0.01728
	0.2	0.06393	0.02630	0.18625	0.01822	0.21971	0.01648	0.21971	0.01648
	0.4	0.10230	0.04901	0.31808	0.02757	0.35959	0.02064	0.35959	0.02064
	0.8	0.20385	0.09668	0.63905	0.04445	0.72513	0.02178	0.72513	0.02178
MSE		2.50707		0.62413		0.31226		0.31226	
0.5	0.0	0.02520	0.01120	0.07434	0.02222	0.07565	0.02530	0.07565	0.02530
	0.1	0.02988	0.01521	0.12458	0.01873	0.14022	0.02006	0.14022	0.02006
	0.2	0.05970	0.02746	0.18715	0.02045	0.22438	0.01961	0.22438	0.01961
	0.4	0.09445	0.04990	0.31025	0.03054	0.34586	0.02530	0.34586	0.02530
	0.8	0.17984	0.09789	0.61851	0.04538	0.70636	0.02399	0.70636	0.02399
MSE		2.61252		0.72805		0.44246		0.44246	
0.7	0.0	0.03759	0.01581	0.11163	0.02780	0.11723	0.03073	0.11723	0.03073
	0.1	0.03963	0.01675	0.14117	0.02351	0.16157	0.02493	0.16157	0.02493
	0.2	0.06801	0.02819	0.18568	0.02332	0.22593	0.02314	0.22593	0.02314
	0.4	0.09917	0.04982	0.29229	0.03410	0.32776	0.02927	0.32776	0.02927
	0.8	0.20379	0.09593	0.58347	0.04987	0.65915	0.03311	0.65915	0.03311
MSE		2.55001		0.95592		0.69195		0.69195	
0.9	0.0	0.05136	0.01859	0.17949	0.03486	0.21356	0.03809	0.21356	0.03809
	0.1	0.06283	0.01780	0.18461	0.02241	0.21684	0.02302	0.21684	0.02302
	0.2	0.04737	0.02795	0.15783	0.02346	0.20109	0.02179	0.20109	0.02179
	0.4	0.07633	0.05073	0.25388	0.03453	0.29482	0.02799	0.29482	0.02799
	0.8	0.14872	0.10104	0.47422	0.06376	0.52754	0.05286	0.52754	0.05286
MSE		2.83244		1.37271		1.10987		1.10987	

Although as level of multi-collinearity increases, so does the MSE values, the proposed PTMR estimator has smaller MSE among all. Further, the PTMR and shrinkage MR estimators perform better than the M and MR estimators in multi-collinear situations.

### HIV data analysis

In a similar framework as explained in Temesgen and Kebede [4], this section focuses on analyzing HIV data using the linear mixed model. In particular, in this study, we analyze the performance of the proposed estimators using the aids dataset taken from “JMbayes” package in R. The dataset consists of seven covariates for each  $n = 467$  patients. The response variable is the CD4, and we consider the gender and prevOI variables as the random effects, in our study. Didanosine versus zalcitabine in HIV patients, a randomized

clinical trial data were collected to compare the efficacy and safety of two antiretroviral drugs in treating patients who had failed or were intolerant of zidovudine (AZT) therapy. Table 3 introduces variables of this study:

Variables have been measured  $n_i$  time for each individuals, so we have a longitudinal dataset; some of the variables like gender in this set will put the subjects in special groups, so we can consider these variables as random effects and we can use mixed models to analyze this dataset, but we know the high degree of correlation among predictors is expected, because variables are measured several times, so we have multi-collinearity data; to combat this difficulty, we can use mixed ridge regression (see Eliot et al. [13]). Information on the important variables is summarized in Table 4.

In addition, we use shrinkage methods to increase estimation efficiency. For the purpose of utilizing the mixed model in “Shrinkage approach” section, the log transform was applied to the CD 4 counts.

