COMPREHENSIVE WRITTEN EXAMINATION, PAPER III

FRIDAY AUGUST 26, 2005, 9:00 A.M.–1:00 P.M.

STATISTICS 174 QUESTION

Answer all parts. Closed book, calculators allowed. It is important to show all working, especially with numerical calculations. Statistical tables are provided. You may freely quote results from the course notes or text without proof, but to the extent that it is feasible to do so, state precisely the result you are quoting.

Although all parts of the question have a common theme, parts (a) and (f) (data analysis) can be answered without reference to the other parts. Tentative mark scheme: 15 points for each of (a), (c), (d), (e); 10 for (b), 30 for (f); total 100.)

(a) Suppose linear model coefficients $\beta_1, ..., \beta_p$ are themselves treated as random variables, independent with $\beta_j \sim N[0, \sigma^2/\alpha^2]$ for $j = 1, ..., p$. Also, conditionally on $\beta_1, ..., \beta_p$, we have the usual linear model equation: $y_i \sim N[\sum_j x_{ij}\beta_j, \sigma^2]$ (independently) for $i = 1, ..., n$. Here x_{ij} $(1 \le i \le n,\ 1 \le j \le p)$ are assumed known; σ^2 is unknown and $\alpha \in (0, \infty)$ is assumed known.

Suppose the above system is rewritten in the form

$$
\begin{aligned}
y_1 &= \sum_j x_{1j}\beta_j + \epsilon_1, \\
&\vdots \\
y_n &= \sum_j x_{nj}\beta_j + \epsilon_n, \\
0 &= \alpha\beta_1 + \epsilon_{n+1}, \\
&\vdots \\
0 &= \alpha\beta_p + \epsilon_{n+p},
\end{aligned}
$$

with $\epsilon_1, ..., \epsilon_{n+p}$ independent $N[0, \sigma^2]$.

Based on this formulation, what are the least squares estimation equations for $\beta_1, ..., \beta_p$? What is the connection with ridge regression?

(b) We now revert to the usual formulation of ridge regression, in which β is treated as a fixed set of regression coefficients and the usual least squares estimator is replaced by $\tilde{\beta} = (X^TX + cI_p)^{-1}X^TY$ for some $c > 0$. Let $b_c = \mathrm{E}\{\tilde{\beta} - \beta\}$ denote the bias and V_c the covariance matrix of $\tilde{\beta}$. Write down expressions for b_c and V_c.

(c) Suppose we are interested in predicting a new vector of observations $Y^* = X\beta + \epsilon^*$ where $\epsilon^* \sim N[0, \sigma^2 I_n]$ independent of ϵ. Suppose we use $\tilde{Y}^* = X\tilde{\beta}$ as a predictor. A natural criterion for the quality of the predictor is the sum of mean squared prediction errors, or $\mathrm{E}\{(Y^* - \tilde{Y}^*)^T(Y^* - \tilde{Y}^*)\}$. Show that

$$
\begin{aligned}
\mathrm{E}\{(Y^* - \tilde{Y}^*)^T(Y^* - \tilde{Y}^*)\} &= c^2\beta^T(X^TX + cI_p)^{-1}X^TX(X^TX + cI_p)^{-1}\beta \\
&\quad +\sigma^2\left[p + \mathrm{tr}\left\{X^TX(X^TX + cI_p)^{-1}X^TX(X^TX + cI_p)^{-1}\right\}\right].
\end{aligned}
$$

(d) Find an expression for $\mathrm{E}\{\tilde{\beta}^T A \tilde{\beta}\}$ where A is an arbitrary matrix.

(e) Is it possible to find a matrix A and a scalar constant B (not depending on β, though they may depend on σ^2) for which $\tilde{\beta}^T A \tilde{\beta} + B$ is an unbiased estimator of $\mathrm{E}\{(Y^* - \tilde{Y}^*)^T (Y^* - \tilde{Y}^*)\}$? How might this be relevant to the selection of c?

(*Note:* You are not required to discuss the practical calculation of A and B or even to prove that they exist, but you should state equations that A and B must satisfy if they exist.)

