Random and Fixed Effects Selection for Weighted Ridge

Lulah Alnaji*

Department of Mathematics, College of Sciences, University of Hafr Al Batin, Saudi Arabia

**Corresponding author:* laalnaji@uhb.edu.sa

Abstract. Using penalized profiled log-likelihood and penalized limited profiled log-likelihood, respectively, together with the weighted ridge penalized term, we offer a method in this study for choosing the fixed and random effects in linear mixed models. Then, we use the penalized restricted profiled log-likelihood to perform in the random effects depending on the chosen tuning parameter. Second, we use the penalized profiled log-likelihood to choose the fixed effect parameters. There is no closed-form solution for the choice of the fixed and random effects, hence the Newton-Raphson technique is employed to iteratively estimate the parameters. We use a simulation study to show how well the suggested strategy works. Lastly, we use two separate datasets to use the methods to further evaluate the newly proposed model.

1. Introduction

With longitudinal data, each individual is followed repeatedly across various times in time. Thus, the independence assumption is not optimal because of the associated observations of each participant. The linear mixed model is a common option for longitudinal research and is a helpful tool since it incorporates the random effects to account for the within-subjects correlation. [1].

Due to the fixed and random effect parameters' increased dimension during the past 20 years, there have been certain challenges. The issue of variable selection has been researched, and many different approaches have been put forth in order to narrow the selection of parameters to those that are most crucial, such as Ridge Regression [2], LASSO Method [3], Adaptive LASSO [4], Elastic Net [5] and SCAD [6] among many others.

Selection criteria like Akaike information criterion (AIC [7]) and Bayesian information criterion (BIC [8]) have been used for variable selection and proved to provide a consistent selection model rules.

In particular, BIC is asymptotically consistent for model selection. Many research have extensively explored and used the asymptotic qualities when the number of possible aggressors is fixed [9].

Fixed effects selection in the linear mixed-effects model using adaptive ridge process for $L0$ penalty performance is one study that has focused a lot of emphasis on one-stage variable selection using penalized log-likelihood approach for the fixed effects [10]; and model selection in linear mixed effect m els [11]. Although the methods listed above are highly helpful, the fixed and random effects' fundamental features are quite different, therefore the methods listed above might not reveal these differences.

Adaptive LASSO has been the subject of various studies for variable selection methods such as [12], they investigated pathwise coordinate optimization and fixed and random effects selection via REML. Also, profile log-likelihood-based adaptive LASSO for linear mixed model selection [13]. They discuss selecting the fixed and random effects using ML and REML procedures.

For weighted ridge, [14] studied a weighted ridge procedure for $L0$ regularization in fixed models. The linear mixed-effects model was used to study the selection of fixed effects, and the fixed effects, random effects, and variance components were estimated using the weighted ridge approach for $L0$ penalty performance [10]. We provide a model selection process for a weighted Ridge in mixed models for both random and fixed effects, respectively, to further enhance the behavior of the current penalized techniques.

The remaining sections of this article are structured as follows. Section 2 presents variable selection for the linear mixed-effects model. The methodology of the weighted ridge is presented in Section 3. Section 4 presents simulation studies and Section 5 presents the important conclusions from this study.

2. Variable selection for the linear mixed-effects model

Maximum likelihood (ML) and restricted maximum likelihood (REML) are major methods have been proposed to estimate the parameters in 2.1 when assuming that λ_j and ϵ_j are normally distributed. See Random-effects models for longitudinal data [15], Unbalanced repeated-measures models with structured covariance matrices [16], and Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data [17].

In this article, restricted maximum likelihood (REML) and Maximum likelihood (ML) are used to select the random effects and the fixed effects, respectively.

2.1. **Literature review of classical linear mixed-effect mode.** In this section, we consider the classical linear mixed-effect model setting to establish a selection method for weighted ridge mixed model.

$$Y_j = X_j \psi + Z_j \lambda_j + \epsilon_j, \quad j = 1, \cdots, m, \tag{2.1}$$

where Y_j is the $m_j \times 1$ response vector for the observations of subject j, $\psi \in R^r$ is the fixed effects vector corresponding to its $m_j \times r$ ull rank design matrix X_j, $\lambda_j \in R^l$ is the random effects vector corresponding to its $m_j \times l$ design matrix Z_j and ϵ_j is the $m_j \times 1$ vector of the model errors.

