Journal of Data Science and Modeling, Vol. 1, No. 1, 159-170, December 2022

Research Manuscript

Semiparametric Ridge Regression for Longitudinal Data

Mozhagn Taavoni[1*], Mohammad Arashi[2]

[1] Ph.D. candidate in statistics, Shahrood University of Technology, Shahrrod, Iran,

[2] Associate professor of statistics, Shahrood University of Technology, Shahrrod, Iran.

Received: 28/10/2018 Accepted: 29/11/2018

Abstract: This paper considers an extension of the linear mixed model, called semiparametric mixed-effects model, for longitudinal data, when multicollinearity is present. To overcome this problem, a new mixed ridge estimator is proposed, while the nonparametric function in the semiparametric model is approximated by the kernel method. The proposed approach integrates the ridge method into the semiparametric mixed-effects modeling framework to account for both the correlation induced by repeatedly measuring an outcome on each individual over time, as well as the potentially high degree of correlation among possible predictor variables. The asymptotic normality of the exhibited estimator is established. To improve efficiency, the estimation of the covariance function is accomplished using an iterative algorithm. Performance of the proposed estimator is compared through a simulation study and analysis of the CD4 data.

Keywords: Kernel, Longitudinal data, Mixed-effect, Ridge regression, Semiparametric.

Mathematics Subject Classification (2010): 62J07, 62620.

*Corresponding author: taavonimozhgan@yahoo.com

1. Introduction

Longitudinal data frequently arise in biological and economic applications. A major difficulty in the analysis of longitudinal data is that the data are subject to within subject correlation among repeated measurements over time. In recent years, various traditional methods for longitudinal data have been developed. However, mixed-effect models and semiparametric models are popularly applied. In this paper, we consider the following semiparametric mixed-effects model

$$Y_i(t_{ij}) = X_i^\top(t_{ij})\beta + g(t_{ij}) + Z_i^\top(t_{ij})b_i + \varepsilon_i(t_{ij}), \tag{1.1}$$

where $Y_i(t_{ij})$ is the response for the ith subject at time point t_{ij}, $\beta = (\beta_1, \dots, \beta_p)^\top$ is a $p \times 1$ vector of regression coefficients associated with the covariates $X_i(t_{ij})$; $g(t_{ij})$ is an unknown twice differentiable smooth function of time; the b_i are *i.i.d.*, unobservable $q \times 1$ vectors of the random effects associated with covariates $Z_i(t_{ij})$ with mean zero and covariance matrix D_i; and the $\varepsilon_i(t_{ij})$, independent of b_i, is random error with $E\{\varepsilon_i(t_{ij})\} = 0$. Without loss of generality, we assume every t_{ij} is scaled into the interval $[0, 1]$. Similar to Fan and *et al.* (2007), we assume $Var\{\varepsilon_i(t_{ij})\} = \sigma^2(t_{ij})$, which is a nonparametric smooth function, but the correlation function between $\varepsilon_i(t_{ij})$ and $\varepsilon_i(t_{ik})$ has a parametric form, $corr\{\varepsilon_i(t_{ij}), \varepsilon_i(t_{ik})\} = \rho(t_{ij}, t_{ik}, \theta)$, where $\rho(t_{ij}, t_{ik}, \theta)$ is a positive definite function of t_{ij} and t_{ik}, and θ is an unknown parameter vector. We assume both the random effects and the errors are normally distributed and we have a random sample of n subjects with the ith subject having n_i observations over time and total observation of $N = \sum_{i=1}^{n} n_i$.

