SUPPLEMENTARY MATERIAL FOR

Group least squares regression for linear models with strongly correlated predictor variables

Min Tsao

Abstract This document contains supplementary material for AISM paper "Group least squares regression for linear models with strongly correlated predictor variables". It consists of (*i*) a fuller version of the remarks in Section 2.3 of the paper on group effects that are meaningful and can be accurately estimated, (*ii*) a simulation study on the group approach to the least squares regression, and (*iii*) a more detailed analysis of the Hald cement data. The simulation study examines the performance of the group approach in estimation, variable selection and prediction, and compares this approach with the traditional non-group based variable selection and ridge regression. For simplicity of presentation, we use a small model (3) for this simulation study. We emphasize that the good performance and advantages of the group approach we observe in this simulation study do not depend on the examples used; similar results can be obtained when the group approach is applied to any linear model containing one or more groups of strongly correlated variables.

Notation Throughout this document, equation numbers with a "†" sign refer to equations in the paper "Group least squares regression for linear models with strongly correlated predictor variables". Equation numbers without this sign, such as the (3) in the abstract, refer to equations in this document.

Keywords Strongly correlated predictor variables · Multicollinearity · Group effects · Linear models · Least squares regression.

Supported by the Natural Sciences and Engineering Research Council of Canada.

Min Tsao
Department of Mathematics & Statistics
University of Victoria
Victoria, British Columbia
Canada V8W 2Y2
E-mail: mtsao@uvic.ca

1 Effects that are meaningful and can be accurately estimated

Remarks [a] and [b] in Section 2.3 of the paper discussed group effects that are meaningful and can be accurately estimated. Here, we expand on these remarks and also add a discussion on general linear combinations of parameters of strongly correlated variables that can be accurately estimated.

Remark [a] For the q strongly correlated variables in APC arrangement in $\mathbf{X}_1'$ of the standardized model $(8)^\dagger$, let $\xi'(\mathbf{w}) = w_1\beta_1' + w_2\beta_2' + \cdots + w_q\beta_q'$ be a group effect. Its minimum-variance unbiased linear estimator is

$$\hat{\xi}'(\mathbf{w}) = w_1\hat{\beta}_1' + w_2\hat{\beta}_2' + \cdots + w_q\hat{\beta}_q'.$$

When the level of multicollinearity is high (r_M close to 1), by Theorem 1 in the paper, the eigen-effect ξ_E in $(12)^\dagger$ is accurately estimated with $var(\hat{\xi}_E) \approx \sigma^2/q$. The corresponding normalized eigen-effect ξ_E^* is also accurately estimated with $var(\hat{\xi}_E^*) \approx \sigma^2/q^2$, and Theorem 2 implies all (normalized) effects that can be accurately estimated are in a small neighbourhood of ξ_E^*. The average group effect ξ_A in $(15)^\dagger$ has simple expression and interpretation. Since $\xi_E^* \to \xi_A$ as $r_M \to 1$, ξ_A is in general in the small neighbourhood of ξ_E^* containing all effects that can be accurately estimated and $var(\hat{\xi}_A) \approx \sigma^2/q^2$. Because of this and the simplicity of ξ_A, we use it as the reference point to characterize the set of all effects that can be accurately estimated. Specifically, at high levels of multicollinearity, such effects are in a small neighbourhood of ξ_A,

$$\mathcal{N}_A = \{\xi'(\mathbf{w}) : ||\mathbf{w} - \mathbf{w}_a|| < \delta_1\} \tag{1}$$

where δ_1 is a small positive constant and $\mathbf{w}_a = \frac{1}{q}\mathbf{1}_q$ is the weight vector of ξ_A. Incidentally, there are same number of effects that can be accurately estimated (in the sense of having a 1-to-1 correspondence) even when variables in $\mathbf{X}_1'$ are not in an APC arrangement, but these effects would be difficult to characterize. The APC arrangement made the simple characterization (1) possible.

For the q variables in APC arrangement in $\mathbf{X}_1$ of the unstandardised model $(3)^\dagger$, the variability weighted average ξ_W in $(18)^\dagger$ is accurately estimated by $\hat{\xi}_W$ in $(19)^\dagger$ as $var(\hat{\xi}_W)$ is substantially smaller than σ^2. Other effects $\xi(\mathbf{w})$ in $(4)^\dagger$ that can be accurately estimated are in a neighbourhood of ξ_W

$$\mathcal{N}_W = \{\xi(\mathbf{w}) : ||\mathbf{w} - \mathbf{w}^*|| < \delta_2\}, \tag{2}$$

where δ_2 is a small positive constant. An alternative way to characterize $\mathcal{N}_W$ is to use $\mathcal{N}_A$ as follows. Let $\xi(\mathbf{w}) = \kappa \times \xi'(\mathbf{w}')$ where $\kappa = \sum_{i=1}^q |w_i s_i^{-1}|$ and $\xi'(\mathbf{w}')$ is a group effect for $\mathbf{X}_1'$ in the corresponding standardized model with weights $\mathbf{w}' = (w_1', w_2', \ldots, w_q')^T$ where $w_i' = w_i s_i^{-1}/\kappa$. Usually, κ is small as s_i is in general much larger than w_i. Thus, $\xi(\mathbf{w})$ can be accurately estimated if $\xi'(\mathbf{w}')$ can be accurately estimated, so an alternative expression for $\mathcal{N}_W$ is

$$\mathcal{N}_W = \{\xi(\mathbf{w}) : \xi(\mathbf{w}) \text{ such that the corresponding } \xi'(\mathbf{w}') \in \mathcal{N}_A\}.$$