(f) Consider the following dataset:

x1	x2	x3	x4	x5	x6	x7	y
1	8	1	1	1	0.541	-0.099	10.006
1	8	1	1	0	0.130	0.070	9.737
1	8	1	1	0	2.116	0.115	15.087
1	0	0	9	1	-2.397	0.252	8.422
1	0	0	9	1	-0.046	0.017	8.625
1	0	0	9	1	0.365	1.504	16.289
1	2	7	0	1	1.996	-0.865	5.958
1	2	7	0	1	0.228	-0.055	9.313
1	2	7	0	1	1.380	0.502	12.960
1	0	0	0	10	-0.798	-0.399	5.541
1	0	0	0	10	0.257	0.101	8.756
1	0	0	0	10	0.440	0.432	10.937

The covariates are $x1 - x7$ and the response variable is y. In order to keep a consistent notation for the covariates, the intercept is treated by including a column of ones ($x1$) among the covariates, and the regression run with the /NOINT option. The regression was performed in SAS together with the usual diagnostics. The output is included at the end of this exam.

Write a report summarizing the SAS output, concentrating on the interpretation of regression diagnostics. Is there a problem with multicollinearity in this dataset, and if so, what feature of the dataset is responsible for it?

The analysis in part (a) was fitted with four different values of α, with the following sets of parameter estimates and residual mean squared errors:

```
Alpha=0.25, Root MSE=0.90361:
```

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
x1	1	1.83781	3.38124	0.54	0.5967	168.90059
x2	1	0.87935	0.33106	2.66	0.0209	27.39085
x3	1	0.65138	0.35318	1.84	0.0900	22.92515
x4	1	0.71724	0.34691	2.07	0.0610	36.26701
x5	1	0.63550	0.34135	1.86	0.0873	43.82067
x6	1	1.08999	0.30870	3.53	0.0041	2.05034
x7	1	4.96967	0.57360	8.66	<.0001	1.52135

Alpha=0.5, Root MSE=1.12042:

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
x1	1	0.75802	2.17132	0.35	0.7331	46.00655
x2	1	0.97789	0.22729	4.30	0.0010	8.40562
x3	1	0.74905	0.24685	3.03	0.0104	7.29305
x4	1	0.84881	0.23635	3.59	0.0037	10.95810
x5	1	0.74437	0.22599	3.29	0.0064	12.49971
x6	1	1.12130	0.37737	2.97	0.0117	2.01408
x7	1	4.63634	0.68508	6.77	<.0001	1.48162

Alpha=1, Root MSE=1.61342:

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
x1	1	0.46306	1.57788	0.29	0.7742	12.43357
x2	1	0.99580	0.20137	4.95	0.0003	3.19321
x3	1	0.74991	0.22649	3.31	0.0062	2.97567
x4	1	0.94320	0.20317	4.64	0.0006	3.91660
x5	1	0.77656	0.18254	4.25	0.0011	3.94245
x6	1	1.17309	0.51747	2.27	0.0427	1.90351
x7	1	3.66320	0.86947	4.21	0.0012	1.36873

Alpha=2, Root MSE=2.37034:

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
x1	1	0.38761	1.16200	0.33	0.7445	3.84512
x2	1	0.99794	0.22025	4.53	0.0007	1.79590
x3	1	0.72775	0.25417	2.86	0.0143	1.77073
x4	1	1.03918	0.20729	5.01	0.0003	1.91188
x5	1	0.78470	0.17790	4.41	0.0008	1.75174
x6	1	1.09292	0.65622	1.67	0.1217	1.64823
x7	1	2.03947	0.92978	2.19	0.0487	1.18677

Write a brief summary of how the parameter estimates are affected by the ridge regression, and suggest which of the four values of α you would recommend (with reasons).

(*Note:* In this example, contrary to what is sometimes recommended, the ridge regression has been performed without any centering or rescaling of the X variables.)

Appendix: SAS Program and Output

Program

```
infile 'data.txt';
input x1-x7 y;
run;
;
proc reg;
model y=x1-x7 /noint collin influence r cli clm vif covb;
output r=resid p=predval;
run;
;
proc plot;
plot resid*predval;
run;
;
proc univariate plot normal;
var resid;
run;
```