In the mixed-effects model, random component takes care the correlation among observations from the same subject, [see Fitzmaurice and *et al.* (2004) and Laird and Ware (1982)]. Since the parametric assumption in this model is not always adequate, it is of interest to model the time effect nonparametrically while accounting for the correlation within the same subject. In such cases, the semiparametric models are widely used in longitudinal studies. Related works include Fan and Li (2004), Hu and *et al.* (2004) and Lin and Ying (2001). Model (1.1) is a natural extension of linear mixed models and semiparametric models called semiparametric mixed-effect model that uses parametric fixed effects to present the covariate effects and an arbitrary smooth function to model the time effect to account for the within-subject correlation using random effects. The parametric component provides a simple summary of covariates effects, which are of main scientific interest, while the baseline function is included for flexibility. With both parametric and nonparametric components, the proposed model is more flexible than the traditional linear model. Zeger and Diggle (1994) used a semiparametric random intercept model to analyze the CD4 cell numbers in HIV seroconverters, where the nonparametric function is estimated by backfitting method. Our work is an extension of their model. Indeed, we consider a more general class of mixed-effects models with a kernel estimator of nonparametric function.

In addition to proposing a more flexible model (1.1), we consider the problem of multicollinearity of predictor variables and estimation in the proposed model. Ridge regression (RR), also known as Tikhonov regularization [see Tikhonov (1943)], is a well-known penalized regression approach to handling multiple colinear predictor variables that involves adding a regularization term to the least-squares equation in order to derive parameter estimates in the context of an ill-conditioned or singular design matrix [see Hoerl and Kennard (1970a, b)]. The resulting shrinkage estimates, while biased, offer improved prediction accuracy, that is, reduced prediction variance, and thus may be preferable in settings with a large number of highly correlated independent variables, in which unique least-squares solutions are not tenable. Recent applications of RR to longitudinal data include, for example, Zhang and Horvath (2006), Malo and *et al.* (2008), and Eliot and *et al.* (2011). In the present paper, we extend RR to the correlated response setting in which a single outcome of interest is measured repeatedly over time at potentially unevenly spaced time intervals. In the case of uncorrelated or minimally correlated predictor variables, mixed-effects models with a person-specific random intercept and slopes terms can be applied to account for the within-person correlation in making inferences about the association. Here we integrate RR and the mixed-effects modeling frameworks in order to account both for the correlation induced by repeatedly measuring the outcome on each individual over time, as well the potentially high degree of correlation among potential predictor variables.

We begin in Section 2 by providing a brief background on kernel estimation of nonparametric function while proposed a new estimator, which we term the mixed ridge (MR) estimator, for longitudinal data. Further, its asymptotic normality is provided. In Section 3, an iterative algorithm is provided to calculate the estimator. A simulation study to evaluate the performance and application to real data are then provided in Sections 4. We conclude your paper in section 5.

2. Estimation procedure

Denote $Y_i = \big(Y_i(t_{i1}), \ldots, Y_i(t_{in_i})\big)^\top$; X_i, Z_i, ε_i and t_i similarly. The model (1.1) can be presented as

$$Y_i = X_i^\top \beta + g(t_i) + Z_i^\top b_i + \varepsilon_i,$$

where $g(t_i) = \big(g(t_{i1}), \ldots, g(t_{in_i})\big)^\top$. Let $V_i = Z_i D_i Z_i^\top + \Sigma_i$ be the $n_i \times n_i$ covariance matrix of Y_i and $\Sigma_i = A_i R_i(\theta) A_i$ be the $n_i \times n_i$ working covariance matrix of ε_i, where $A_i = diag(\sigma(t_{i1}), \ldots, \sigma(t_{in_i}))$ and $R_i(\theta)$ is the correlation matrix of ε_i with (j,k) element equaling $\rho(t_{ij}, t_{ik}, \theta)$. Under normality assumption of error term and random effects,

$$Y_i \sim N\big(X_i\beta + g(t_i), V_i\big). \tag{2.2}$$

Let $\mu_Y(t_{ij}) = E\{Y_i(t)|t = t_{ij}\}$; $\mu_X(t_{ij})$ defined equivalently. It is evident that $\mu_Y(t_{ij}) = \mu_X^\top(t_{ij})\beta + g(t_{ij})$. This together with equation (2.2) yields

$$Y_i - \mu_Y(t_i) \sim N\big([X_i - \mu_X(t_i)]^\top \beta, V_i\big).$$

Further, denoting $Y = (Y_1^\top, \ldots, Y_n^\top)^\top$, and X, ε similarly and $Z = diag(Z_1^\top, \ldots, Z_n^\top)$, we have matrix notation of model (1.1) as