Remark [b] Set $\mathcal{N}_A$ in (1) is also the set of practically important and meaningful group effects for variables in $\mathbf{X}_1'$ in that $\mathbf{w}$ values in the neighbourhood of $\mathbf{w}_a$ represent the most probable changes of the variables in $\mathbf{X}_1'$. Two extreme examples illustrate this point. (I) Effect $\beta_1' \notin \mathcal{N}_A$ as its weight vector is $\mathbf{w}_1 = (1, 0, \ldots, 0)$. It represents the group impact on response when x_1' increases by 1 unit but the other variables do not change. (II) Effect ξ_A has $\mathbf{w}_a = (1/q, 1/q, \ldots, 1/q)$, so $\xi_A \in \mathcal{N}_A$. It represents the group impact when all variables increase by $(1/q)$th of a unit. With strong positive correlations and in standardized units, the variables are likely to increase at the same time and in similar amounts. So ξ_A is practically important and meaningful whereas β_1' is not. In fact, estimating β_1' alone amounts to extreme extrapolation and β_1' by itself is neither meaningful nor interpretable as one cannot just increase x_1' by 1 unit while holding other variables constant under strong correlations among variables. Another example showing individual parameters are not meaningful is the extreme case of perfect correlation with $x_1' = \cdots = x_q' = x'$. Let $c = \beta_1' + \cdots + \beta_q'$. Then, the collective impact of these q variables on the response is cx'. There are infinitely many sets of β_i' that sum up to c. The data $(\mathbf{X}', \mathbf{y}')$ contains no information on which set is in the true model. In this sense, it contains no information about the individual β_i'. Similarly, the data contains little information about the individual β_i' when the level of multicollinearity is high. The large variances of the least squares estimators for β_i' are warnings for this lack of information. As such they should not be viewed as merely a numerical problem caused by the ill-conditioning of the $\mathbf{X}'^T\mathbf{X}'$ matrix. This lack of information is always a problem regardless the method of regression used. With this understanding, we should focus on estimating c, or equivalently $\xi_A = c/q$, and group effects in $\mathcal{N}_A$, not individual β_i'.

For the strongly correlated variables in $\mathbf{X}_1$ in the unstandardised model $(3)^\dagger$, a group effect is meaningful if and only if the corresponding effect in the standardized model is meaningful, so $\mathcal{N}_W$ is the set of meaningful effects.

Remark [c] Set $\mathcal{N}_A$ leads to the following geometric characterization of linear combinations $c_1\beta_1' + c_2\beta_2' + \cdots + c_q\beta_q'$ that can be accurately estimated for the standardized model $(8)^\dagger$. A linear combination can be expressed as $c_t\xi'(\mathbf{w})$ where $c_t = \sum_{i=1}^q |c_i|$ and $\mathbf{w} = {c_t}^{-1}(c_1, c_2, \ldots, c_q)^T$. Its minimum-variance unbiased linear estimator is $c_t\hat{\xi}'(\mathbf{w})$, so it can be accurately estimated when $var(c_t\hat{\xi}'(\mathbf{w})) = c_t^2 var(\hat{\xi}'(\mathbf{w}))$ is smaller than or comparable to σ^2. This happens under one of the following two conditions: (i) $\xi'(\mathbf{w}) \in \mathcal{N}_A$ and c_t is not too large, or (ii) $\xi'(\mathbf{w}) \notin \mathcal{N}_A$ but c_t is very small. These two conditions and $\mathcal{N}_A$ imply that in the 2-dimensional case where $q = 2$, points (c_1, c_2) representing linear combinations that can be accurately estimated form a band centred around the line $c_1 = c_2$. In higher dimensions where $q > 2$, they form a hyper-cylinder centred around the line $c_1 = c_2 = \cdots = c_q$. This observation will be used for discussing prediction accuracy in the next section.

Table 1 Correlation coefficients of the 6 variables in data matrix $\mathbf{X}_d$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$
$\mathbf{x}_1$	1.00	0.90	-0.34	-0.34	-0.06	0.14
$\mathbf{x}_2$	0.90	1.00	-0.27	-0.20	-0.25	0.38
$\mathbf{x}_3$	-0.34	-0.27	1.00	0.96	-0.41	-0.53
$\mathbf{x}_4$	-0.34	-0.20	0.96	1.00	-0.49	-0.44
$\mathbf{x}_5$	-0.06	-0.25	-0.41	-0.49	1.00	0.03
$\mathbf{x}_6$	0.14	0.38	-0.53	-0.44	0.03	1.00

2 Simulation study on group least squares regression

2.1 The linear model used in this study

Throughout this simulation study, we will use model (3) below with 6 predictor variables in 4 groups $\mathbf{X}_1 = [\mathbf{x}_1, \mathbf{x}_2]$, $\mathbf{X}_2 = [\mathbf{x}_3, \mathbf{x}_4]$, $\mathbf{X}_3 = [\mathbf{x}_5]$ and $\mathbf{X}_4 = [\mathbf{x}_6]$,

$$\mathbf{y} = \beta_0 \mathbf{1}_n + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{X}_4\boldsymbol{\beta}_4 + \boldsymbol{\varepsilon}, \tag{3}$$

where $\beta_0 = 3$, $\boldsymbol{\beta}_1 = (\beta_1, \beta_2)^T = (0, 0)^T$, $\boldsymbol{\beta}_2 = (\beta_3, \beta_4)^T = (1, 2)^T$, $\boldsymbol{\beta}_3 = \beta_5 = 0$, $\boldsymbol{\beta}_4 = \beta_6 = 3$ and $\boldsymbol{\varepsilon}$ is the n-variate standard normal random error, so $\sigma^2 = 1$. We use 6 independent n-variate standard normal random vectors $\mathbf{z}_i$ and three parameters (v_1, v_2, γ) to generate the 6 variables as follows so that groups $\mathbf{X}_1 = [\mathbf{x}_1, \mathbf{x}_2]$ and $\mathbf{X}_2 = [\mathbf{x}_3, \mathbf{x}_4]$ are, respectively, strongly correlated groups:

$$\begin{aligned} &\mathbf{x}_1 = \mathbf{z}_1, \quad \mathbf{x}_2 = \gamma[v_1\mathbf{z}_1 + (1 - v_1)\mathbf{z}_2]; \\ &\mathbf{x}_3 = \mathbf{z}_3, \quad \mathbf{x}_4 = \gamma[v_2\mathbf{z}_3 + (1 - v_2)\mathbf{z}_4]; \\ &\mathbf{x}_5 = \mathbf{z}_5, \quad \mathbf{x}_6 = \gamma\mathbf{z}_6. \end{aligned} \tag{4}$$

The theoretical non-zero correlation coefficients among the variables are:

$$\begin{aligned} \rho_{12} = \rho_{21} &= v_1[v_1^2 + (1 - v_1)^2]^{-1/2}, \\ \rho_{34} = \rho_{43} &= v_2[v_2^2 + (1 - v_2)^2]^{-1/2}. \end{aligned}$$

We see from the above formulas that $\rho_{12} \to 1$ when $v_1 \to 1$ and $\rho_{34} \to 1$ when $v_2 \to 1$, so large values of the weights v_1 and v_2 generate strong correlations among variables of the two groups. For the simulation study, we need to use observed values of $\mathbf{x}_i$. The sample correlation coefficients of the observed values of $\mathbf{x}_i$ differ somewhat from the theoretical values given by the formulas.