Output (edited)

```
                              The REG Procedure
                           Dependent Variable: y

            NOTE: No intercept in model. R-Square is redefined.

                            Analysis of Variance

                                      Sum of           Mean
  Source                   DF        Squares         Square    F Value    Pr > F

  Model                     7     1343.99981      191.99997     150.50    <.0001
  Error                     5        6.37871        1.27574
  Uncorrected Total        12     1350.37852

            Root MSE              1.12949    R-Square     0.9953
            Dependent Mean       10.13592    Adj R-Sq     0.9887
            Coeff Var            11.14342

                            Parameter Estimates

                    Parameter      Standard                               Variance
Variable     DF      Estimate         Error    t Value    Pr > |t|       Inflation

x1            1      16.65995      14.04655       1.19      0.2889      1855.91286
```

x2	1	-0.53126	1.34179	-0.40	0.7085	287.89610
x3	1	-0.83845	1.42063	-0.59	0.5807	237.29697
x4	1	-0.77534	1.40942	-0.55	0.6059	383.04595
x5	1	-0.84396	1.40313	-0.60	0.5737	473.77140
x6	1	1.02325	0.39094	2.62	0.0472	2.09711
x7	1	5.04705	0.72766	6.94	0.0010	1.54106

Covariance of Estimates

Variable	x1	x2	x3	x4
x1	197.30550925	-18.79941384	-19.88749485	-19.75171835
x2	-18.79941384	1.800397842	1.8941694943	1.8794267473
x3	-19.88749485	1.8941694943	2.018198214	1.9880420213
x4	-19.75171835	1.8794267473	1.9880420213	1.986454603
x5	-19.68775366	1.8757064986	1.9836520292	1.9706535485
x6	-0.743368976	0.0544266957	0.0522444787	0.0907311986
x7	-0.621315373	0.0684284374	0.0835915808	0.0221143033

Variable	x5	x6	x7
x1	-19.68775366	-0.743368976	-0.621315373
x2	1.8757064986	0.0544266957	0.0684284374
x3	1.9836520292	0.0522444787	0.0835915808
x4	1.9706535485	0.0907311986	0.0221143033
x5	1.9687625723	0.0751728053	0.0595242705
x6	0.0751728053	0.1528359427	-0.071946284
x7	0.0595242705	-0.071946284	0.5294882908

Collinearity Diagnostics

		Condition	-------Proportion of Variation-------		
Number	Eigenvalue	Index	x1	x2	x3
1	2.63287	1.00000	0.00006953	0.00026211	0.00029752
2	1.82065	1.20255	0.00000957	0.00006777	0.00020920
3	1.03335	1.59622	0.00001529	0.00021404	0.00000725
4	0.65826	1.99994	0.00002004	0.00045646	0.00004750
5	0.60573	2.08485	9.681367E-9	0.00254	0.00350
6	0.24884	3.25280	2.283311E-8	0.00116	0.00234
7	0.00030936	92.25341	0.99989	0.99529	0.99360

	-----------------Proportion of Variation----------------			
Number	x4	x5	x6	x7
1	0.00005709	0.00006476	0.02169	0.00434
2	0.00048377	0.00004444	0.05233	0.09490
3	0.00015594	0.00128	0.02556	0.10103

4	0.00051279	0.00028111	0.19059	0.39582
5	0.00013989	0.00006576	0.00114	0.00023133
6	0.00277	0.00000692	0.69089	0.40030
7	0.99589	0.99826	0.01781	0.00338

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean	
1	10.0060	10.0060	1.1295	7.1026	12.9094
2	9.7370	11.2824	0.8864	9.0039	13.5608
3	15.0870	13.5416	0.8864	11.2632	15.8201
4	8.4220	7.6571	0.9203	5.2914	10.0227
5	8.6250	8.8767	0.8468	6.7000	11.0534
6	16.2890	16.8022	0.9480	14.3654	19.2390
7	5.9580	6.5610	0.9400	4.1447	8.9773
8	9.3130	8.8400	0.7653	6.8729	10.8072
9	12.9600	12.8300	0.7946	10.7874	14.8726
10	5.5410	5.3900	0.7550	3.4493	7.3307
11	8.7560	8.9931	0.6614	7.2928	10.6933
12	10.9370	10.8509	0.7159	9.0105	12.6913