$$Y = X^\top \beta + g(t) + Z^\top b + \varepsilon,$$

where $g(t) = \big(g(t_1)^\top, \ldots, g(t_n)^\top\big)$ and $b = (b_1^\top, \ldots, b_n^\top)^\top$ is $(n \times q) \times 1$ vector of random effects with mean zero and covariance matrix $D = diag(D_1, \ldots, D_n)$. Let $V = ZDZ^\top + \Sigma$ be the $N \times N$ block diagonal covariance matrix of Y and Σ be the $N \times N$ block diagonal covariance matrix of ε.

Further, define $\breve{Y}_i = Y_i - \mu_Y(t_i)$, $\mu_Y(t_i) = \big(\mu_Y(t_{i1}), \ldots, \mu_Y(t_{in_i})\big)^\top$; $\breve{X}_i$ and $\mu_X(t_i)$ similarly. Finally, for matrix notation, we have $\breve{Y} = (\breve{Y}_1^\top, \ldots, \breve{Y}_n^\top)^\top$ and $\breve{X} = (\breve{X}_1^\top, \ldots, \breve{X}_n^\top)^\top$.

A weighted least squares (WLS) criterion for β is

$$l(\beta) = \sum_{i=1}^{n} (\breve{Y}_i - \breve{X}_i^\top \beta)^\top V_i^{-1} (\breve{Y}_i - \breve{X}_i^\top \beta).$$

the term $l(\beta)$ cannot be directly used in statistical inference because it contains unknown functions $\mu_Y(.)$, $\mu_X(.)$. We employ the kernel smoother to estimate them, and the estimators are respectively defined as

$$\widehat{\mu}_X(t) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \omega_{ij} X_i(t_{ij}), \qquad \widehat{\mu}_Y(t) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \omega_{ij} Y_i(t_{ij}),$$

where $\omega_{ij} = K_h(t_{ij}-t)/\sum_{k=1}^{n}\sum_{l=1}^{n_k} K_h(t_{kl}-t)$, h is a bandwidth, $K_h(.) = K(./h)$ and $K(.)$ is a kernel function.

Therefore, an estimator $\widehat{l}(\beta)$ of $l(\beta)$ can be obtained by substituting $\mu_Y(t)$ and $\mu_X(t)$ of $l(\beta)$ with $\widehat{\mu}_Y(t)$ and $\widehat{\mu}_X(t)$, that is

$$\widehat{l}(\beta) = \sum_{i=1}^{n}(\tilde{Y}_i - \tilde{X}_i^\top\beta)^\top V_i^{-1}(\tilde{Y}_i - \tilde{X}_i^\top\beta),$$

where $\tilde{Y}_i = Y_i - \widehat{\mu}_Y(t_i)$, $\widehat{\mu}_Y(t_i) = (\widehat{\mu}_Y(t_{i1}), ..., \widehat{\mu}_Y(t_{in_i}))^\top$; $\tilde{X}_i$ and $\widehat{\mu}_X(t_i)$ similarly defined.

Ridge regression, designed specifically to handle correlated predictors, involves introducing a shrinkage penalty λ to the WLS, then the weighted ridge regression (WRR) estimator of β is the minimizer of the objective function

$$l^*(\beta) = \widehat{l}(\beta) + \lambda\beta^\top\beta,$$

the WRR estimator is then defined as

$$\begin{aligned}\widehat{\beta}_{MR} &= arg\,\min_{\beta}\,\{(\tilde{Y} - \tilde{X}\beta)^\top V^{-1}(\tilde{Y} - \tilde{X}\beta) + \lambda\beta^\top\beta\} \qquad (2.3)\\ &= (\tilde{X}^\top V^{-1}\tilde{X} + \lambda I)^{-1}\tilde{X}^\top V^{-1}\tilde{Y}\\ &= R(\lambda)\widehat{\beta}_{kernel},\end{aligned}$$

where $R(\lambda) = \Big((\tilde{X}^\top V^{-1}\tilde{X})^{-1}\lambda + I\Big)^{-1}$ and $\widehat{\beta}_{kernel}$ is the minimizer of $\widehat{l}(\beta)$ w.r.t β.

