The International Journal of Biostatistics

Volume 7, Issue 1

2011

Article 37

Ridge Regression for Longitudinal Biomarker Data

Melissa Eliot, *University of Massachusetts Amherst*

Jane Ferguson, *University of Pennsylvania*

Muredach P. Reilly, *University of Pennsylvania*

Andrea S. Foulkes, *University of Massachusetts Amherst*

Recommended Citation:

Eliot, Melissa; Ferguson, Jane; Reilly, Muredach P.; and Foulkes, Andrea S. (2011) "Ridge Regression for Longitudinal Biomarker Data," *The International Journal of Biostatistics*: Vol. 7: Iss. 1, Article 37.

DOI: 10.2202/1557-4679.1353

Ridge Regression for Longitudinal Biomarker Data

Melissa Eliot, Jane Ferguson, Muredach P. Reilly, and Andrea S. Foulkes

Abstract

Technological advances facilitating the acquisition of large arrays of biomarker data have led to new opportunities to understand and characterize disease progression over time. This creates an analytical challenge, however, due to the large numbers of potentially informative markers, the high degrees of correlation among them, and the time-dependent trajectories of association. We propose a mixed ridge estimator, which integrates ridge regression into the mixed effects modeling framework in order 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. An expectation-maximization algorithm is described to account for unknown variance and covariance parameters. Model performance is demonstrated through a simulation study and an application of the mixed ridge approach to data arising from a study of cardiometabolic biomarker responses to evoked inflammation induced by experimental low-dose endotoxemia.

KEYWORDS: biomarkers, cardiovascular disease (CVD), mixed effects, repeated measures, ridge regression

Author Notes: Support for this research was provided by an NIH/NHLBI individual research award to A.S.F. (R01 HL107196) and an NIH/NHLBI SCCOR award to M.P.R. (P50 HL083799).

1 Introduction

Ridge regression (RR), also known as Tikhonov regularization (Tikhonov, 1943), is a well-described penalized regression approach to handling multiple co-linear 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 (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 data arising from genetic association studies include, for example, Zhang and Horvath (2006) and Malo, Libiger, and Schork (2008).

In this manuscript 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. The data motivating our research arise from the Genetics of Evoked-Response to Niacin and Endotoxemia (the GENE) study and are described in more detail in Section 3.2 below. In the case of uncorrelated or minimally-correlated predictor variables, mixed effects models with person-specific random intercept and slopes terms can be applied to account for the within person correlation in making inferences about association (Fitzmaurice, Laird, and Ware, 2004). Here we integrate RR and the mixed effects modeling framework in order to account both for the correlation induced by repeatedly measuring the outcome on each individual over time, as well the potential high degree of correlation among potential predictor variables.

We begin in Section 2.1 by providing a brief background on RR and mixed effects modeling while defining the notation used throughout this manuscript. We then introduce a new estimator, which we term the mixed ridge (MR) estimator, for longitudinal data, as well as an associated testing framework in Section 2.2. Results of a simulation study to evaluate performance and an application to data arising from the GENE study are then provided in Sections 3.1 and 3.2, respectively. Finally, a discussion of our findings and alternative approaches is provided in Section 4.

2 Methods

2.1 Background and Notation

2.1.1 Ridge regression

We begin with the simple linear regression model given by:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \quad (1)$$

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ is an $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times p$ design matrix comprised of p columns representing each of the potential predictor variables, n is the number of individuals in our sample and $\varepsilon \sim MVN(0, \sigma^2 \mathbf{I})$ is an $n \times 1$ vector of independent errors. It is straightforward to show that the least squares and maximum likelihood solutions to this equation are given by $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ and $Var(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2$ (Christensen, 2002). Notably, in the case that the columns of $\mathbf{X}$ are highly correlated, $\mathbf{X}^T \mathbf{X}$ will be singular and we replace $(\mathbf{X}^T \mathbf{X})^{-1}$ with $(\mathbf{X}^T \mathbf{X})^-$ where ‘ $-$ ’ denotes generalized inverse, and a unique solution to Equation 1 does not exist. Further, in the case of high correlation where $\mathbf{X}^T \mathbf{X}$ 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 λ to the least squares equation, and subsequently solving for the value of $\hat{\beta}$ such that:

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

It has been shown that the solution to Equation 2 is given by $\hat{\beta}_R = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{Y}$ and we have $Var(\hat{\beta}_R) = [(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T] [(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T]^T$ (Hoerl and Kennard, 1970a,b). 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 given by $EDF = tr \left([\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}]^{-1} \mathbf{X}^T \mathbf{X} \right)$ (Malo et al., 2008). Applying the generalized cross validation (GCV) method of Craven and Wahba (1979), we estimate λ of Equation 2 by:

$$\hat{\lambda} = \arg \min_{\lambda} \left\{ n^{-1} (\mathbf{Y} - \hat{\mathbf{Y}})^T (\mathbf{Y} - \hat{\mathbf{Y}}) / (1 - tr(\mathbf{S})/n)^2 \right\} \quad (3)$$

where $\mathbf{S} = \mathbf{X}(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T$ and $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}_R$. The optimal value of $\hat{\lambda}$ is found through an iterative procedure that involves recursively updating $\hat{\mathbf{Y}}$ and $\mathbf{S}$ based on

the current value of $\hat{\lambda}$. Notably, as described in Gentle, Hardle, and Mori (2004) for $tr(\mathbf{S})/\mathbf{n}$ small, $\hat{\lambda}$ of Equation 3 is approximated by:

$$\hat{\lambda} \approx \arg \min_{\lambda} \left\{ n^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}}) + 2n^{-2}tr(\mathbf{S})(\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}}) \right\} \quad (4)$$

2.1.2 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:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{z}_i^T b_i + \boldsymbol{\varepsilon}_i \quad (5)$$

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 $b_i \sim MVN(0, \mathbf{D})$ are person-specific random effects, $\mathbf{z}_i$ is the corresponding random effects design matrix, and $\boldsymbol{\varepsilon}_i \sim MVN(0, \sigma^2 \mathbf{I}_{n_i \times n_i})$ are independent random errors. Finally, we let $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{Z}$ be appropriately defined matrices representing the concatenation of the corresponding variables over all individuals i .

Notably, the log-likelihood function of $\mathbf{Y}$ based on this model is given by (Fitzmaurice et al., 2004):

$$l(\mathbf{Y}) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log(V_i) - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad (6)$$

where $\mathbf{V} = Var(\mathbf{Y}) = \mathbf{ZDZ}^T + \sigma^2 \mathbf{I}$ and V_i is component corresponding to individual i . Maximizing this function with respect to the fixed effects parameter vector, $\boldsymbol{\beta}$, in the non-penalized setting is equivalent to minimizing the least squares objective function:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \{ (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \} \quad (7)$$

It is straightforward to show that the resultant estimate of $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} \quad (8)$$

Further, the best unbiased predictor (BLUP) of the random effects is given by $\hat{b} = \mathbf{DZ}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, $var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$ and $var(\hat{b}) = \mathbf{DZ}^T [\mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}] \mathbf{ZD}$ (Laird and Ware, 1982).

2.2 Mixed Ridge (MR) Regression

In this section we introduce a penalized regression approach to estimation for the mixed model given in Equation 5. To begin, we assume the variance parameters $\theta = (\sigma, D)$ are known and add a penalization term to Equation 7, similar to the ridge regression setting, which yields:

$$\hat{\beta}_{MR} = \arg \min_{\beta} \{ (\mathbf{Y} - \mathbf{X}\beta)^T V^{-1} (\mathbf{Y} - \mathbf{X}\beta) + \lambda \beta^T \beta \} \quad (9)$$

Differentiating the objective function in Equation 9, setting the resultant equal to 0, and solving for $\hat{\beta}$, we have:

$$\hat{\beta}_{MR} = (\mathbf{X}^T V^{-1} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T V^{-1} \mathbf{Y} \quad (10)$$