<table>
<tr><th>Obs</th><th>95% CL Predict</th><th></th><th>Residual</th><th>Std Error Residual</th><th>Student Residual</th><th>-2-1 0 1 2</th></tr>
<tr><td>1</td><td>5.8999</td><td>14.1121</td><td>-2.25E-13</td><td>2.97E-7</td><td>-759E-9</td><td>| | |</td></tr>
<tr><td>2</td><td>7.5917</td><td>14.9731</td><td>-1.5454</td><td>0.700</td><td>-2.207</td><td>| ****| |</td></tr>
<tr><td>3</td><td>9.8509</td><td>17.2323</td><td>1.5454</td><td>0.700</td><td>2.207</td><td>| |**** |</td></tr>
<tr><td>4</td><td>3.9119</td><td>11.4023</td><td>0.7649</td><td>0.655</td><td>1.168</td><td>| |** |</td></tr>
<tr><td>5</td><td>5.2479</td><td>12.5055</td><td>-0.2517</td><td>0.747</td><td>-0.337</td><td>| | |</td></tr>
<tr><td>6</td><td>13.0117</td><td>20.5927</td><td>-0.5132</td><td>0.614</td><td>-0.836</td><td>| *| |</td></tr>
<tr><td>7</td><td>2.7836</td><td>10.3384</td><td>-0.6030</td><td>0.626</td><td>-0.963</td><td>| *| |</td></tr>
<tr><td>8</td><td>5.3329</td><td>12.3471</td><td>0.4730</td><td>0.831</td><td>0.569</td><td>| |* |</td></tr>
<tr><td>9</td><td>9.2800</td><td>16.3800</td><td>0.1300</td><td>0.803</td><td>0.162</td><td>| | |</td></tr>
<tr><td>10</td><td>1.8977</td><td>8.8823</td><td>0.1510</td><td>0.840</td><td>0.180</td><td>| | |</td></tr>
<tr><td>11</td><td>5.6284</td><td>12.3577</td><td>-0.2371</td><td>0.916</td><td>-0.259</td><td>| | |</td></tr>
<tr><td>12</td><td>7.4133</td><td>14.2885</td><td>0.0861</td><td>0.874</td><td>0.0986</td><td>| | |</td></tr>
</table>

Obs	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS
1	1.193	-6.787E-7	1.0000	6.916E13	-2.5850
2	1.116	-12.3685	0.6158	0.0000	-15.6594
3	1.116	12.3685	0.6158	0.0000	15.6594
4	0.385	1.2252	0.6639	1.5245	1.7218
5	0.021	-0.3047	0.5620	9.2727	-0.3451
6	0.238	-0.8059	0.7044	5.6268	-1.2441

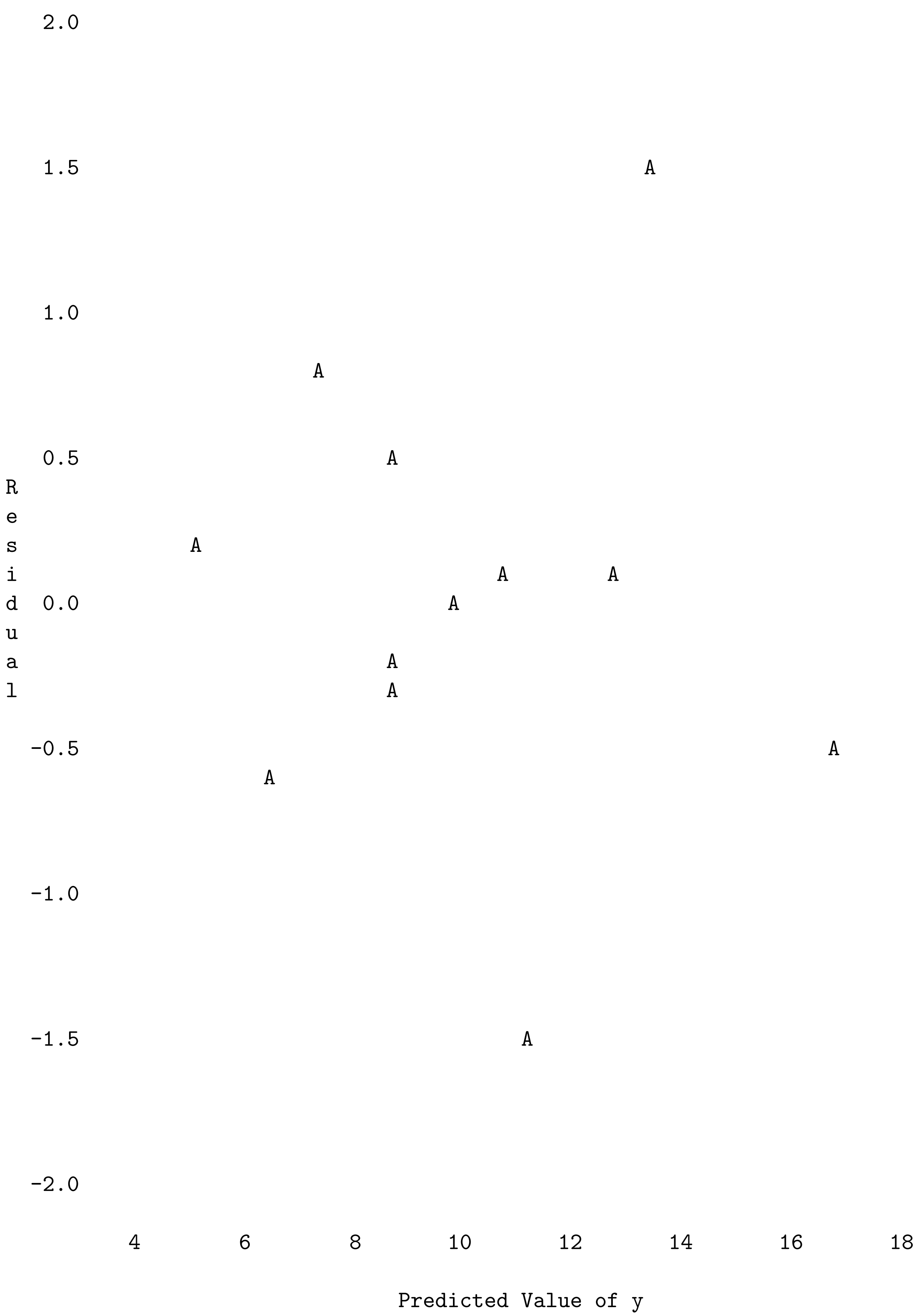
7	0.298	-0.9542	0.6926	3.6911	-1.4323
8	0.039	0.5266	0.4590	5.5134	0.4851
9	0.004	0.1452	0.4949	9.1000	0.1438
10	0.004	0.1613	0.4468	8.2369	0.1449
11	0.005	-0.2332	0.3429	6.6026	-0.1684
12	0.001	0.0882	0.4018	7.8633	0.0723