It is straightforward to show that the estimator of $g(t_{ij})$ is

$$\widehat{g}(t_{ij}) = \widehat{\mu}_Y(t_{ij}) - \widehat{\mu}_X^\top(t_{ij})\widehat{\beta}_{MR}.$$

Further, the best linear unbiased estimator (BLUE) of the random effects b_i may be proceeded by calculating their conditional expectations given Y_i while estimating β by $\widehat{\beta}_{MR}$. This gives

$$\widehat{b_i} = E[b_i|Y_i] = D_iZ_i^\top V_i^{-1}(\tilde{Y}_i - \tilde{X}_i^\top\widehat{\beta}_{MR}) = D_iZ_i^\top V_i^{-1}\big(Y_i - X_i^\top\widehat{\beta}_{MR} - \widehat{g}(t_i)\big).$$

Fan and Li (2001) suggested the GCV to select the tuning parameter λ, the GCV statistic is defined by

$$GCV(\lambda) = \frac{RSS(\lambda)/n}{(1 - tr\big(d(\lambda)\big)/n)^2},$$

where $RSS(\lambda) = (\tilde{Y} - \tilde{X}^\top\widehat{\beta}_{MR})^\top V^{-1}(\tilde{Y} - \tilde{X}^\top\widehat{\beta}_{MR})$ and $d(\lambda) = \tilde{X}(\tilde{X}^\top V^{-1}\tilde{X} + \lambda I)^{-1}\tilde{X}^\top V^{-1} + ZDZ^\top(I - \tilde{X}[(\tilde{X}^\top V^{-1}\tilde{X} + \lambda I)^{-1}\tilde{X}^\top V^{-1}])$. We select $\widehat{\lambda} = arg\,\min_{\lambda}$ $GCV(\lambda)$.

In the following, we establish the limiting distribution of the WRR estimator. The Theorem 2.1 concerning the kernel method is given in Lin and Carroll (2001). We quote it here to ease comparison with properties of the MR method given in Theorem 2.2.

Theorem 2.1. *Suppose that* $h \propto n^{-\alpha}$, $\frac{1}{5} \leqslant \alpha \leqslant \frac{1}{3}$ *and* $n \longrightarrow \infty$ *and define*

$$\tilde{X} = X = \lim_{n \longrightarrow \infty} \frac{\partial \widehat{g}(t;\beta)}{\partial \beta}. \tag{2.4}$$

Then $\widehat{\beta}_K$ *converges in distribution:*

$$\sqrt{n}\{\widehat{\beta}_{kernel} - \beta + h^2 b_{kernel}(\beta, g)/2\} \xrightarrow{D} N(0, V_{kernel})$$

where

$$\begin{aligned} b_{kernel}(\beta, g) &= E\{\tilde{X}^T V^{-1} \tilde{X}\}^{-1} E\{\tilde{X}^T V^{-1} g^{(2)}(t)\}, \quad (2.5) \\ V_{kernel} &= E\{\tilde{X}^T V^{-1} \tilde{X}\}^{-1} E\{(Z_1 - Z_2)^T \Sigma (Z_1 - Z_2)\} E\{\tilde{X}^T V^{-1} \tilde{X}\}^{-1}. \end{aligned}$$

Here $\tilde{X} = \big(X - \mu_X(t)\big)$, $\Sigma = var(Y|X,t)$, $Z_1 = V^{-1}\tilde{X}$, $Z_2 = (Z_2^1, \ldots, Z_2^m)$ *with*

$$Z_2^j = \frac{\{\sum_{k=1}^m \sum_{l=1}^m E(\tilde{X}^k V^{kl} | t^l = t^j)\} f_j(t^j)}{\sum_{l=1}^m f_l(t^j)}, \tag{2.6}$$

and V^{kl} *denotes the* (k, l) *entry of* V^{-1}.

