

THE UNIVERSITY OF CHICAGO
Graduate School of Business
 Business 41912-01, Spring Quarter 2008, Mr. Ruey S. Tsay

Solutions to Homework Assignment #4

1. Problem 7.3 of the textbook, p.420

Proof: Let $\mathbf{V}^{1/2}$ be the symmetric square-root matrix of \mathbf{V} . Pre-multiplying the model by $\mathbf{V}^{-1/2}$, we have

$$\mathbf{V}^{-1/2}\mathbf{Y} + \mathbf{V}^{-1/2}\mathbf{Z}\boldsymbol{\beta} = \mathbf{V}^{-1/2}\boldsymbol{\epsilon}. \quad (1)$$

Since $\text{Cov}(\mathbf{V}^{-1/2}\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$, the prior model is a multiple linear regression so that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_w &= [(\mathbf{V}^{-1/2}\mathbf{Z})'(\mathbf{V}^{-1/2}\mathbf{Z})]^{-1}(\mathbf{V}^{-1/2}\mathbf{Z})'\mathbf{V}^{-1/2}\mathbf{Y} \\ &= (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Y}). \end{aligned}$$

Also, from (??), an unbiased estimate of σ^2 is

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-r-1}(\mathbf{V}^{-1/2}\mathbf{Y} - \mathbf{V}^{-1/2}\mathbf{Z}\hat{\boldsymbol{\beta}}_w)'(\mathbf{V}^{-1/2}\mathbf{Y} - \mathbf{V}^{-1/2}\mathbf{Z}\hat{\boldsymbol{\beta}}_w) \\ &= \frac{1}{n-r-1}[\mathbf{V}^{-1/2}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_w)]'[\mathbf{V}^{-1/2}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_w)] \\ &= \frac{1}{n-r-1}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_w)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_w). \end{aligned}$$

2. Problem 7.17 of the textbook, p. 423

Answer: (a) The fitted linear regression model is

$$\text{Profits} = .013 + .068\text{sales} + 0.006\text{Assets} + \epsilon$$

The residual standard error is 3.863 and $R^2 = 0.56$. (b) The plot of standardized residuals looks fine. The leverages of the data are given below:

```
> print(diag(H), digits=3)
[1] 0.626 0.101 0.243 0.222 0.251 0.275 0.279 0.364 0.203 0.436
```

From the values, the first data point, Citigroup, appears to be influential. (c) The 95% predictive interval is $[-1.55, 20.95]$. (d) The t-ratio of β_2 is 1.17 with p-value 0.28. Thus, the model can be modified by removing the “Assets” variable.

3. Problem 7.25 of the textbook, p. 426-427

Answer: (a) The fitted linear regression model is

$$y_1 = -2879 + 675.7z_1 + 0.285z_2 + 10.27z_3 + 7.25z_4 + 7.598z_5 + \epsilon$$

with $\sigma = 281.2$ and $R^2 = 0.887$. The residual plot looks ok and all estimates are significant at the 10% level (close to 5% level). The predictive interval is [41.43,1417.62].

(b) The fitted linear regression is

$$y_2 = -2729 + 763.0z_1 + 0.306z_2 + 8.896z_3 + 7.206z_4 + 4.987z_5 + \epsilon$$

with residual standard error 292.5 and $R^2 = 87.6$. The residuals also look reasonable, but coefficient of z_5 is insignificant and can be removed. The 95% predictive interval is [0,1291.23].

(c) The estimates of multivariate linear regression are given below:

```
[1] "LSE of parameters"
[1] "  est    s.d.   t-ratio  prob"
      [,1]    [,2]    [,3]    [,4]
[1,] -2879.478 893.2603 -3.22 0.008108
[2,]  675.651 162.0556  4.17 0.001565
[3,]   0.285   0.0609  4.68 0.000675
[4,]  10.272   4.2549  2.41 0.034358
[5,]   7.251   3.2252  2.25 0.046026
[6,]   7.598   3.8489  1.97 0.074006
[7,] -2728.709 928.8465 -2.94 0.013502
[8,]  763.030 168.5117  4.53 0.000861
[9,]   0.306   0.0633  4.84 0.000521
[10,]  8.896   4.4244  2.01 0.069515
[11,]  7.206   3.3536  2.15 0.054782
[12,]  4.987   4.0022  1.25 0.238622
```