Denote the total of observations $M = \sum_{j=1}^{m} m_j$. Assume that $Y_1, \cdots, Y_m$ are independent and that ϵ_j and λ_j are independent with $\epsilon_j \sim N(0, \sigma^2 I_{m_j})$, $\lambda_j \sim MVN(0, G)$, hence $Y_j \sim N(X_j\psi, \sigma^2 D_j(\alpha))$ where $D_j(\alpha) = Z_j G Z_j' + I_{m_j}\sigma^2$ and α is the $k = 1/2 * (l(l+1))$ vector that consists of the unknown covariance parameters which characterizes the matrix G.

The model 2.1 is a classical model and the estimation of the fixed and random effect parameters can be done by using the well-known methods, unbiased estimation(BLUE) and best linear unbiased prediction (BLUP), respectively, using maximum likelihood (ML) approach [18].

2.2. **Selection of Random Effects Parameters.**

Due to its unpredictable nature, which presents greater difficulties in estimating the variance-covariance matrix's structure, random effects selection hasn't garnered as much attention as fixed effects selection, as was previously indicated. Although the parameters for the fixed effects are not very sensitive to the choice of random effects, choosing random effects incorrectly can have an impact on how effectively the fixed effects are estimated.

To implement the selection of the random effects by selecting a parameter α that maximizes the penalized restricted profiled log-likelihood. Consider the weighted ridge [14] for the penalty term

$$
\begin{aligned}
T_{ran}(\alpha) = & -\frac{1}{2}\log\left|\sum_{j=1}^{m} X_j' D_j X_j\right| - \frac{1}{2}\sum_{j=1}^{m} \log|D_j| \\
& -\frac{1}{2}M \times \log\left(\sum_{j=1}^{m} e' D_j^{-1} e\right) \\
& +\frac{1}{2}r \times \log\left(\sum_{j=1}^{m} e' D_j^{-1} e)\right) \\
& -\gamma_{1m}\sum_{j=1}^{l} w_{1j} g_j^2
\end{aligned}
\tag{2.2}
$$

where $\gamma_{1m} \geq 0$ is the tuning parameter (also called the regularization parameter), $W_1 = \text{diag}(w_{11}, \cdots, w_{1l})$ is $l \times l$ diagonal matrix and w_{1j} is the j-th element of W_1's diagonal (See Section 3), $e = (Y_j - X_j\hat{\psi})$ and g_j is the j-th element of the diagonal of the matrix G.

The primary interest lies on the selection of ψ and α. The proposed selection variable method has no closed-form solution and can be solved iteratively. For example, Newton-Raphson is a popular iterative method to be used. The first and the second derivatives of Newton-Raphson method can be found in details in [16].

2.3. **Selection of Fixed Effects Parameters.** To select the important covariates of the fixed effects, we propose to maximize the penalized profiled log-likelihood

$$
\begin{aligned}
T_{fix}(\psi) = & -\frac{1}{2}\sum_{j=1}^{m}\log|D_j| \\
& -\frac{1}{2}M \times \log\left(\sum_{j=1}^{m}(Y_j - X_j\psi)'D_j^{-1}(Y_j - X_j\psi)\right) \\
& -\gamma_{2n}\sum_{j=1}^{r}w_{2j}\psi_j^2
\end{aligned}
\tag{2.3}
$$

where the tuning parameter (also called the regularization parameter) is $\gamma_{2n} \geq 0$, $W_2 = \mathrm{diag}(w_{21}, \cdots, w_{2q})$ is $l \times l$ diagonal matrix and w_{2j} is the j-th element of W_2's diagonal (See Section 3). Note that at this step, the penalized profiled log-likelihood 2.3 is a function of ψ only, since the matrix D_j always selected by the first stage using the penalized restricted profiled log-likelihood 2.2.

After selecting the matrix D_j in the first stage by maximizing the penalized restricted profiled log-likelihood 2.2, in the second stage the primary interest lies in the selection of the covariates of ψ. Similarly, the proposed selection variable method has no closed-form solution and can be solved iteratively by the Newton-Raphson method of the form

$$
\psi_{t+1} = \psi_t - H^{-1}S_\psi, \qquad t = 0, 1, 2 \cdots
\tag{2.4}
$$

where ψ_t is the current result of 2.3, ψ_{t+1} is the updated one, S_ψ is the score vector and $H_{\psi\psi}$ is the Hessian matrix.