A simple proof of this result is provided in Appendix A.1. Additionally, it can be shown that $Var(\hat{\beta}_{MR}) = (\mathbf{X}^T V^{-1} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T V^{-1} \mathbf{X} (\mathbf{X}^T V^{-1} \mathbf{X} + \lambda \mathbf{I})^{-1}$ and the empirical Bayes estimate of b is given by:

$$\begin{aligned} \hat{b} &= E[b|\mathbf{Y}] = D\mathbf{Z}^T V^{-1} (\mathbf{Y} - \mathbf{X}\hat{\beta}_{MR}) \\ &= D\mathbf{Z}^T V^{-1} (\mathbf{I} - \mathbf{X}[(\mathbf{X}^T V^{-1} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T V^{-1}]) \mathbf{Y} \end{aligned} \quad (11)$$

Estimation of λ is achieved using the approach described in Section 2.1.1, where now we have $\mathbf{S} = \mathbf{X}(\mathbf{X}^T V^{-1} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T V^{-1} + \mathbf{Z}D\mathbf{Z}^T (\mathbf{I} - \mathbf{X}[(\mathbf{X}^T V^{-1} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T V^{-1}])$.

More generally, consider the setting in which the variance parameters θ are unknown. We propose an extension of the expectation-maximization (EM) algorithm described by Laird and Ware (1982), that includes an additional step for estimation of the ridge component. This approach is summarized by the following step-by-step procedure:

1. **(E-step)** Initialize $\hat{\theta}^{(t)} = \theta_0$ and $\hat{\lambda}^{(t)} = \lambda_0$. Solve for $\hat{\beta}_{MR}^{(t)}$ and the sufficient statistics $\hat{t}_1^{(t)}$ and $\hat{t}_2^{(t)}$ given by:

$$\hat{t}_1^{(t)} = E \left(\sum_{i=1}^n \varepsilon_i^T \varepsilon_i | \mathbf{Y}_i, \hat{\beta}_{MR}^{(t)}, \hat{\theta}^{(t)} \right) \quad (12)$$

$$\hat{t}_2^{(t)} = E \left(\sum_{i=1}^n b_i b_i^T | \mathbf{Y}_i, \hat{\beta}_{MR}^{(t)}, \hat{\theta}^{(t)} \right) \quad (13)$$

2. **(M-Step)** 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 $\hat{V}^{(t+1)} = \mathbf{Z}\hat{D}^{(t+1)}\mathbf{Z}^T + \hat{\sigma}^{2(t+1)}\mathbf{I}$
3. Update $\hat{\lambda}^{(t+1)}$ and let

$$\hat{\beta}_{\text{MR}}^{(t+1)} = \left(\mathbf{X}^T \hat{V}^{-1(t+1)} \mathbf{X} + \hat{\lambda}^{(t+1)} \mathbf{I} \right)^{-1} \mathbf{X}^T \hat{V}^{-1(t+1)} \mathbf{Y} \quad (14)$$

4. Repeat Steps (a) – (c) a large number of times and until a convergence criterion is met.

In order to determine the significance of each of the predictor variables, we begin by calculating Wald test statistics. These are denoted Z_k for each $k = 1, \dots, p$, and given by the corresponding value of $\hat{\beta}_{\text{MR}}$ divided by the square root of its variance. Since an EM algorithm is employed to estimate β , we use Louis' formula (Louis, 1982) to determine the corresponding variance. Finally, Westfall and Young's free step-down resampling approach (Westfall and Young, 1993, Foulkes, 2009), also known as the maxT procedure, is applied to adjust for multiple testing.

3 Examples

3.1 Simulation Study

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. For simplicity of presentation, the simulation study assumes repeatedly measured outcomes, while the predictor variables are measured at a single, baseline time point. We further assume $n_i = 4$ measurements for each subject i and generate data according to the model of Equation 5 where $\beta = (0, 0, 0.2, 0.4, 0.6)$, $b_i \sim N(0, 0.6)$, 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 ρ in Table 1, is assumed to take on values between 0 and 0.99. Starting values for the variance components are derived based on fitting a mixed model with no ridge component. In total, $M = 500$ simulations are conducted for each condition based on sample sizes of $n = 500$ individuals.