We set $n = 12$, $v_1 = 0.7$, $v_2 = 0.8$ and $\gamma = 2$. Matrix $\mathbf{X}_d = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_6]$ containing observed values of the 6 variables randomly generated using (4) is given in "R display 1" below. The full design matrix is $\mathbf{X} = [\mathbf{1}_n, \mathbf{X}_d]$. Table 1 contains the sample correlations of the 6 variables in matrix $\mathbf{X}_d$. It shows strong within-group correlations for groups $\mathbf{X}_1$ and $\mathbf{X}_2$ but weak between-group correlations. This simulation study involves unstandardised model (3). A standardized example is given in the Hald cement analysis.

```
R display 1: The design matrix Xd for all examples in this simulation study.
             The full design matrix is X=[1,Xd].
> Xd
             [,1]        [,2]        [,3]       [,4]        [,5]        [,6]
 [1,]  1.33247194  2.38707243  0.35045404  1.1355655 -1.66362725  0.82837127
 [2,]  0.82081027 -0.04932373 -1.81765385 -3.3503997  1.76569602  0.43909989
 [3,] -0.29595458 -0.27168960  0.04750956  0.7710956  0.50504306 -1.07289930
 [4,] -0.45687467 -0.96368003  0.79497781  1.6863252 -0.22227593 -1.92318639
 [5,]  0.62474607  0.01700248  1.68893821  2.4008808 -0.82581051 -2.15037060
 [6,]  0.05469564  0.40265862 -0.71020015 -1.1235155 -0.80982723  1.37227484
 [7,]  0.30456557  0.37345144 -1.47371005 -1.7492288  0.93406886  0.82796429
 [8,]  0.48008957  1.35339554 -0.42040266  0.2643296 -0.01488494  3.73023350
 [9,] -0.68291613 -0.56048771  1.58447035  2.3769584 -0.90045687 -0.57890494
[10,]  1.61956212  2.33300610  0.09129845  0.2557185 -0.36214200  0.07201769
[11,]  2.84612051  3.24706230 -0.95907566 -1.1348475 -0.31756247 -0.26719905
[12,]  0.60236279  0.73704811  0.86278183  1.0274744  1.91966047 -0.32319049
```

2.2 Group approach to estimation and inference

For a group of strongly correlated variables in an unstandardised model, the group approach studies only meaningful group effects in the neighbourhood of its variability weighted average (2). To compare such effects with effects not in the neighbourhood, we consider the following 6 effects for model (3):

1. $\xi_1 = w_{11}^*\beta_1 + w_{12}^*\beta_2$: variability weighted average for group $\mathbf{X}_1$.
2. $\xi_2 = w_{21}^*\beta_3 + w_{22}^*\beta_4$: variability weighted average for group $\mathbf{X}_2$.
3. $\xi_3 = \frac{1}{2}(\beta_1 - \beta_2)$: half difference effect for group $\mathbf{X}_1$.
4. $\xi_4 = \frac{1}{2}(\beta_5 - \beta_6)$: half difference effect between $\mathbf{x}_5$ and $\mathbf{x}_6$.
5. $\xi_5 = \frac{1}{2}(\beta_3 + \beta_4)$: average group effect for group $\mathbf{X}_2$.
6. $\xi_6 = (w_{21}^* - \delta)\beta_3 + (w_{22}^* + \delta)\beta_4$: an effect in the neighbourhood of ξ_2.

Using $(17)^\dagger$ and the data in "R display 1", the weight vectors for ξ_1 and ξ_2 are found to be $(w_{11}^*, w_{12}^*) = (0.42847, 0.57152)$ and $(w_{21}^*, w_{22}^*) = (0.39177, 0.60822)$, respectively. The exact values of the 6 effects are $0, 1.60822, 0, -1.5, 1.5, 1.65822$, respectively. Table 2 gives the means and variances of 1000 minimum-variance unbiased linear estimates for these group effects and the six parameters β_i of model (3). The minimum-variance unbiased linear estimates for each effect are computed by replacing each β_i in the effect with its least squares estimate $\hat{\beta}_i$; for example, that for ξ_3 is $\hat{\xi}_3 = \frac{1}{2}(\hat{\beta}_1 - \hat{\beta}_2)$. We used the same design matrix $\mathbf{X}_d$ in "R display 1" and model (3) to randomly generate 1000 $\mathbf{y}$'s. Each estimate is computed by using one of the 1000 $(\mathbf{X}_d, \mathbf{y})$ pairs.

Table 2 shows ξ_1 and ξ_2 are accurately estimated with very small variances relative to the error variance $\sigma^2 = 1$. Effect ξ_3 is the half difference effect for $\mathbf{X}_1$ which is not in the neighbourhood of ξ_1 as its weight vector $(0.5, -0.5)$ is not close to (w_{11}^*, w_{12}^*), so it is poorly estimated with a large variance. But since ξ_3 measures the expected change in the response when x_1 increases by half a unit and x_2 decreases by half a unit at the same time which is unlikely to occur given the strong positive correlation between x_1 and x_2, it

Table 2 Mean and variance of 6 estimated group effects and 6 estimated individual effects based on 1000 simulated values.

Effect	Mean	Variance	Effect	Mean	Variance
ξ_1	0.01009	0.02643	β_1	0.01604	2.16007
ξ_2	1.61319	0.03534	β_2	0.00526	1.37544
ξ_3	0.05936	1.68234	β_3	1.01535	1.66295
ξ_4	-1.49600	0.08343	β_4	1.98636	0.82435
ξ_5	1.50585	0.06974	β_5	0.00688	0.13240
ξ_6	1.66424	0.05442	β_6	3.00181	0.14773

is not a practically meaningful effect, so we are not interested in ξ_3 and thus not concerned that it cannot be accurately estimated. Effect ξ_4 is also a half difference effect but for weakly correlated $\mathbf{x}_5$ and $\mathbf{x}_6$. It is accurately estimated. Effect ξ_5 is the average group effect of $\mathbf{X}_2$. It is accurately estimated as it is in the neighbourhood of the variability weighted average effect ξ_2. Effect ξ_6 of $\mathbf{X}_2$ will be in the neighbourhood of ξ_2 when δ is small. For the ξ_6 in Table 2, $\delta = 0.05$, so it is accurately estimated. Parameters $\beta_1, \beta_2, \beta_3$ and β_4 for the two strongly correlated groups are poorly estimated but β_5 and β_6 are accurately estimated. In real applications, there is only one response vector $\mathbf{y}$ and thus only one estimated value $\hat{\xi}(\mathbf{w}) = w_1\hat{\beta}_1 + w_2\hat{\beta}_2 + \cdots + w_6\hat{\beta}_6$ for an effect $\xi(\mathbf{w})$. To assess whether $\hat{\xi}(\mathbf{w})$ is accurate, we may use the estimated variance $\widehat{var}(\hat{\xi})$ which can be computed by using (14) with $\boldsymbol{x}_+ = (0, w_1, \ldots, w_6)$.