3. The methodology of the weighted ridge

[14] proposed a weighted ridge strategy that improves the performance of the L_0-penalty for fixed effect models, motivated by Least absolute shrinkage is equivalent to quadratic penalization [19]; One-step sparse estimates in non-concave penalized likelihood models [20] and Visualization of genomics changes by segmented smoothing using an L0 penalty [21]. In (2017), [14] has been extended to a linear mixed effects model by [10]. They proposed a selection strategy for the fixed effects while the covariance matrix for the random effects is Cholesky factorized. In this article, we proposed a selection approach for both the fixed and random effects using an iteratively weighted ridge strategy following the procedure of the weight matrix in [14]

Consider the objective function 2.2 where the weight matrices for for w_{1j} and w_{2j} are calculated by

$$
w_{1j}^t = \frac{1}{|\alpha_j^{(t)}|^2 + \delta^2}
\tag{3.1}
$$

and

$$w_{2j}^t = \frac{1}{|\psi_j^{(t)}|^2 + \delta^2}$$ (3.2)

where w_{1j} is the j-th element of diag(W_1), α_j is the j-th element of the diagonal of the selected covariance matrix in the first stage, t is the number of iteration, δ is a constant, w_{2j} is the j-th element of diag(W_2) and ψ_j is the j-th element of the vector of the fixed effect parameters. In numerical practices small positive choices for δ seem to perform better than $\delta = 0$ (for further details see [10], [22], [14], [21]).

Although, the selection of the tuning parameter γ would add more computational work to search among lattice of γ's but it is an important step. The selection of the tuning parameter is an influential part of penalized methods.

Selection criteria like Akaike information criterion (AIC [7]), Bayesian information criterion (BIC [8]) and Generalized Cross Validation (GCV [23]) have been used for variable selection. For compression we employ the three criteria for the fixed and random effect parameters.

4. Simulation studies

A simulation study is conducted in order to examine the asymptotic properties and the performance of our newly developed method. All of the simulated data are generated according to the model 2.1 using R statistical software. Following the examples in [24], the simulation study assumes repeatedly observation per each subject.

(I) Assume $m_j = 5$ per each subject j with $m = 30$ and consider the true fixed effects vector to be $\psi = (1, 1, 0, 0, 0, 0, 0, 0, 0)$ with $r = 9$. We further consider $l = 4$ for random effects with the assumption of normal distribution for the error of the model and the random effects as, $\epsilon_j \sim N(0, \sigma^2 I)$ and $\lambda_j = (\lambda_{j0}, 0)'$ with $r_{j0} \sim N(0, G)$ where $\sigma^2 = 1$ and the true covariance matrix

$$G = \begin{bmatrix} 9 & 4.8 & 0.6 & 0 \\ 4.8 & 4 & 1 & 0 \\ 0.6 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The design covariates matrix X_j is assumed to arise from a uniform(-2,2). The first column in matrix Z_j are ones for the subject-specific intercept, while the remaining columns are assumed to arise from a uniform(-2,2) as well.

(II) This case follows case (I) with an increase of the sample size. Assume $m = 60$ with $m_j = 10$ and generate 200 dataset following the same methodology.)

(III) This case follows case (I) with an increase of the sample size. Assume $m = 60$ with $m_j = 10$ and generate 500 dataset following the same methodology.)

Table 1. : The Simulation Results for Case (I)

Criteria	BIC	AIC	GCV
%Correct	64	62	65
%CR	69	67	58
%CF	73	70	70

Table 2. : The Simulation Results for Case (II)

Criteria	BIC	AIC	GCV
%Correct	87	82	85
%CR	90	77	71
%CF	94	69	71

Table 3. : The Simulation Results for Case (III)

Criteria	BIC	AIC	GCV
%Correct	90	83	85
%CR	90	80	75
%CF	97	73	78

Table 4. : The simulation results are for case (I) compared to some existing studies

$m = 30$ $m_j = 5$	WRidge	[13]	[12]
%Correct	64	73	61
%CR	69	81	79
%CF	73	88	79

%Correct denotes the percentage of times that the correct model (fixed and random effects) is selected, %CR denotes the percentage of times that the random effects is selected, and %CF denotes the percentage of times that the fixed effects is selected. $m = 30$,$m_j = 5$. While there is a large body

Table 5. : The simulation results are for case (II) compared to some existing studies

$m = 60$ $m_j = 10$	WRidge	[13]	[12]
%Correct	87	92	88
%CR	90	92	91
%CF	94	100	97

of literature on the estimation the parameters, only countable references studied parameters selection. In particular, the selection of the random effect has received less attention than the selection of the fixed effects. [14] studied one-stage of an weighted ridge procedure for L0 regularization in fixed models. Also, [10] studied selection of fixed effects with estimation fixed effects, random effects and variance components in the linear mixed-effects model using weighted ridge procedure for L0 penalty performance.