Table 1: Results of simulation study

ρ	β	Mixed		Mixed Ridge	
		Estimate (stan dev)	Power	Estimate (stan dev)	Power
0.0	0.0	0.002 (0.031)		0.005 (0.023)	
	0.0	-0.001 (0.031)		0.002 (0.023)	
	0.2	0.196 (0.031)	1.0	0.197 (0.023)	1.0
	0.4	0.399 (0.031)	1.0	0.397 (0.023)	1.0
	0.6	0.602 (0.031)	1.0	0.597 (0.023)	1.0
0.2	0.0	-0.004 (0.031)		0.003 (0.023)	
	0.0	-0.004 (0.031)		0.002 (0.023)	
	0.2	0.197 (0.034)	1.0	0.203 (0.025)	1.0
	0.4	0.402 (0.034)	1.0	0.398 (0.025)	1.0
	0.6	0.601 (0.035)	1.0	0.595 (0.025)	1.0
0.4	0.0	-0.004 (0.030)		-0.002 (0.022)	
	0.0	0.004 (0.030)		0.006 (0.022)	
	0.2	0.199 (0.039)	1.0	0.200 (0.029)	1.0
	0.4	0.402 (0.039)	1.0	0.400 (0.029)	1.0
	0.6	0.602 (0.039)	1.0	0.595 (0.029)	1.0
0.6	0.0	0.004 (0.029)		0.005 (0.022)	
	0.0	-0.001 (0.030)		-0.001 (0.022)	
	0.2	0.191 (0.047)	0.98	0.206 (0.031)	1.0
	0.4	0.401 (0.047)	1.0	0.398 (0.031)	1.0
	0.6	0.603 (0.046)	1.0	0.592 (0.031)	1.0
0.8	0.0	0.003 (0.029)		0.005 (0.022)	
	0.0	-0.002 (0.029)		0.000 (0.022)	
	0.2	0.198 (0.065)	0.85	0.211 (0.047)	0.90
	0.4	0.394 (0.065)	1.0	0.393 (0.046)	1.0
	0.6	0.607 (0.064)	1.0	0.591 (0.046)	1.0
0.9	0.0	0.003 (0.029)		0.005 (0.021)	
	0.0	-0.002 (0.029)		0.000 (0.021)	
	0.2	0.204 (0.091)	0.52	0.227 (0.062)	0.74
	0.4	0.399 (0.091)	0.99	0.399 (0.062)	1.0
	0.6	0.596 (0.091)	1.0	0.568 (0.062)	1.0
0.99	0.0	0.003 (0.029)		0.005 (0.021)	
	0.0	-0.002 (0.029)		0.000 (0.021)	
	0.2	0.210 (0.288)	0.16	0.254 (0.143)	0.61
	0.4	0.402 (0.288)	0.28	0.404 (0.143)	0.73
	0.6	0.587 (0.286)	0.51	0.536 (0.142)	0.90

Figure 1: Power for detecting associaton

Simulation results are presented in Table 1 and Figure 1. As expected, MR estimates tend to exhibit slightly more bias ($|\text{Estimate} - \beta|$) than mixed estimates; however, the MR estimates have smaller variances, resulting ultimately in higher power of the MR model. Power for detecting association when $\beta = 0.20$, $\sigma = 1$ and $D = 0.6$, is illustrated in Figure 1. Here we see that power begins to diverge between the two models at approximately $\rho = 0.8$. This difference is more dramatic at extreme levels of correlation ($\rho = 0.99$), where power for the mixed model when the true value of β is 0.20 is 0.160 compared to a power of 0.610 for the MR model; when true value of β is 0.60 the power for the mixed model is 0.51 compared with 0.90 for the MR model. The average type 1 error rates across all assumed values of ρ and both β_1 and β_2 are 0.050 and 0.054 for the mixed and MR models, respectively.