To test hypotheses or construct confidence intervals for $\xi(\mathbf{w})$, we use

$$T = \frac{\hat{\xi}(\mathbf{w}) - \xi(\mathbf{w})}{\sqrt{\widehat{var}(\hat{\xi})}} \tag{5}$$

which has a t_{n-7} distribution under the null hypothesis. To summarize, for strongly correlated variables in an unstandardised model, meaningful group effects in the neighbourhood of the variability weighted average are accurately estimated. For variables not strongly correlated with other variables, least squares estimates for their parameters and effects are not affected by multicollinearity and are accurate. Hypothesis test and confidence interval for group effects can be conducted/constructed by using the t statistic in (5).

2.3 Group approach to variable/model selection

Traditional methods of variable selection such as all subsets regression and stepwise selection allow variables to be selected one at a time. Multicollinearity creates problems for these methods as often only one variable from a strongly correlated group is selected and different methods may choose very different models. The group approach does variable selection at the group level so that variables in a group are either all in or all out. We now illustrate this through all

subsets regression for model (3). Recall that $\beta_0 = 2$, $\boldsymbol{\beta}_1 = (0,0)^T$, $\boldsymbol{\beta}_2 = (1,2)^T$, $\boldsymbol{\beta}_3 = 0$ and $\boldsymbol{\beta}_4 = 3$, so the "true model" is the 3-variable model:

$$\mathbf{y} = \beta_0 \mathbf{1}_n + \beta_3 \mathbf{x}_3 + \beta_4 \mathbf{x}_4 + \beta_6 \mathbf{x}_6 + \boldsymbol{\varepsilon}.$$

There are $2^6 - 1 = 63$ non-empty models with at least one variable. Among these, 15 are what we call "group models" where x_1 and x_2 are in or out at the same time, and x_3 and x_4 are in or out at the same time. Using R package "leaps" by Lumley and Miller (2017), we performed all subsets regression with the adjusted R^2 criterion 100 times using 100 sets of simulated data from model (3). In each run, the model with the highest adjusted R^2 value among all 63 models is the choice of the traditional all subsets regression and that among the 15 group models is the choice of the group approach to all subsets regression. Table 3 summarizes the results of the 100 runs. It contains the 21 models that had been chosen at least once by either method. We make the following observations based on results of these 100 runs:

Table 3 Percentage of times a model is chosen by the traditional all subsets regression (Pct_1) and group approach to all subsets regression (Pct_2). Only the 21 models that were chosen at least once by either method are listed in this table.

Model	Group model?	Pct_1	Pct_2
x_3, x_4, x_6	Yes	14%	45%
x_3, x_4, x_5, x_6	Yes	3%	22%
x_1, x_2, x_3, x_4, x_6	Yes	2%	18%
$x_1, x_2, x_3, x_4, x_5, x_6$	Yes	5%	15%
x_4, x_6	No	18%	0%
x_4, x_5, x_6	No	11%	0%
x_1, x_5, x_6	No	4%	0%
x_1, x_3, x_4, x_6	No	2%	0%
x_1, x_2, x_4, x_6	No	8%	0%
x_2, x_4, x_6	No	2%	0%
x_2, x_3, x_4, x_6	No	6%	0%
x_1, x_2, x_4, x_5, x_6	No	4%	0%
x_1, x_3, x_4, x_5, x_6	No	5%	0%
x_1, x_4, x_5, x_6	No	1%	0%
x_2, x_3, x_4, x_5, x_6	No	1%	0%
x_2, x_4, x_5, x_6	No	5%	0%
x_1, x_2, x_3, x_6	No	1%	0%
x_1, x_2, x_3, x_5, x_6	No	3%	0%
x_3, x_6	No	1%	0%
x_3, x_5, x_6	No	3%	0%
x_2, x_3, x_6	No	1%	0%

1. In the 100 simulation runs, 4 of the 15 group models (roughly 1/4) were chosen at least once by the group approach, but 21 of 63 models (or 1/3) were chosen by the traditional method, so the group approach is more stable in its selection. The true model containing $\{x_3, x_4, x_6\}$ was chosen 45% of the time by the group approach but only 14% of the time by the traditional method, so the group approach is also more accurate.

2. When the traditional and group approach picked two different models, their adjusted R^2 values typically differ by less than 1%. This shows the group approach is competitive in terms of the adjusted R^2 of the chosen model.
3. All 4 models picked by the group approach at least once contain all relevant variables (variables with $\beta_i \neq 0$). In contrast, 80% of the models picked by the traditional method have missed at least one relevant variable.

The above example involves all subsets regression with the adjusted R^2 criterion. We may apply the group approach to all subsets regression with the Akaike Information Criterion or to the forward selection. Numerical results show that under the group approach, different model selection methods are more consistent in that they are more likely to select the same model.

2.4 Group approach to prediction accuracy analysis

Multicollinearity often leads to poor predictions, but it is known that accurate predictions may be achieved in an area of the predictor variable space. This area is usually expressed through an approximate linear constraint involving all predictor variables; see for example (9.1) on page 286 and remarks about prediction accuracy on page 290 in Montgomery, Peck and Vining (2012). However, such a constraint provides only a vague description of the area where accurate predictions can be achieved. We now take the group approach to characterize this area and also address the misconception about prediction accuracy of the least squares estimated model mentioned in the paper.