In Tables 1 and 2, we use AIC, BIC and GCV criteria and it can be seen that the empirical results confirm the asymptotic properties, that is the selection percentages get higher as the sample size increases. In Tables, 4 and 5 we use BIC criteria and compare the results of our newly proposed approach (WRidge) to the results of [13] and [12]. It can be seen that the percentage of selection in for our method is somehow less and that due to the nature of our method, It is known that the weighted ridge method doesn't eliminate some predictors to zero as LASSO does and that

Our newly proposed method performs well across the simulations experiments and the weighted ridge performs well for model selection for both fixed and random effects.

5. Real Data Applications

To show the efficiency of the suggested penalized technique in the mixed model selection, two dataset set are applied.

5.1. **First Dataset.** We use the data from Amsterdam, Growth and Health study, which is a distinctive, interdisciplinary cohort research established to investigate development and health among teens [25]. The information was gathered to investigate the connection between adolescent and early adulthood lifestyle and health. The study had 147 people in all, who were assessed at six different time periods, for a total of 882 observations. The five factors considered are age, gender, body fat, fitness, and smoking. This paper follows [26] and [13] for compression purposes. The random intercept is permitted to have intercept, but this is not the case for the fixed effects, since the response variable $cholesterol$ has been centered and normalized all of the inputs, and hence the fixed effects have no intercept. Table 6 exhibits that the $bodyfat$ and $time$ are selected by all approaches to be important

fixed effects. For random effects selection, *gender* is selected by all approaches, while *fitness* is selected by SAW and WRidge methods, but *smoking* is selected by HARD only.

Table 6. Comparison of All Methods.

	Fixed Effect				Random Effect			
	HARD	SAW	PS	WRidge	HARD	SAW	PS	WRidge
intercept	-	-	-	-	0.405	0.347	0.017	0.211
fitness	0	0	0	0	0	0.006	0	0.001
body_fat	0.174	0.165	0.170	0.168	0	0	0	0
smoking	0	0	0	0	0.149	0	0	0
gender	0	0	0	0	0.668	0.624	0.888	0.691
time	0.156	0.167	0.165	0.161	0	0	0	0

5.2. **Second Dataset.** The data was collected from 27 children to study, the distance (mm) between the pituitary gland's core and the pterygomaxillary fissure for both gender [27]. The distances were measured in 16 boys and 11 girls at ages 8, 10, 12, and 14 for this study. The purpose of the study was to examine the basic age functions for boys and girls before describing the gap between boys and girls as a function of age. The two factors are gender and age while distance is the response. the response variable *distance* has been centered and normalized all of the inputs and hence there is no intercept for fixed effects.

Table 7. Results of WRidge Selection Method.

	Fixed Effect	Random Effect
	WRidge	WRidge
intercept	-	.021
gender	0.623	0.861
age	0.117	0

Table 7 demonstrates that the *gender* and *age* are selected to be important fixed effects. For random effects selection, *gender* is selected by the newly proposed model to be important random effects.

6. Conclusion

In this paper, we propose weighted ridge selection for a linear mixed model and we focus on the case of longitudinal data with N observations coming from n subjects. The simulation studies of

low-dimensional settings show that the newly established method is an efficient method in general, and the percentages of times the random effects, fixed effects, and both of them combined are high percentages.

Conflicts of Interest: The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] G. Verbeke, Linear Mixed Models for Longitudinal Data, in: Linear Mixed Models in Practice, Springer, New York, NY, 1997: pp. 63-153. https://doi.org/10.1007/978-1-4612-2294-1_3.

[2] A.E. Hoerl, R.W. Kennard, Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics. 12 (1970), 55-67.

[3] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, J. R. Stat. Soc.: Ser. B (Methodol.) 58 (1996), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[4] H. Zou, The Adaptive Lasso and Its Oracle Properties, J. Amer. Stat. Assoc. 101 (2006), 1418–1429. https://doi.org/10.1198/016214506000000735.

[5] H. Zou, T. Hastie, Regularization and Variable Selection via the Elastic Net, J. R. Stat. Soc. Ser. B: Stat. Methodol. 67 (2005), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[6] J. Fan, R. Li, Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties, J. Amer. Stat. Assoc. 96 (2001), 1348–1360. https://doi.org/10.1198/016214501753382273.

[7] H. Akaike, A New Look at the Statistical Model Identification, IEEE Trans. Automat. Contr. 19 (1974), 716-723. https://doi.org/10.1109/tac.1974.1100705.