Table 2: Observed correlations among potential predictor variables

	TNF	Apo-A	Apo-B	Chol	HDL	Trig	Phos	LDL
TNF	1.00	0.13	0.04	0.04	0.07	-0.05	0.02	0.03
Apo-A	0.13	1.00	0.07	0.40	0.87	0.01	0.66	0.01
Apo-B	0.04	0.07	1.00	0.88	-0.12	0.40	0.56	0.95
Chol	0.04	0.40	0.88	1.00	0.29	0.31	0.79	0.90
HDL	0.07	0.87	-0.12	0.29	1.00	-0.32	0.46	-0.11
Trig	-0.05	0.01	0.40	0.31	-0.32	1.00	0.42	0.27
Phos	0.02	0.66	0.56	0.79	0.46	0.42	1.00	0.55
LDL	0.03	0.01	0.95	0.90	-0.11	0.27	0.55	1.00

3.2 Data Example: The GENE study

The Genetics of Evoked-Responses to Niacin and Endotoxemia (the GENE) study is an ongoing trial designed to characterize the effects of genetic factors on the response to niacin therapy and endotoxin. Healthy volunteers ($n = 189$) were administered low-dose endotoxin (LPS) infusion, which produces a mild-inflammatory response over a 6-8 hour period. Vital signs, including systolic blood pressure, as well as apolipoprotein A1 (Apo-A1), apolipoprotein B (Apo-B), total cholesterol (Chol), high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, phospholipids (Phos), triglycerides (Trig) and tumor necrosis factor-alpha (TNF) were measured repeatedly over a 24-hour period.

Systolic blood pressure at 0, 6, 12, and 24 hours after the evoked inflammation is modeled using all other variables listed above as potential predictor variables. Pearson correlations between all pairs of potential predictor variables range in absolute value from 0.01 to 0.95, as described in Table 2. The high degree of correlation indicates that MR regression is appropriate for this setting. As the change in systolic blood pressure over time is expected to be piecewise-linear, a linear spline (Fitzmaurice et al., 2004) model is applied with change points at times 6 and 12, denoted Time_1 and Time_2 respectively. Random within-subject effects for intercept and slope are also included in the model. Each subject is measured between 2 and 4 times.

The coefficient estimates, standard deviations and corresponding observed test statistics (corresponding to 2-sided tests that the coefficient is equal to 0) for the MR and mixed modeling approaches are given in Table 3. In both modeling frameworks, the overall effect of time on systolic blood pressure is significantly different than 0. After adjusting for multiple testing, the MR approach additionally identifies Apo-B as significantly associated with systolic blood pressure over time, while the

Table 3: Results from the GENE study comparing MR to mixed model

Variable name	Mixed		Mixed Ridge	
	Estimate (sd)	T-stat (adj p-val)	Estimate (sd)	T-stat (adj p-val)
Time	0.46 (0.16)	2.90 (0.043)	0.47 (0.12)	3.99 (0.006)
Time ₁	-0.40 (0.26)	-1.55 (0.387)	-0.40 (0.20)	-2.00 (0.366)
Time ₂	-0.17 (0.18)	-0.93 (0.649)	-0.18 (0.15)	-1.15 (0.696)
Apo-A1	0.08 (0.05)	1.79 (0.271)	0.07 (0.04)	1.96 (0.366)
Apo-B	-0.18 (0.09)	-1.94 (0.271)	-0.24 (0.07)	-3.59 (0.018)
Chol	0.49 (1.20)	0.41 (0.769)	1.16 (0.94)	1.23 (0.696)
HDL	-0.55 (1.20)	-0.46 (0.769)	-1.16 (0.94)	-1.23 (0.696)
LDL	-0.35 (1.20)	-0.29 (0.769)	-0.98 (0.94)	-1.04 (0.696)
Phos	-0.05 (0.03)	-1.35 (0.336)	-0.06 (0.02)	-2.67 (0.104)
Trig	-0.08 (0.24)	-0.34 (0.769)	-0.21 (0.19)	-1.09 (0.696)
TNF	0.15 (0.20)	0.75 (0.712)	0.11 (0.13)	0.84 (0.696)

usual mixed modeling approach is unable to detect this association. This finding is consistent with known associations between Apo-B and blood pressure [see, for example, Genest and Cohn (1995)]. Apo-B can be suppressed by inflammation in humans while blood pressure tends to rise with inflammation stress, supporting the observed inverse correlation. This is likely an indirect association with inflammation driving both changes in Apo-B and blood pressure.