Theorem 2.2. *Under the assumptions of Theorem 2.1, the mixed ridge estimator* $\widehat{\beta}_{MR}$ *converges in distribution:* $\sqrt{n}\{\widehat{\beta}_{MR} - R(\lambda)\big(\beta + h^2 b_{kernel}(\beta, g)/2\big)\} \xrightarrow{D} N\big(0, R(\lambda) V_{kernel} R^\top(\lambda)\big)$.

3. E-M Algorithm

Whereas, improving the efficiency of regression coefficients are related to estimation of the covariance function, the solution is found by an iterative algorithm between β, b_i, $g(t)$ and inverse of the estimated covariance matrix V. The proposed algorithm is an extention of E-M algorithm with a structured covariance matrix of random errors. We apply the Newton–Raphson method to minimize equations (2.3) and get the updating formula

$$\begin{aligned} \widehat{\beta}_{MR}^{[r+1]} &= \Big[\sum_{i=1}^n \tilde{X}_i^\top V_i^{-1[r]} \tilde{X}_i + \lambda I\Big]^{-1} \Big[\sum_{i=1}^n \tilde{X}_i^\top V_i^{-1[r]} \tilde{Y}_i\Big], \quad (3.7) \\ \widehat{g}^{[r+1]}(t_{ij}) &= \widehat{\mu}_Y(t_{ij}) - \widehat{\mu}_X^\top(t_{ij}) \widehat{\beta}_{MR}^{[r+1]}, \\ \widehat{b}_i^{[r+1]} &= D_i Z_i^\top V_i^{-1[r]} (\tilde{Y}_i - \tilde{X}_i^\top \widehat{\beta}_{MR}^{[r+1]}), \end{aligned}$$

Hence, we get the following iterative algorithm:

Step 0: Set $r = 0$. Let $\widehat{D}_i^{[r]} = I$, $\widehat{\sigma}^{2[r]}(t) = 1$ and $\widehat{\theta}^{[r]} = 1$.

Step 1: Set $r = r + 1$. Update $\widehat{\beta}^{[r+1]}$, $\widehat{b}_i^{[r+1]}$ and $\widehat{g}^{[r+1]}(t)$ by the computed formulas (3.7) based on current values of $\widehat{V}^{[r]} = diag(\widehat{V}_1^{[r]}, \ldots, \widehat{V}_n^{[r]})$ where $\widehat{V}_i^{[r]} = Z_i D_i^{[r]} Z_i^\top + \widehat{\Sigma}_i^{[r]}$, $\widehat{\Sigma}_i = \widehat{A}_i R_i(\widehat{\theta}) \widehat{A}_i$, $\widehat{A}_i = diag(\widehat{\sigma}(t_{i1}), \ldots, \widehat{\sigma}(t_{in_i}))$ and $R_i(\theta)$ is the correlation matrix of ε_i with (j, k) element equaling $\rho(t_{ij}, t_{ik}, \widehat{\theta})$.

Step 2: Update $\widehat{D}_i^{[r]}$, $\widehat{\sigma}^{2[r]}(t)$ and $\widehat{\theta}^{[r]}$.

At this step, the ML estimator of D_i would be

$$\widehat{D}_i^{[r]} = n_i^{-1} \widehat{b}_i^{[r]} \widehat{b}_i^{[r]\top}.$$

The kernel estimator of σ^2 is

$$\widehat{\sigma}^{2[r]} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \widehat{r}_{ij}^2 K_{h_1}(t - t_{ij})}{\sum_{i=1}^{n} \sum_{j=1}^{n_i} K_{h_1}(t - t_{ij})},$$

where $\widehat{r}_{ij} = Y_i(t_{ij}) - X_i^\top(t_{ij}) \widehat{\beta}^{[r]} - Z_i^\top(t_{ij}) \widehat{b}_i^{[r]} - \widehat{g}^{[r]}(t_{ij})$. To construct the estimator of θ, we adopt the quasi-likelihood approach. By the quasi-likelihood approach, the estimator of θ is defined by

$$\widehat{\theta} = \underset{\theta}{arg\ max} \Big(-\frac{1}{2} \sum_{i=1}^{n} \{ log|R_i(\theta)| + \widehat{b}_i^\top \widehat{A}_i^{-1} R_i^{-1}(\theta) \widehat{A}_i^{-1} \widehat{b}_i \} \Big).$$

Step 3: Repeat Steps 1 and 2 until the selected covariates converge to a stable value.