Consider the expected response at $\boldsymbol{x} = (x_1, \ldots, x_6)$ under model (3),

$$E(y|\boldsymbol{x}) = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + x_5\beta_5 + x_6\beta_6, \tag{6}$$

where β_j are the unknown parameters and $\boldsymbol{x}$ is a row vector containing values of the 6 predictor variables. The predicted value for $E(y|\boldsymbol{x})$ by the least squares estimated model is

$$\hat{y} = \hat{\beta}_0 + x_1\hat{\beta}_1 + x_2\hat{\beta}_2 + x_3\hat{\beta}_3 + x_4\hat{\beta}_4 + x_5\hat{\beta}_5 + x_6\hat{\beta}_6, \tag{7}$$

where $\hat{\beta}_j$ are the least squares estimates of β_j. Let $\mathbf{y}' = \mathbf{y} - \bar{y}$ be the centred version of $\mathbf{y}$ and $\mathbf{x}_i'$ be the standardized version of $\mathbf{x}_i$ in model (3). Then,

$$\mathbf{y}' = \mathbf{x}_1'\beta_1' + \mathbf{x}_2'\beta_2' + \mathbf{x}_3'\beta_3' + \mathbf{x}_4'\beta_4' + \mathbf{x}_5'\beta_5' + \mathbf{x}_6'\beta_6' + \boldsymbol{\varepsilon} \tag{8}$$

is the standardized version of model (3). Let $\hat{\beta}_i'$ be the least squares estimates for parameters of (8). They are related to $\hat{\beta}_j$ in (7) as follows,

$$\hat{\beta}_0 = \bar{y} - \sum_{i=1}^{6} \bar{x}_i\hat{\beta}_i'/s_i \quad \text{and} \quad \hat{\beta}_i = \hat{\beta}_i'/s_i \quad \text{for } i = 1, 2, \ldots, 6, \tag{9}$$

where $\bar{x}_i$ and s_i are defined above equation (7)† in the paper. By (7) and (9),

$$\hat{y} = (\bar{y} - \sum_{i=0}^{6} \bar{x}_i\hat{\beta}_i'/s_i) + x_1(\hat{\beta}_1'/s_1) + \cdots + x_6(\hat{\beta}_6'/s_6). \tag{10}$$

Define the "standardized" version of $\boldsymbol{x}$, $\boldsymbol{x}' = (x'_1, x'_2, \ldots, x'_6)$, as

$$x'_i = \frac{x_i - \bar{x}_i}{s_i} \quad \text{for } i = 1, 2, \ldots, 6. \tag{11}$$

Using (10) and (11), we obtain an expression of $\hat{y}$ in terms of $\hat{\beta}'_i$ and x'_i,

$$\hat{y} = \bar{y} + (x'_1\hat{\beta}'_1 + x'_2\hat{\beta}'_2) + (x'_3\hat{\beta}'_3 + x'_4\hat{\beta}'_4) + (x'_5\hat{\beta}'_5) + (x'_6\hat{\beta}'_6). \tag{12}$$

Since $\hat{y}$ is unbiased for $E(y|\boldsymbol{x})$, taking expectation on both sides of (12) shows that $E(y|\boldsymbol{x})$ is the sum of the expectations of the 5 terms in the right-hand side of (12). Thus, if all 5 terms accurately estimate their respective expectations, then $\hat{y}$ is an accurate estimate of $E(y|\boldsymbol{x})$. As a sample mean, the $\bar{y}$ accurately estimates $E(y)$. Also, $\hat{\beta}'_5$ and $\hat{\beta}'_6$ are accurate estimators as they are for parameters of variables not strongly correlated with others, so $x'_5\hat{\beta}'_5$ and $x'_6\hat{\beta}'_6$ accurately estimate their expected values. Since x'_1 and x'_2 are strongly correlated, by Remark [c] in Section 1, $(x'_1\hat{\beta}'_1 + x'_2\hat{\beta}'_2)$ accurately estimates its expectation $(x'_1\beta'_1 + x'_2\beta'_2)$ if $(x'_1, x'_2) \in \mathcal{C}'_1$ where $\mathcal{C}'_1$ is a band centred around the line $x'_1 = x'_2$. Similarly, $(x'_3\hat{\beta}'_3 + x'_4\hat{\beta}'_4)$ accurately estimates $(x'_3\beta'_3 + x'_4\beta'_4)$ if $(x'_3, x'_4) \in \mathcal{C}'_2$ where $\mathcal{C}'_2$ is a band centred around the line $x'_3 = x'_4$. Thus, the region of $\boldsymbol{x}'$ over which $\hat{y}$ is an accurate estimation for $E(y|\boldsymbol{x})$ is

$$\mathcal{R}'_{FP} = \mathcal{C}'_1 \times \mathcal{C}'_2 \times \mathbb{R}^2 \tag{13}$$

where the $\mathbb{R}^2$ represents no restrictions on variables x'_5 and x'_6 as they are not strongly correlated with other variables. We call the region in (13) the *feasible prediction region* for the least squares estimated model (7). In terms of the unstandardised variable $\boldsymbol{x}$, the feasible prediction region is

$$\mathcal{R}_{FP} = \{\boldsymbol{x} : \boldsymbol{x} \text{ such that its corresponding } \boldsymbol{x}' \in \mathcal{R}'_{FP}\}.$$

In simple terms, the feasible prediction region is the region in the predictor variable space where each group of strongly correlated variables in their APC arrangement are approximately equal after standardization (11). The least squares estimated model gives accurate predictions over this region.

The variance of a predicted value $var(\hat{y})$ is estimated by

$$\widehat{var}(\hat{y}) = \hat{\sigma}^2\boldsymbol{x}_+(\mathbf{X}^T\mathbf{X})^{-1}\boldsymbol{x}_+^T, \tag{14}$$

where $\boldsymbol{x}_+ = (1, \boldsymbol{x}) = (1, x_1, \ldots, x_6)$ and $\hat{\sigma}^2$ is the mean squared error. The accuracy of $\widehat{var}(\hat{y})$ depends only on the accuracy of $\hat{\sigma}^2$ as an estimator for σ^2 which is known to be good and unaffected by multicollinearity. Thus, $\widehat{var}(\hat{y})$ is in general accurate and unaffected by the multicollinearity in the data.