4 Discussion

In this manuscript, we describe an extension and application of ridge regression for longitudinally measured biomarkers. Alternative penalized regression approaches include the least absolute shrinkage and selection operator (LASSO) and the elastic net (Tibshirani, 1996, Zou and Hastie, 2005). LASSO extends ridge regression by introducing a mixture distribution on the coefficient effects with a point mass at zero, while the elastic net is a further generalization that incorporates two shrinkage parameters. Further extensions of the approach described herein is feasible in these alternative settings; however, estimation is less straightforward given the additional parameter constraints that would need to be applied.

We have proposed a new estimator to use when data sets may have correlated variables and within-subject effects. The mixed-effects ridge estimator combines the usefulness of hierarchical linear models and ridge regression. To account

for unobserved variance parameters, we have proposed an EM algorithm. As expected, the mixed-ridge estimation results in coefficients with smaller variances than the mixed model. Furthermore, as correlation among predictors increases, the power for mixed modeling decreases more rapidly than the power for our MR model. In summary, the mixed ridge model appears to perform better than the usual mixed model in terms of power for detecting association in the context of highly correlated predictors.

Appendix (A.1): Differentiating the objective function of Equation 9 and setting this equal to 0, we have:

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

or equivalently, $\beta^T (\mathbf{X}^T V^{-1} \mathbf{X} + \lambda \mathbf{I}_{p \times p}) = \mathbf{Y}^T V^{-1} \mathbf{X}$. Solving for β results in Equation 10.

References

- Christensen, R. (2002): *Plane Answers to Complex Questions*, Springer.
- Craven, P. and G. Wahba (1979): “Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation,” *Numerische Mathematik*, 31, 377–403.
- Fitzmaurice, G., N. Laird, and J. Ware (2004): *Applied Longitudinal Analysis*, Wiley Series in Probability and Statistics.
- Foulkes, A. (2009): *Applied Statistical Genetics with R: for population-based association studies*, Springer, NY.
- Genest, J. and J. Cohn (1995): “Clustering of cardiovascular risk factors: targeting high-risk individuals,” *American Journal of Cardiology*, 76, 8A–20A.
- Gentle, J., W. Hardle, and Y. Mori (2004): *Handbook of Computational Statistics*, Springer-Verlag Berlin Heidelberg.
- Hoerl, A. and R. Kennard (1970a): “Ridge regression: Applications to nonorthogonal problems,” *Technometrics*, 12, 69–82.
- Hoerl, A. and R. Kennard (1970b): “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, 12, 55–67.

- Laird, N. and J. Ware (1982): “Random-effects models for longitudinal data,” *Biometrics*, 38, 963–974.
- Louis, T. A. (1982): “Finding the observed information matrix when using the EM algorithm,” *Journal of the Royal Statistical Society, Series B: Methodological*, 44, 226–233.
- Malo, N., O. Libiger, and N. Schork (2008): “Accommodating linkage disequilibrium in genetic-association analyses via ridge regression,” *The American Journal of Human Genetics*, 82, 375–85.
- Tibshirani (1996): “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–88.
- Tikhonov, A. (1943): “On the stability of inverse problems,” *Proceedings of the USSR Academy of Sciences*, 39, 267–88.
- Westfall, P. and S. Young (1993): *Resampling-based multiple testing*, Wiley-Interscience.
- Zhang, B. and S. Horvath (2006): “Ridge regression based hybrid genetic algorithms for multi-locus quantitative trait mapping,” *Bioinformatics Research and Applications*, 1, 261–72.
- Zou, H. and T. Hastie (2005): “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 67, 301–320.