Output Statistics

	DFBETAS						
Obs	x1	x2	x3	x4	x5	x6	x7
1	2.0786	-2.1760	-2.0552	-2.0716	-2.0808	0.0000	0.0000
2	-8.9689	8.0419	8.8950	9.1314	8.9884	6.7854	-1.4453
3	7.0769	-6.5782	-7.5579	-6.8602	-7.0749	6.7854	-1.4453
4	0.1968	-0.1702	-0.1523	-0.1525	-0.1977	-1.1299	-0.1462
5	0.0036	0.0031	0.0036	-0.0270	-0.0042	-0.1461	0.1963
6	0.1265	-0.1114	-0.1232	-0.1521	-0.1252	-0.3221	-0.7346
7	0.0152	0.0162	-0.0470	-0.0606	-0.0180	-0.6769	0.9243
8	0.0303	-0.0260	0.0086	-0.0389	-0.0308	-0.2511	0.0999
9	-0.0068	0.0062	0.0160	0.0039	0.0067	-0.0087	0.0812
10	0.0173	-0.0162	-0.0170	-0.0172	-0.0115	-0.0417	-0.0475
11	-0.0032	0.0041	0.0052	0.0035	-0.0046	-0.0263	-0.0031
12	-0.0012	0.0008	0.0006	0.0003	0.0042	0.0115	0.0237

Sum of Residuals	0
Sum of Squared Residuals	6.37871
Predicted Residual SS (PRESS)	56.44282

Plot of resid*predval. Legend: A = 1 obs, B = 2 obs, etc.



The UNIVARIATE Procedure
Variable: resid (Residual)

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.963202	Pr < W	0.8284
Kolmogorov-Smirnov	D	0.17142	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.050106	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.30846	Pr > A-Sq	>0.2500

Extreme Observations

------Lowest------		------Highest-----	
Value	Obs	Value	Obs
-1.545357	2	0.130005	9
-0.602998	7	0.150976	10
-0.513211	6	0.472993	8
-0.251693	5	0.764904	4
-0.237074	11	1.545357	3

```
Stem Leaf                    #              Boxplot
   1 5                       1                 0
   1
   0 58                      2                 |
   0 0112                    4              +--+--+
  -0 32                      2              +-----+
  -0 65                      2                 |
  -1
  -1 5                       1                 0
     ----+----+----+----+
```

```
                    Normal Probability Plot
  1.75+                                          *   ++++++
      |                                        ++++++
      |                                 +++++++
      |                         *******+ *
      |                    ++*+++
      |             +++++++
      |       ++++++
 -1.75+++++++   *
       +----+----+----+----+----+----+----+----+----+----+
           -2        -1         0        +1        +2
```