To illustrate $\mathcal{R}_{FP}$, we make predictions using the least squares estimated model (7) and the ridge regression at the following three points:

$$\begin{aligned}
\boldsymbol{x}_1 &= (0.60413, 0.75045, 0.00328, 0.21336, 1, 2),\\
\boldsymbol{x}_2 &= (0.93025, 1.27245, 0.75025, 1.48901, 1, 2),\\
\boldsymbol{x}_3 &= (1.58247, 1.18545, 0.75025, 3.11257, 1, 2).
\end{aligned}$$

Table 4 Comparison of the least squares predictor and Ridge regression predictor for $E(y)$ at predictor vector values $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ in terms of estimated bias (in absolute value) and MSE based on 1000 simulated values of each predictor.

$\boldsymbol{x}$ values	Exact $E(y)$	Least squares		Ridge regression	
		Bias	MSE	Bias	MSE
$\boldsymbol{x}_1$	9.43000	0.02184	0.78324	0.29714	0.83051
$\boldsymbol{x}_2$	12.72829	0.03562	1.41920	0.42563	1.55798
$\boldsymbol{x}_3$	15.97541	0.10922	9.91271	1.02438	7.84208

Using (11) and $\mathbf{X}_d$ in "R display 1" in Section 1 of this document, we can find the standardized versions of the three points, and they are

$$\begin{aligned}\boldsymbol{x}_1' &= (0, 0, 0, 0, *, *),\\ \boldsymbol{x}_2' &= (0.10, 0.12, 0.20, 0.22, *, *),\\ \boldsymbol{x}_3' &= (0.30, 0.10, 0.20, 0.50, *, *),\end{aligned}$$

where the standardized values of x_5 and x_6 are not shown as they are irrelevant for the present discussion. From the standardized values of the first four variables which are in strongly correlated groups, we see that $\boldsymbol{x}_1$ is at the centre of $\mathcal{R}_{FP}$ as $\boldsymbol{x}_1'$ is at the centre of $\mathcal{R}_{FP}'$; $\boldsymbol{x}_2$ is also in $\mathcal{R}_{FP}$ as $\boldsymbol{x}_2'$ is in $\mathcal{R}_{FP}'$ ($0.10 \approx 0.12$ and $0.20 \approx 0.22$), but $\boldsymbol{x}_3$ is not in $\mathcal{R}_{FP}$ as $\boldsymbol{x}_3'$ is not in $\mathcal{R}_{FP}'$ ($0.30 \not\approx 0.10$ and $0.20 \not\approx 0.50$).

Table 4 contains the bias and MSE of the least squares predictor (7) and the ridge regression predictor based on 1000 simulated values of the two predictors computed by using the same design matrix $\mathbf{X}_d$ but 1000 different $\mathbf{y}$ values simulated using model (3). The least squares predictor has small bias at all three $\boldsymbol{x}_i$ points as it is unbiased. Its MSE is small at $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ but large at $\boldsymbol{x}_3$ because $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are in $\mathcal{R}_{FP}$ but $\boldsymbol{x}_3$ is not. The ridge regression predictions were computed by using R package "glmnet" by Friedman *et al.* (2017) with the optimal λ value in $(0.01, 1000)$. It has bigger biases than the least squares predictor at all three points. At $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, its MSE is larger than that of the least squares predictor. At $\boldsymbol{x}_3$, its MSE is smaller but is still large in absolute terms. We have compared the two predictors using other examples and observed the same behaviour: at an $\boldsymbol{x} \in \mathcal{R}_{FP}$, both predictors are accurate but the least squares predictor is more accurate with smaller bias and smaller MSE. Outside $\mathcal{R}_{FP}$, the ridge regression predictor has a smaller MSE but a larger bias, and neither estimator is very accurate.

The misconception that the ridge regression gives more accurate predictions than the least squares regression was based on comparing prediction accuracy outside $\mathcal{R}_{FP}$ which was unknowingly done as the concept of feasible prediction region $\mathcal{R}_{FP}$ was previously unavailable. From (12), we see that making a prediction amounts to estimating a set of group effects. Making predictions over $\mathcal{R}_{FP}$ involves estimating meaningful effects, but doing so outside $\mathcal{R}_{FP}$ involves estimating effects that are not meaningful (see Remarks [a] and [b] in Section 1). Thus, predictions outside $\mathcal{R}_{FP}$ are also not meaningful,

Table 5 Correlations of original Hald cement data (left) and renamed data (right)

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$		$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
$\mathbf{x}_1$	1.00	0.22	-0.82	-0.24	$\mathbf{x}_1$	1.00	0.82	0.22	0.24
$\mathbf{x}_2$	0.22	1.00	-0.13	-0.97	$\mathbf{x}_2$	0.82	1.00	0.13	0.02
$\mathbf{x}_3$	-0.82	-0.13	1.00	0.02	$\mathbf{x}_3$	0.22	0.13	1.00	0.97
$\mathbf{x}_4$	-0.24	-0.97	0.02	1.00	$\mathbf{x}_4$	0.24	0.02	0.97	1.00

and they should not be used for comparison. When we compare meaningful predictions over $\mathcal{R}_{FP}$, the least squares predictor is more accurate.

Finally, as an example of estimating the variance of the least squares predictor with formula (14), for the 3 points in Table 4, the average of 1000 estimates by (14) are 0.72335, 1.38200 and 9.21323, respectively, which match the MSE's in Table 4 closely. On the other hand, there is no simple formula for estimating the variance of the ridge regression predictor when λ is optimized through cross-validation. There is also no formula for estimating its bias.

3 The Hald cement data analysis

The Hald cement data has been widely used in the literature to illustrate multicollinearity; see, for example, Draper and Smith (1998). The data set contains 13 observations with 4 predictor variables and a response y:

y = heat evolved in calories per gram of cement;
x_1 = amount of tricalcium aluminate;
x_2 = amount of tricalcium silicate;
x_3 = amount of tetracalcium alumino ferrite;
x_4 = amount of dicalcium silicate.

The Hald cement data set is available from various public sources. For convenience, we give this data set in "R display 2" at the end of this section.

We first illustrate the APC arrangement of a group of strongly correlated variables using this data set. In Table 5, the correlation matrix on the left is that of the four predictor variables. It shows that there are two strongly correlated groups $\{x_1, x_3\}$ and $\{x_2, x_4\}$ with negative correlation within each group, so $\{x_1, -x_3\}$ and $\{x_2, -x_4\}$ are their APC arrangements. For convenience, we rename the variables so that x_1 is still the same but the old $-x_3$ is now called x_2, the old x_2 now called x_3, and the old $-x_4$ now called x_4. The renamed data is in "R display 3" at the end of this section. The correlation matrix of the renamed variables is on the right of Table 5. The strongly correlated groups are now $\{x_1, x_2\}$ and $\{x_3, x_4\}$, both in APC arrangement. There are no strong correlations between variables from different groups.