4. Illustrations

4.1 Simulation studies

A simple simulation study is conducted to characterize the relative performances of mixed ridge regression and the usual mixed-effects modeling approach in the context of multiple, correlated predictors. To achieve different degrees of collinearity, the explanatory variables are generated using the following formula

$$x_{ik} = (1 - \gamma^2)^{\frac{1}{2}} \omega_{ik} + \gamma^2 \omega_{ip}, \quad i = 1, \ldots, n = 50, \qquad k = 1, \ldots, p = 5, \tag{4.8}$$

where ω_{ik} are independent and generated from a standard normal distribution, γ is specified so that the correlation between any two explanatory variables is given by γ^2. Four different set of correlations corresponding to $\gamma = 0.70, 0.80, 0.90$ and 0.99 are considered. For the nonparametric component in (1.1), we selecte a smooth function of the form $g(t_{ij}) = 2sin(2\pi t_{ij})$. We further assume $n_i = 4$ measurements for each subject i and simulate data from the following semiparametric mixed model

$$\begin{aligned} Y_i(t_{ij}) = \quad & x_{1,i}(t_{ij})\beta_1 + x_{2,i}(t_{ij})\beta_2 + x_{3,i}(t_{ij})\beta_3 + x_{4,i}(t_{ij})\beta_4 + x_{5,i}(t_{ij})\beta_5 \\ & + g(t_{ij}) + b_i + \varepsilon_i(t_{ij}), \end{aligned} \tag{4.9}$$

where b_i independently follow normal distribution $N(0, \sigma_b^2)$, where $\sigma_b^2 = 0.25$. The true value of β was taken to be $\beta^T = (0.5, 1, 1.5, 2, 0.1, 0.2)$. The time t_{ij} is generated from the $U(0, 1)$ distribution. The random error process $\varepsilon_i(t)$ is taken to be a Gaussian process with mean 0, variance function $\sigma^2(t) = exp(\frac{t}{12})$ and $AR(1)$ correlation structure $corr(\varepsilon_i(s), \varepsilon_i(t)) = \rho^{|t-s|}$ for $s \neq t$, where $\rho = 0.9$ to capture strong correlation errors.

We use the Epanechnikov kernel function $K(u) = 0.75(1-u^2)_+$. To investigate the impact on the performance of two methods under a misspecified correlation structure, we compare the performance of $\widehat{\beta}_{kernel}$ and $\widehat{\beta}_{MR}$ using the exchangable working correlation structure ($kernel - I$) or ($MR - I$) when the true correlation structure is $AR(1)$ ($kernel - C$) or ($MR - C$). For each of the estimators $\widehat{\beta}_{kernel}$ and $\widehat{\beta}_{MR}$, its estimation accuracy is measured by the mean squared error (MSE) defined by

$$MSE = (\widehat{\beta} - \beta)^T (\widehat{\beta} - \beta).$$

The simulation results for MSE of the proposed methods are presented in Table 2. Table 1 reports the empirical biases and standard deviations (SDs) of the estimated β from the kernel and backfitting methods. We can take the following observations:

(i) Mixed ridge method outperforms mixed method in terms of the MSE criterion, even for the cases of misspecified correlation structure.

(ii) Mixed ridge estimator has a smaller bias and SD in all cases than the mixed estimator. Hence, our proposed method becomes more stable and efficient.

(iii) Furthermore, as the correlation among predictors increases, the performance of mixed modeling decreases more rapidly than the mixed ridge model in terms of the biases, SDs and MSEs criterion. This result is more dramatic at extreme levels of correlation ($\gamma = 0.99$).

Table 1: Estimated regression coefficients for important variates, bias(SD) based on 500 replications.