SOLUTIONS

(a) Suppose $X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ as usual in linear regression, and let $\tilde{X}$, $\tilde{Y}$ denote the extended system including $0 = \alpha\beta_j + \epsilon_{n+j},\ 1 \le j \le p$. Thus

$$\tilde{X} = \begin{pmatrix} X \\ \alpha I_p \end{pmatrix}, \qquad \tilde{Y} = \begin{pmatrix} Y \\ 0_p \end{pmatrix}.$$

(I_p denotes the $p \times p$ identity matrix; 0_p denotes the p-dimensional vector of zeros.) So

$$\begin{aligned} \tilde{X}^T\tilde{X} &= \begin{pmatrix} X^T & \alpha I_p \end{pmatrix} \begin{pmatrix} X \\ \alpha I_p \end{pmatrix} = X^TX + \alpha^2 I_p, \\ \tilde{X}^T\tilde{Y} &= \begin{pmatrix} X^T & \alpha I_p \end{pmatrix} \begin{pmatrix} Y \\ 0_p \end{pmatrix} = X^TY, \\ \tilde{\beta} &= (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y} = (X^TX + \alpha^2 I_p)^{-1}X^TY, \end{aligned}$$

which is the ridge regression estimator in which the constant usually denoted c has been replaced by α^2. [*Note:* The course text states that one derivation of ridge regression is through a Bayesian approach where $\beta_j \sim N[0, \sigma^2/c]$ *a priori*, but does not include any derivation of this result.]

(b) Following the usual calculation of bias and variance in ridge regression, the bias of $\tilde{\beta}$ is

$$\begin{aligned} b_c &= (X^TX + cI_p)^{-1}X^TX\beta - \beta \\ &= (X^TX + cI_p)^{-1}(X^TX + cI_p - cI_p)\beta - \beta \\ &= -c(X^TX + cI_p)^{-1}\beta \qquad (1) \end{aligned}$$

while the covariance matrix is

$$V_c = (X^TX + cI_p)^{-1}X^TX(X^TX + cI_p)^{-1}\sigma^2. \qquad (2)$$

(*Note:* These formulae are in the course text, p. 221, and it would be acceptable if they were quoted from memory.)

(c) We have

$$Y^* - \tilde{Y}^* = X(\beta - \tilde{\beta}) + \epsilon^*$$

where the two terms are independent, and hence

$$\mathrm{E}\left\{(Y^* - \tilde{Y}^*)^T(Y^* - \tilde{Y}^*)\right\} = \mathrm{E}\left\{(\tilde{\beta} - \beta)^TX^TX(\tilde{\beta} - \beta)\right\} + \mathrm{E}\left\{\epsilon^{*T}\epsilon^*\right\}. \qquad (3)$$

If we write $\tilde{\beta} - \beta = b_c + \eta$, say, where $\eta \sim N[0, V_c]$ the first term in (3) becomes

$$b_c^TX^TXb_c + \mathrm{E}\left\{\eta^TX^TX\eta\right\} = b_c^TX^TXb_c + \mathrm{tr}\left\{X^TXV_c\right\}$$

Also, the second term in (3) is $p\sigma^2$. If we now substitute from (1) and (2) into (3), we get the claimed result.

(d) Writing $\mathrm{E}\{\tilde{\beta}\} = \beta + b_c = (I_p - c(X^T X + cI_p)^{-1})\beta$ and hence $\tilde{\beta} = (I_p - c(X^T X + cI_p)^{-1})\beta + \eta$ with $\eta \sim N[0, V_c]$, we have

$$\begin{aligned}\mathrm{E}\{\tilde{\beta}^T A \tilde{\beta}\} &= \beta^T (I_p - c(X^T X - cI_p)^{-1}) A (I_p - c(X^T X - cI_p)^{-1})\beta + \mathrm{E}\{\eta^T A \eta\} \\ &= \beta^T (I_p - c(X^T X - cI_p)^{-1}) A (I_p - c(X^T X - cI_p)^{-1})\beta + \mathrm{tr}(AV_c).\end{aligned}$$