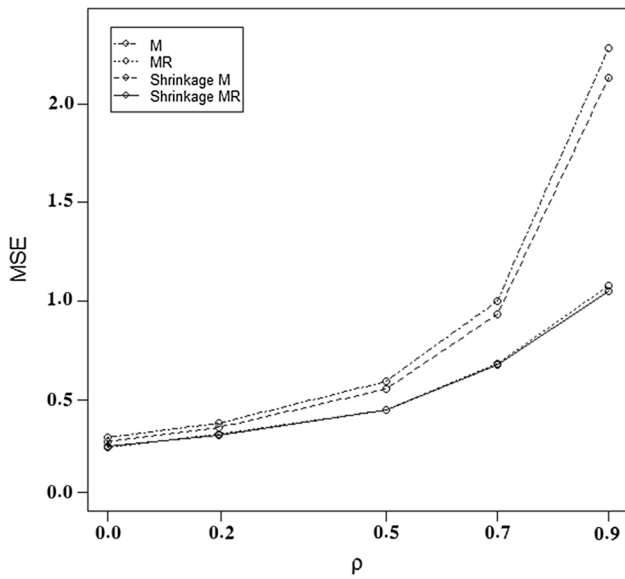


Fig. 1 MSE of estimators versus  $\rho$

Table 5 shows the mixed, MR, shrinkage  $M$ , shrinkage MR and PTMR estimators, respectively, denoted by  $\hat{\beta}_M$ ,  $\hat{\beta}_{MR}$ ,  $\hat{\beta}_M^S$ ,  $\hat{\beta}_{MR}^S$  and  $\hat{\beta}_{MR}^{PT}$ . From Table 6, it is clear that the drug ddI is more effective in curing the infected cells than ddC. In fact, the rate of improvement through using ddI increases by 63.41% as compared to using ddC. This conforms with the negative sign of the parameter “obstime” which indicates that the proliferation of the HIV virus is under control as the time point increases by an approximate rate of 5.715%. The negative sign in the “death” parameter also denotes that using ddC, the rate of deaths and the time to survival improve by 47.83% and 8.36% respectively. Besides, the standard deviation estimates illustrate that these explanatory variables are significant, and hence based on the above dataset, the drug ddI (didanosine) is a recommendable remedy.

From the medical point of view as well, it is shown that ddI yields better treatment in controlling the growth of the HIV virus in the human body (see Molina et al. [26]) while the drug ddC (zalcitabine) has been strongly recommended to be unused due to its countereffects as discussed in the book “clinical neurotoxicology” by Dobbs [27] and Bilgrami and O’Keefe [28].

Table 3 Introduction to data and variables format

Variables	Description
Patient	Patients identifier, in total there are 467 patients
Time	The time to death or censoring
Death	A numeric vector with 0 denoting censoring and 1 death
CD4	The CD4 cells counts
Obstime	The time points at which the CD4 cells count was recorded
Drug	A factor with levels ddC denoting zalcitabine and ddI denoting didanosine (ref)
Gender	A factor with levels female (ref) and male
prevOI	A factor with levels AIDS denoting previous opportunistic infection (AIDS diagnosis) at study entry, and noAIDS denoting no previous infection
AZT	A factor with levels intolerance (ref) and failure denoting AZT intolerance and AZT failure, respectively

Table 4 Summary of dataset

	Time	CD4	Obstime	
Min	0.47	0.00	0.00	
1st. Q	12.23	3.16	0.00	
Median	14.07	5.47	2.00	
Mean	13.89	7.02	4.21	
3rd. Q	17.00	10.44	6.00	
Max	21.40	24.12	18.00	
death	drug	gender	prevOI	AZT
Death: 412	ddI: 688	Male: 1288	AIDS: 863	Failure: 491
Censoring: 993	ddC: 717	Female: 117	noAIDS: 542	Intolerance: 914

To compare the performance of the shrinkage MR estimator, we evaluate the MPE; the lesser, the better. In what follows, we describe the scheme we used to derive the MPE. 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,  $\{(\mathbf{X}^{\text{test}}, \mathbf{y}^{\text{test}})\}$ , termed as test set, while the remaining  $K - 1$  subsets, called the training set, are used to fit model. The resultant estimator is called  $\hat{\beta}^{\text{train}}$ . The fitted model is then used to predict the responses of test dataset. Finally, prediction errors are obtained by taking the squared deviation of the observed and predicted values in the test set, i.e.,

$$PE^k = \|\mathbf{X}_k^{\text{test}} \beta - \hat{\mathbf{y}}_k^{\text{test}}\|^2$$

where  $\hat{\mathbf{y}}_k^{\text{test}} = \mathbf{X}_k^{\text{test}} \hat{\beta}_k^{\text{train}}$ . 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 the average prediction error is estimated which is given by

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

where  $PE_i^k$  is the prediction error of considering  $k$ th test set in  $i$ th iteration. If the value of MPE is lesser, the estimator is preferred, comparatively.