Methods	parameters	$\gamma = 0.70$	$\gamma = 0.80$	$\gamma = 0.90$	$\gamma = 0.99$
MR-C	β_1	0.0316(0.0400)	0.0376(0.0475)	0.0514(0.0649)	0.1401(0.1748)
	β_2	0.0310(0.0390)	0.0368(0.0464)	0.0506(0.0635)	0.1446(0.1706)
	β_3	0.0282(0.0355)	0.0333(0.0422)	0.0452(0.0577)	0.1359(0.1572)
	β_4	0.0306(0.0388)	0.0364(0.0461)	0.0498(0.0630)	0.1397(0.1745)
	β_5	0.0351(0.0454)	0.0454(0.0592)	0.0690(0.0910)	0.2208(0.2644)
MR-I	β_1	0.0337(0.0414)	0.0402(0.0492)	0.0553(0.0677)	0.1686(0.2059)
	β_2	0.0330(0.0408)	0.0393(0.0484)	0.0540(0.0665)	0.1624(0.2011)
	β_3	0.0295(0.0375)	0.0351(0.0446)	0.0482(0.0613)	0.1453(0.1847)
	β_4	0.0362(0.0452)	0.0431(0.0537)	0.0593(0.0739)	0.1800(0.2236)
	β_5	0.0342(0.0427)	0.0444(0.0556)	0.0687(0.0865)	0.2518(0.3219)
kernel-C	β_1	0.0315(0.0400)	0.0375(0.0476)	0.0517(0.0656)	0.1596(0.2027)
	β_2	0.0313(0.0390)	0.0373(0.0464)	0.0513(0.0639)	0.1585(0.1975)
	β_3	0.0296(0.0358)	0.0352(0.0426)	0.0485(0.0586)	0.1499(0.1811)
	β_4	0.0308(0.0389)	0.0366(0.0463)	0.0504(0.0638)	0.1558(0.1971)
	β_5	0.0356(0.0456)	0.0463(0.0596)	0.0720(0.0930)	0.2782(0.3608)
kernel-I	β_1	0.0337(0.0414)	0.0402(0.0493)	0.0553(0.0679)	0.1708(0.2097)
	β_2	0.0332(0.0407)	0.0395(0.0485)	0.0543(0.0667)	0.1679(0.2062)
	β_3	0.0296(0.0376)	0.0353(0.0447)	0.0485(0.0615)	0.1500(0.1902)
	β_4	0.0362(0.0452)	0.0431(0.0538)	0.0594(0.0740)	0.1835(0.2288)
	β_5	0.0345(0.0428)	0.0444(0.0557)	0.0688(0.0868)	0.2633(0.3380)

4.2 CD4 data analysis

We now apply the semiparametric mixed model to the longitudinal CD4 cell count data through the proposed methods. In this data set, there are a total of 2,376 CD4 measurements from 369 subjects available. The first objective of this analysis is to characterize the population average time course of CD4 decay, while accounting for the following additional predictor variables: smoking (packs per day); recreational drug use (yes or no); numbers of dangerous relationships; and depression symptoms as measured by the CESD scale (larger values indicate increased depressive symptoms). The analysis was conducted on square-root transformed CD4 numbers whose distribution is more nearly Gaussian. We refer to Zeger and Diggle (1994) and Wang et al (2005) for more detailed descriptions of the data.

Since there seems to exist a positive correlation among responses from the same patient, we need to incorporate a correlation structure into the estimation scheme.

Table 2: Mean square error based on 500 replications.

Methods	$\gamma = 0.70$	$\gamma = 0.80$	$\gamma = 0.90$	$\gamma = 0.99$
MR-C	0.0079	0.0118	0.0238	0.2103
MR-I	0.0087	0.0128	0.0258	0.2704
kernel-C	0.0081	0.0121	0.0249	0.2871
kernel-I	0.0087	0.0129	0.0261	0.2919

Following Zeger and Diggle (1994), it is found that the compound symmetry covariance matrix fitted the data reasonably well. The estimated correlation is $\rho = 0.509$.