(e) Suppose we can find a matrix A such that

$$\begin{aligned}&\left\{I_p - c(X^T X + cI_p)^{-1}\right\} A \left\{I_p - c(X^T X + cI_p)^{-1}\right\} \\ =\ & c^2 (X^T X + cI_p)^{-1} X^T X (X^T X + cI_p)^{-1}.\end{aligned} \qquad (4)$$

Then

$$\begin{aligned}\mathrm{E}\{\tilde{\beta}^T A \tilde{\beta}\} &= c^2 \beta^T (X^T X + cI_p)^{-1} X^T X (X^T X + cI_p)^{-1} \beta + \mathrm{tr}(AV_c) \\ &= \mathrm{E}\{(Y^* - \tilde{Y}^*)^T (Y^* - \tilde{Y}^*)\} + \mathrm{tr}(AV_c) \\ &\quad -\sigma^2 \left[p + \mathrm{tr}\left\{X^T X (X^T X + cI_p)^{-1} X^T X (X^T X + cI_p)^{-1}\right\}\right].\end{aligned}$$

The result will be satisfied if we define A by (4) and

$$B = \sigma^2 \left[p + \mathrm{tr}\left\{X^T X (X^T X + cI_p)^{-1} X^T X (X^T X + cI_p)^{-1}\right\}\right] - \mathrm{tr}(AV_c).$$

This suggests an alternative algorithm for estimating c, as the value that minimizes $\tilde{\beta}^T A \tilde{\beta} + B$. (The matrix A will exist if $I_p - c(X^T X + cI_p)^{-1}$ is invertible. That depends entirely on X and c, not on any unknowns of the model. As for σ^2, that of course would be unknown in practice but the value of s^2 derived from ordinary least squares is an unbiased estimator, and would therefore be a possible choice as an estimator of σ^2.)

(f) *VIFs*: Each of the coefficients for $x1, x2, x3, x4, x5$ has high VIF (but not $x6$ or $x7$).

Collinearity diagnostics: The condition index of 92.25 shows that there is a multicollinearity problem, and the high variance proportions in $x1, x2, x3, x4, x5$ again shows that all five of these contribute to the problem (but not $x6$ or $x7$).

Residuals: Observation 1 is clearly unusual, with a residual of -2.25×10^{-13} but also that the residual has a standard error of 2.97×10^{-7}. This suggests that for some reason connected with the design matrix, the regression line is forced to go through observation 1. We also have exceptional *RStudent* values for observations 2 and 3, suggesting that these are major outliers.

Leverage: The usual $\frac{2p}{n}$ criterion for a large h_i evaluates to 1.17, which is meaningless given that $0 \le h_i \le 1$ always. However we clearly have a major problem with observation 1 because of $h_i = 1$ in this case (this explains why the residual came out to be 0).

DFFITS: $2\sqrt{\frac{p}{n}} = 1.528$ so there are high DFFITS values in observations 1–4 and 7, but especially so with observations 2 and 3.

Cook's D: Large for observations 1,2,3

DFBETAS: $\frac{2}{\sqrt{n}} = 0.577$ so there are numerous problematic values — note in particular that observations 2,3,4 and 7 have significant DFBETAS in column $x6$, and observations 2,3,6 and 7 in column $x7$. So there are some problems with columns $x6$ and $x7$ as well as the earlier multicollinearity issues associated with $x1 - x5$.

Residual v. Predicted value plot and tests of normality: No apparent problems here.

The diagnostics point up a whole host of problems with this analysis, but especially a major multicollinearity involving columns $x1 - x5$.

Ridge regression results: the main point is that as α increases, the estimates get much more stable, but also in the case of $x6$ and $x7$, the results stray rather far from their initial least squares values (for $x6$, the least squares value is 1.02 with a standard error 0.39; for $x7$, estimate is 5.05 and standard error 0.73; these ought to be fairly reliable estimates given that there is no multicollinearity with respect to those two variables). Based on those values, would probably not want to go beyond $\alpha = 0.5$. On the other hand, the VIFs show there is still a problem with multicollinearity at $\alpha = 0.5$ but much less so at $\alpha = 1$. It's not a clear-cut answer but if I had to take a shot at one of the given values of α, it would probably be 0.5.

(The data set is an artificial one used by Phil Brown in his book *Measurement, Regression and Calibration* to illustrate the concept of ridge regression. The multicollinearity comes from the fact that $x2 + x3 + x4 + x5 = 10$ in every row except the first, where it is 11. This also explains why the OLS regression leads to zero residual in the first row.)