For model $(8)^\dagger$ with the standardized renamed variables, the matrix $\mathbf{X}'^T\mathbf{X}'$ in $(10)^\dagger$ is just the correlation matrix on the right of Table 5. Matrix $\mathbf{R}_{11}$ in $(10)^\dagger$ is the upper-left quarter of this correlation matrix, $\mathbf{R}_{22}$ is the lower-right quarter, and $\mathbf{R}_{12}$ the upper-right quarter. For $i \neq j$, r_{ij} in $\mathbf{R}_{11}$ are close to 1.

Table 6 Estimated parameter values and average group effects for the standardized model $(8)^{\dagger}$; ξ_A^1 is the average group effect for group $\{x_1', x_2'\}$, and ξ_A^2 is that for $\{x_3', x_4'\}$.

	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
β_1'	31.607	14.308	2.209	0.055
β_2'	-2.261	15.788	-0.143	0.889
β_3'	27.500	36.784	0.748	0.473
β_4'	8.353	38.762	0.215	0.834
ξ_A^1	14.673	1.456	10.072	0.000
ξ_A^2	17.927	1.571	11.409	0.000

Also, elements in $\mathbf{R}_{12}$ are all small, and this leads to small elements in

$$\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} = \begin{pmatrix} 0.06 & -0.01 \\ -0.01 & 0.22 \end{pmatrix}.$$

Thus, Theorem 1(ii) applies to group $\{x_1', x_2'\}$ in that $var(\hat{\xi}_E) \approx \sigma^2/2$ and consequently $var(\hat{\xi}_A) \approx \sigma^2/2^2$. Similarly, it also applies to group $\{x_3', x_4'\}$.

Table 6 shows the estimated values of the 4 parameters β_i' and the 2 average group effects ξ_A^i in $(15)^{\dagger}$. The β_i' are poorly estimated with large standard errors due to multicollinearity generated by the two groups of strongly correlated variables. The t-test shows they are not significantly different from zero at the 5% level. The average group effects, on the other hand, are very accurately estimated with small standard errors and are highly significant. The estimated error variance is $\hat{\sigma}^2 = 2.306^2$, so the estimated standard errors of the two average group effects based on Theorem 1(ii) is $\hat{\sigma}/2 = 1.153$. We see from Table 6 that the standard errors of the two estimated group effects are indeed close to this value. We write the least squares estimated model as

$$\hat{y}' = (31.607x_1' - 2.261x_2')_G + (27.500x_3' + 8.353x_4')_G, \tag{15}$$

where the $(\dots)_G$ notation indicates that variables inside each $(\dots)_G$ are strongly correlated. Individual estimated parameter values such as 31.607 and -2.261 inside such brackets should not be used as point estimates as the underlying parameters are not meaningful and thus not estimated; they should only be used to estimate or make inference on meaningful group effects, such as ξ_A^1 and ξ_A^2, or make predictions over the feasible prediction region.

Finally, we demonstrate that the least squares estimated model gives accurate predictions over the feasible prediction region $\mathcal{R}_{FP}$, and accurate extrapolation is also possible with this estimated model. Consider 5 points

$$\begin{aligned}
\boldsymbol{x}_1 &= (7.46153, -11.76923, 48.15385, -30.00000),\\
\boldsymbol{x}_2 &= (3.18232, -15.98495, 64.86423, -10.86569),\\
\boldsymbol{x}_3 &= (7.25776, -11.10359, 46.53671, -28.84034),\\
\boldsymbol{x}_4 &= (-4.76478, -25.08204, 75.10608, -1.00862),\\
\boldsymbol{x}_5 &= (13.57470, -18.42563, 75.10608, -47.39482).
\end{aligned}$$

Using formula (11) and the renamed Hald cement data in "R display 3", the standardized values of these 5 points are found to be:

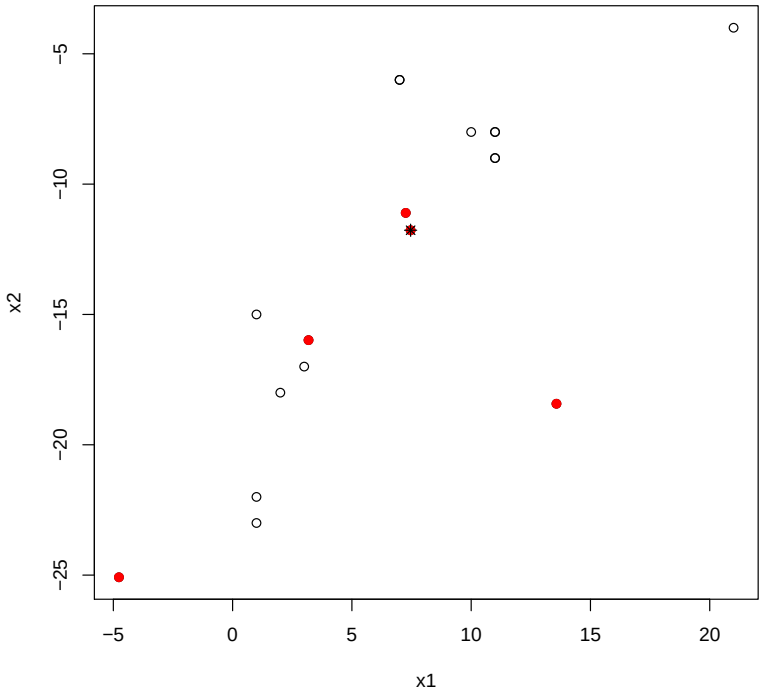

Fig. 1 Points representing (x_1, x_2) of the 13 observations in the Hald cement data are in circles. The "$\star$" symbol represents the mean of the 13 points. Points representing the 5 prediction points are in red dots. Points $\boldsymbol{x}_4$ and $\boldsymbol{x}_5$ are the two red dots outside the circle data hull, and $\boldsymbol{x}_4$ is the one in the lower left corner which is still inside the feasible prediction region. A plot of (x_3, x_4) of these points (not included) gives similar observations.