[8] G. Schwarz, Estimating the Dimension of a Model, Ann. Stat. 6 (1978), 461-464. https://www.jstor.org/stable/2958889.

[9] Y. Yang, Can the Strengths of AIC and BIC Be Shared? A Conflict Between Model Indentification and Regression Estimation, Biometrika. 92 (2005), 937-950. https://doi.org/10.1093/biomet/92.4.937.

[10] E. Adjakossa, G. Nuel, Fixed Effects Selection in the Linear Mixed-Effects Model Using Adaptive Ridge Procedure for L_0 Penalty Performance, arXiv:1705.01308. (2017). https://doi.org/10.48550/ARXIV.1705.01308.

[11] H. Peng, Y. Lu, Model Selection in Linear Mixed Effect Models, J. Multivar. Anal. 109 (2012), 109-129. https://doi.org/10.1016/j.jmva.2012.02.005.

[12] B. Lin, Z. Pang, J. Jiang, Fixed and Random Effects Selection by REML and Pathwise Coordinate Optimization, J. Comput. Graph. Stat. 22 (2013), 341–355. https://doi.org/10.1080/10618600.2012.681219.

[13] J. Pan, J. Shang, Adaptive LASSO for Linear Mixed Model Selection via Profile Log-Likelihood, Commun. Stat. - Theory Methods. 47 (2017), 1882-1900. https://doi.org/10.1080/03610926.2017.1332219.

[14] F. Frommlet, G. Nuel, An Adaptive Ridge Procedure for L_0 Regularization, PLoS ONE. 11 (2016), e0148620. https://doi.org/10.1371/journal.pone.0148620.

[15] N.M. Laird, J.H. Ware, Random-Effects Models for Longitudinal Data, Biometrics. 38 (1982), 963. https://doi.org/10.2307/2529876.

[16] R.I. Jennrich, M.D. Schluchter, Unbalanced Repeated-Measures Models with Structured Covariance Matrices, Biometrics. 42 (1986), 805-820. https://doi.org/10.2307/2530695.

[17] M.J. Lindstrom, D.M. Bates, Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data, J. Amer. Stat. Assoc. 83 (1988), 1014–1022. https://doi.org/10.1080/01621459.1988.10478693.

[18] D.A. Harville, Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems, J. Amer. Stat. Assoc. 72 (1977), 320-338.

[19] Y. Grandvalet, Least Absolute Shrinkage is Equivalent to Quadratic Penalization, in: L. Niklasson, M. Boden, T. Ziemke (Eds.), ICANN 98, Springer London, London, 1998: pp. 201-206. https://doi.org/10.1007/978-1-4471-1599-1_27.

[20] P. Bühlmann, L. Meier, H. Zou, Discussion of "One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models" by H. Zou and R. Li, Ann. Stat. 36 (2008), 1534-1541.

[21] R.C.A. Rippe, J.J. Meulman, P.H.C. Eilers, Visualization of Genomic Changes by Segmented Smoothing Using an L_0 Penalty, PLoS ONE. 7 (2012), e38230. https://doi.org/10.1371/journal.pone.0038230.

[22] E.J. Candes, M.B. Wakin, S.P. Boyd, Enhancing Sparsity by Reweighted ℓ_1 Minimization, J. Fourier Anal. Appl. 14 (2008), 877-905. https://doi.org/10.1007/s00041-008-9045-x.

[23] M. Stone, An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion, J. R. Stat. Soc.: Ser. B (Methodol.) 39 (1977), 44-47. https://doi.org/10.1111/j.2517-6161.1977.tb01603.x.

[24] H.D. Bondell, A. Krishna, S.K. Ghosh, Joint Variable Selection for Fixed and Random Effects in Linear Mixed-Effects Models, Biometrics. 66 (2010), 1069-1077. https://doi.org/10.1111/j.1541-0420.2010.01391.x.

[25] J.W.R. Twisk, H.C.G. Kemper, G.J. Mellenbergh, Longitudinal Development of Lipoprotein Levels in Males and Females Aged 12-28 Years: The Amsterdam Growth and Health Study, Int. J. Epidemiol. 24 (1995), 69-77. https://doi.org/10.1093/ije/24.1.69.

[26] M. Ahn, H.H. Zhang, W. Lu, Moment-Based Method for Random Effects Selection in Linear Mixed Models, Stat. Sinica. 22 (2012), 1539-1562. https://doi.org/10.5705/ss.2011.054.

[27] R.F. Potthoff, S.N. Roy, A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems, Biometrika. 51 (1964), 313-326. https://doi.org/10.1093/biomet/51.3-4.313.