To facilitate choosing the bandwidth h, we adopt the approach of Fan and Li (2004). $\widehat{\sigma}^2(t)$ is a one dimensional kernel regression of the squared residuals over time. we estimate $h = 0.02$ and $h_1 = 0.015$. We use bootstrap resampling to get the SD of estimatrs.

The estimates of the parameters in the model are presented in Table 3. Based on the results, we see that the mixed estimator has a larger SD than the mixed ridge estimator. Age plays little role. The effects of smoking and depression are found to be significant. In contrast, the other three effects are insignificant at level 0.05, which are similar to Wang et al (2005) except that the effect of the dangerous relationships is significant in their analysis.

Table 3: Regression coefficient estimates(SD) in the analysis of the CD4 data.

	methods	
	Mixed Ridge	Mixed
Age	0.0208(0.0021)	0.0212(0.0021)
Smoking	0.5808(0.0080)	0.5941(0.0082)
Drug	0.4582(0.0166)	0.5350(0.0194)
dangerous relationships	0.0597(0.0024)	0.0585(0.0024)
Depression	-0.0531(0.0009)	-0.0532(0.0009)

Conclusions

In this paper, we consider a more flexible model called semiparametric mixed-effects model and described an extension and application of ridge regression for longitudinal data with a multicollinearity problem. We proposed a new estimator to use when data sets may have correlated variables and within-subject effects. The proposed ridge estimator combines the flexibility of nonparametric assumptions, the usefulness of hierarchical linear models and ridge methodology. To improve efficiency for regression coefficients, the estimation of the covariance function is integrated with the iterative algorithm. As expected, the mixed ridge estimation results in coefficients with smaller biases, variances and MSE than the mixed model even for the cases of the misspecified correlation structure. Furthermore, as the correlation among predictors increases, the performance of mixed modeling decreases more rapidly than the mixed ridge model. There were the results obtained from numerical studies.

References

Eliot M., Ferguson J., Reilly M. P. and Foulkes A. S. (2011), Ridge regression for longitudinal biomarker data, *The International Journal of Biostatistics,* **7**, Article 37.

Fan J. Q., Huang T. and Li R. Z. (2007), Analysis of longitudinal data with semiparametric estimation of covariance function, *Journal of the American Statistical Association,* **102**, 632-641.

Fan J. Q. and Li R. Z. (2004), New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis, *Journal of the American Statistical Association,* **99**, 710-723.

Fan J.Q. and Li R. Z. (2001), Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association,* **96**, 1348-1360.

Fitzmaurice G., Laird N. and Ware J. (2004), *Applied Longitudinal Analysis,* Wiley Series in Probability and Statistics.

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

Hoerl A. and Kennard R. (1970b), Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics,* **12**, 55-67.

Hu Z. H., Wang N. Y. and Carroll R.J. Profile-kernel versus backfitting in the partially linear models for longitudinal/ clustered data, *Biomeitrka,* **91**, 251-262.

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

Lin X. and Carroll, R. J. (2001), Semiparametric regression for clustered data using generalised estimating equations, *Journal of the American Statistical Association,* **96**, 1045-56.

Lin D.Y. and Ying Z. (2001), Semiparametric and nonparametric regression analysis of longitudinal data, *Journal of the American Statistical Association,* **96**, 103-113.

Malo N., Libiger O. and Schork N. (2008), Accommodating linkage disequilibrium in genetic-association analyses via ridge regression, *The American Journal of Human Genetics,* **82**, 375-85.

Tikhonov A. (1943), On the stability of inverse problems, *Proceedings of the USSR Academy of Sciences,* **39**, 267-88.

Wang N., Carroll R. and Lin X. H. (2005), Efficient semiparametric marginal estimation for longitudinal/clustered data, *Journal of the American Statistical Association,* **100**, 147-157.

Zeger S. L. and Diggle P. J. (1994), Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, *Biometrics,* **50**, 689–99.

Zhang B. and Horvath S. (2006), Ridge regression based hybrid genetic algorithms for multi-locus quantitative trait mapping, *Bioinformatics Research and Applications,* **1**, 261-72.