$$\begin{aligned}
\boldsymbol{x}_1' &= (0.00, 0.00, 0.00, 0.00),\\
\boldsymbol{x}_2' &= (-0.21, -0.19, 0.31, 0.33),\\
\boldsymbol{x}_3' &= (-0.01, 0.03, -0.03, 0.02),\\
\boldsymbol{x}_4' &= (-0.60, -0.60, 0.50, 0.50),\\
\boldsymbol{x}_5' &= (0.30, -0.30, 0.50, -0.30).
\end{aligned}$$

Since the strongly correlated groups in APC arrangement are $\{x_1, x_2\}$ and $\{x_3, x_4\}$, an $\boldsymbol{x}_i$ is in $\mathcal{R}_{FP}$ if its standardized version $\boldsymbol{x}_i' = (x_1', x_2', x_3', x_4')$ is in $\mathcal{R}_{FP}'$; that is, if $\boldsymbol{x}_i'$ satisfies $x_1' \approx x_2'$ and $x_3' \approx x_4'$. Thus, points $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$ and $\boldsymbol{x}_4$ are in $\mathcal{R}_{FP}$. Plotting (x_1, x_2) of the 5 points and the 13 points in the renamed Hald cement data in Figure 1 finds $\boldsymbol{x}_4$ and $\boldsymbol{x}_5$ outside the data hull of the 13 points, so making predictions at $\boldsymbol{x}_4$ and $\boldsymbol{x}_5$ is extrapolation. Table 7 gives the predicted values given by the least squares estimated model (for the renamed but unstandardised variables) and their estimated variances (14) at the 5 points. The predictions at $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are accurate with

Table 7 Predicted values and their estimated variances at 5 points

	Predicted value	Estimated variance
$\boldsymbol{x}_1$	95.423	0.460
$\boldsymbol{x}_2$	100.496	3.706
$\boldsymbol{x}_3$	94.381	7.359
$\boldsymbol{x}_4$	95.742	5.285
$\boldsymbol{x}_5$	116.827	1689.129

small variances as these points are in both the data hull and $\mathcal{R}_{FP}$. Point $\boldsymbol{x}_5$ is not in $\mathcal{R}_{FP}$ as it violated the strong positive correlation of the data (its $x_1' = 0.3$ but $x_2' = -0.3$, and its $x_3' = 0.5$ but $x_4' = -0.3$), so extrapolation at $\boldsymbol{x}_5$ is highly inaccurate with a large variance. In contrast, extrapolation at $\boldsymbol{x}_4$ is accurate with a small variance as $\boldsymbol{x}_4$ is in $\mathcal{R}_{FP}$. This shows accurate extrapolation with the least squares estimated model is possible, even when there is multicollinearity, provided it is done within $\mathcal{R}_{FP}$.

```
R display 2: The original Hald cement data.
> hald.data
          y x1 x2 x3 x4
 [1,]  78.5  7 26  6 60
 [2,]  74.3  1 29 15 52
 [3,] 104.3 11 56  8 20
 [4,]  87.6 11 31  8 47
 [5,]  95.9  7 52  6 33
 [6,] 109.2 11 55  9 22
 [7,] 102.7  3 71 17  6
 [8,]  72.5  1 31 22 44
 [9,]  93.1  2 54 18 22
[10,] 115.9 21 47  4 26
[11,]  83.8  1 40 23 34
[12,] 113.3 11 66  9 12
[13,] 109.4 10 68  8 12
```

```
R display 3: Renamed Hald cement data where the two groups of strongly correlated
             predictor variables {x1, x2} and {x3, x4} are in APC arrangement.
> renamed.data
         y    x1   x2   x3   x4
 [1,]  78.5    7   -6   26  -60
 [2,]  74.3    1  -15   29  -52
 [3,] 104.3   11   -8   56  -20
 [4,]  87.6   11   -8   31  -47
 [5,]  95.9    7   -6   52  -33
 [6,] 109.2   11   -9   55  -22
 [7,] 102.7    3  -17   71   -6
 [8,]  72.5    1  -22   31  -44
 [9,]  93.1    2  -18   54  -22
[10,] 115.9   21   -4   47  -26
[11,]  83.8    1  -23   40  -34
[12,] 113.3   11   -9   66  -12
[13,] 109.4   10   -8   68  -12
```

The following is the list of all references that have been cited in either the paper or this Supplementary Material.

References

Belsley, D. A., Kuh, E., Welsch, R. E. (2004). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley & Sons, New York.

Conniffe, D., Stone, J. (1973). A critical view of ridge regression. *American Statistician*, 22, 181–187.

Dampster, A. P., Schatzoff, M., Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association*, 72, 77–90.

Draper, N. R., Smith, H. (1998). *Applied Regression Analysis*, 3rd ed., Wiley, New York.

Draper, N. R., Van Nostrand, R. C. (1979). Ridge regression and James-Stein estimators: review and comments. *Technometrics*, 21 451–466.

Friedman, J., Hastie, T., Simon, N., Tibshirani, R. (2017). Package 'glmnet', an R package available at https://cran.r-project.org.

Gunst, R. F., Mason, R. L. (1977). Biased estimation in regression: an evaluation using mean squared error. *Journal of the American Statistical Association*, 72, 616–628.

Gunst, R. F., Webster, J. T., Mason, R. L. (1976). A comparison of least squares and latent root regression estimators. *Technometrics*, 18, 75–83.

Hoerl, A. E., Kennard, R. W. (1970). Ridge Regression: biased estimation for nonorthonogal problems. *Technometrics*, 12, 55–67.

Hoerl, A. E., Kennard, R. W., Baldwin, K. F. (1975). Ridge regression: some simulations. *Communications in Statistics: Theory and Methods*, 4, 105–123.

Horn, R. A., Johnson, C. A. (1985). *Matrix Analysis*, Cambridge University Press.

Jolliffe, I. T. (1986). *Principal component analysis.* Springer-Verlag, New York.

Lawless, J. F. (1978). Ridge and related estimation procedures: theory and practice. *Communications in Statistics: Theory and Methods*, 7, 135–164.

Lumley, T., Miller, A. (2017). Package 'leaps', an R package available at https://cran.r-project.org.

Montgomery, D. C., Peck, E. A., Vining, G. G. (2012). *Introduction to Linear Regression Analysis*, 5th ed., Wiley, New York.

Tsao, M. (2019). Estimable group effects for strongly correlated variables in linear models. *Journal of Statistical Inference and Planning*, 198, 29–42.

Webster J. T., Gunst, R. F., Mason, R. L. (1974). Latent root regression analysis. *Technometrics*, 16, 513–522.