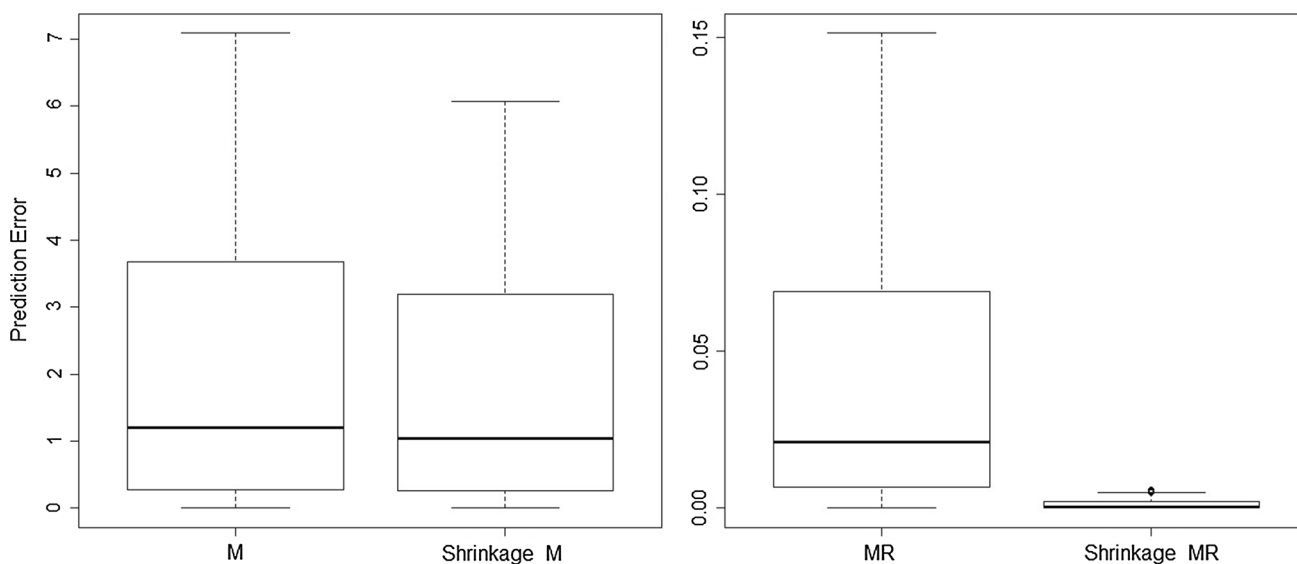
Our results are based on  $N = 500$  case re-sampled bootstrap sample. In Table 5, we report the estimates and MPE

**Table 5** Estimations of real data

Variables	$\hat{\beta}_M$	$\hat{\beta}_{MR}$	$\hat{\beta}_M^S$	$\hat{\beta}_{MR}^S$	$\hat{\beta}_{MR}^{PT}(\alpha = 0.5)$
Time	0.06869698	0.2073583	0.02856897	0.08031050	0.2073583
Death	-2.04918509	-1.0093676	-0.85219338	-0.39093109	-1.0093676
Obstime	-0.14989040	-0.1434991	-0.06233483	-0.05557763	-0.1434991
Drug	0.52531199	0.3335692	0.68193879	0.49110137	0.3335692
AZT	0.66381892	0.1724118	0.27606198	0.06677560	0.1724118
MPE	1.20687	0.02114	1.04644	0.00052	0.02114

**Table 6** Standard deviation (sd) estimations

Variables	$sd(\beta_M)$	$sd(\beta_{MR})$	$sd(\beta_M^S)$	$sd(\beta_{MR}^S)$	$sd(\beta_{MR}^{PT})$
Time	0.0006130127	1.666749e-04	0.0005790592	9.235027e-05	1.666749e-04
Death	0.0157215596	1.170777e-02	0.0152335866	3.478280e-04	1.170777e-02
Obstime	0.0011262431	4.908799e-04	0.0010734227	8.779722e-05	4.908799e-04
Drug	0.0374385567	1.637019e-02	0.0344460129	4.589694e-04	1.637019e-02
AZT	0.0044424233	4.234534e-03	0.0041215722	8.977002e-04	4.234534e-03



**Fig. 2** Box plots for the PEs of the real data

values. Based on the results, the proposed shrinkage MR estimator performs better than the others, in MPE sense. Further, the absolute value of estimates in the shrinkage MR estimates is lesser than the others. The box plots of the PE are shown in Fig. 2.

From the results in the estimation table, it could be deduced that the didanosine drug provides a better treatment.

## Conclusion

In this paper, we developed a preliminary test and Stein-type ridge regression estimation in linear mixed model for longitudinal data analysis. Hence, we considered a penalized likelihood approach and proposed the shrinkage mixed ridge estimator for the vector of regression coefficients. An EM algorithm is also exhibited to solve the penalized likelihood for the unknown parameters. Simulation studies demonstrated the good performance of the proposed estimator for multicollinear situations compared to the maximum likelihood estimator. In addition, the above model has contributed largely to justify the use of didanosine in improving the health states of HIV patients, as stated in various biomedical studies. Henceforth, such model and its estimation step based on shrinkage is highly commendable for medical studies of such genre.

**Acknowledgements** We would like to thank the referees for constructive comments which greatly improved the presentation of paper.

**Author contribution** All authors have contributed equally in the entire work from writing the draft until the last revision.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Mamode Khan, N., Jowaheer, V., Sunecher, Y., Bourignon, M.: Modeling longitudinal INMA(1) with COM–Poisson innovation under non-stationarity: application to medical data. *Comput. Appl. Math.* **6**(17), 1–22 (2018)
- Yuan, S., Zhang, H.H., Davidian, M.: Variable selection for covariate-adjusted semiparametric inference in randomized clinical trials. *Stat. Med.* **31**, 3789–3804 (2012)
- Verbeke, G., Fieuws, S., Molenberghs, G., Davidian, M.: The analysis of multivariate longitudinal data: a review. *Stat. Methods Med. Res.* **23**(1), 42–59 (2014)
- Temesgen, A., Kebede, T.: Joint modeling of longitudinal CD4 count and weight measurements of HIV/tuberculosis co-infected patients at Jimma University specialized hospital. *Ann. Data Sci.* **3**(3), 321–338 (2016)
- Seyoum, A., Ndlovu, P., Temesgen, Z.: Joint longitudinal data analysis in detecting determinants of CD4 cell count change and adherence to highly active antiretroviral therapy at Felege Hiwot

- Teaching and Specialized Hospital, Northwest Ethiopia (Amhara Region). *AIDS Res. Ther.* **14**(14), 1–13 (2017)
- Zeger, S.L., Liang, K.Y.: Longitudinal data analysis for discrete and continuous outcomes. *Biometrika* **42**(1), 121–130 (1986)
- Thall, P., Vail, S.: Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**(3), 657–671 (1990)
- Laird, N., Ware, J.: Random-effects models for longitudinal data. *Biometrics* **38**, 963–974 (1982)
- Sutradhar, B.C.: An overview on regression models for discrete longitudinal responses. *Stat. Sci.* **18**(3), 377–393 (2003)
- Sutradhar, B.C., Jowaheer, V.: Analyzing longitudinal count data from adaptive clinical trials: a weighted generalized quasi-likelihood approach. *J. Stat. Comput. Simul.* **76**(12), 1079–1093 (2006)
- FitzMaurice, G.M., Laird, N.M.: A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80**(1), 141–151 (1993)
- Sutradhar, B.C., Jowaheer, V., Rao, P.: Remarks on asymptotic efficient estimation for regression effects in stationary and non-stationary models for panel count data. *Braz. J. Prob. Stat.* **28**(2), 241–254 (2014)
- Eliot, M., Ferguson, J., Reilly, M.P., Foulkes, A.S.: Ridge regression for longitudinal biomarker data. *Biostatistics* **7**, 1–11 (2011)
- Hossain, S., Thomson, T., Ahmed, E.: Shrinkage estimation in linear mixed models for longitudinal data. *Metrika* **81**(5), 569–586 (2018)
- Saleh, A.K.E., Arashi, M., Tabatabaey, S.M.M.: *Statistical Inference for Models with Multivariate t-Distributed Errors*. Wiley, New Jersey (2014)
- Ali, A.M., Saleh, A.K.E.: Estimation of the mean vector of a multivariate normal distribution under symmetry. *J. Stat. Comput. Simul.* **35**, 209–226 (1990)
- Ahmed, S.E., Fallahpour, S.: Shrinkage estimation strategy in quasi-likelihood models. *Stat. Probab. Lett.* **82**, 2170–2179 (2012)
- Roohbeh, M., Arashi, M.: Shrinkage ridge regression in partial linear models. *Commun. Stat. Theory Methods* **45**(20), 6022–6044 (2016)
- Yuzbasi, B., Ahmed, S.E.: Shrinkage and penalized estimation in semi-parametric models with multicollinear data. *J. Stat. Comput. Simul.* **86**(17), 3543–3561 (2016)
- Yuzbasi, B., Ahmed, S.E., Gungor, M.: Improved penalty strategies in linear regression models. *REVSTAT Stat. J.* **15**(2), 251–276 (2017)
- Asar, Y.: Some new methods to solve multicollinearity in logistic regression. *J. Commun. Stat. Simul. Comput.* **46**(4), 2576–2586 (2017)
- Bancroft, T.A.: On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Stat.* **15**, 190–204 (1944)
- Stein, C.: Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 197–206 (1956)
- James, W., Stein, C.: Estimation with quadratic loss. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 361–379 (1961)
- Hoerl, A., Kennard, R.: Ridge regression: applications to non-orthogonal problems. *Technometrics* **12**, 69–82 (1970)
- Molina, J.M., Marcelin, A.G., Pavie, J., Heripret, L., Boever, C.M.D., Troccaz, M., Leleu, G.: Didanosine in HIV-1 infected patients experiencing failure of antiretroviral therapy: a randomized placebo-controlled trial. *J. Infect. Dis.* **191**(6), 840–847 (2005)
- Dobbs, M.R.: *Clinical Neurotoxicology: Syndromes, Substances, Environments*. Elsevier, Amsterdam (2009)
- Bilgrami, M., O’Keefe, P.: Neurologic diseases in HIV-infected patients. *Handb. Clin. Neurol.* **121**, 1321–1344 (2014)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.