As expected, the results are the same as individual multiple linear regressions. The residuals are also reasonable. The 95% prediction ellipse is centered at $\mathbf{c} = (729.52, 575.73)'$ and given by

$$(\mathbf{Y}_o - \mathbf{c})' \begin{bmatrix} 79091.66 & 69606.95 \\ 69606.95 & 85518.99 \end{bmatrix}^{-1} (\mathbf{Y}_o - \mathbf{c}) \leq 11.157.$$

4. Problem 8.10 of the textbook, p. 473

Answer: (a) The sample covariance matrix is

```

> S=cov(x)
> print(S,digits=3)
      V1      V2      V3      V4      V5
V1 4.33e-04 0.000276 1.59e-04 6.41e-05 8.90e-05
V2 2.76e-04 0.000439 1.80e-04 1.81e-04 1.23e-04
V3 1.59e-04 0.000180 2.24e-04 7.34e-05 6.05e-05
V4 6.41e-05 0.000181 7.34e-05 7.22e-04 5.08e-04
V5 8.90e-05 0.000123 6.05e-05 5.08e-04 7.66e-04

```

Next, perform PCA on the covariance matrix. The result is shown below from R:

```

> m1=princomp(x,cor=F)
> names(m1)
[1] "sdev"      "loadings" "center"    "scale"     "n.obs"     "scores"    "call"
> summary(m1)
Importance of components:
              Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
Standard deviation  0.03680217 0.02635056 0.01585365 0.01188352 0.01085046
Proportion of Variance 0.52926066 0.27133298 0.09821584 0.05518400 0.04600652
Cumulative Proportion 0.52926066 0.80059364 0.89880948 0.95399348 1.00000000

```

```

> m1$loadings

```

```

Loadings:
      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
V1  0.223  0.625  0.326  0.663  0.118
V2  0.307  0.570 -0.250 -0.414 -0.589
V3  0.155  0.345      -0.497  0.780
V4  0.639 -0.248 -0.642  0.309  0.148
V5  0.651 -0.322  0.646 -0.216

```

(b) From the output, the proportion of the total sample variance explained by the first 3 PCs is 90%. From the loading matrix, the first component is a weighted average of the stock returns, the second component represents weighted difference between financial stocks and oil stocks. The third component is a weighted difference between contrasts of the two industries.

(c) The 90% simultaneous Bonferroni C.I.s of λ_1 , λ_2 and λ_3 are given below:

```

Lambda1
[0, 0.04088]
Lambda2

```

```
[0, 0.02844]
Lambda3
[0, 0.01661]
```

where the lower bound is set to zero. (d) Yes, it seems that the stock returns can be summarized using three components.

5. Problem 8.11 of the textbook, p.474.

Answer: The PCA is given below:

```
> x=read.table("T8-5.DAT")
> y=cbind(x[,1:4],x[,5]*10)
> S=cov(y)
> print(S,digits=3)
          V1    V2    V3    V4 x[, 5] * 10
V1      3.397 -1.10  4.306 -2.08    0.272
V2     -1.102  9.67 -1.513 10.95   12.031
V3      4.306 -1.51 55.626 -28.94   -0.436
V4     -2.078 10.95 -28.937 89.07    9.573
x[, 5] * 10 0.272 12.03 -0.436  9.57   31.863
> m2=princomp(y,cor=F,scores=T)
> summary(m2)
Importance of components:
          Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
Standard deviation 10.3197378 6.5140207 5.5456786 2.12666855 1.51966388
Proportion of Variance 0.5709801 0.2275003 0.1648895 0.02424841 0.01238166
Cumulative Proportion 0.5709801 0.7984804 0.9633699 0.98761834 1.00000000
> print(m2$loadings,digits=3)
```

Loadings:

```
          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
V1          0.000 0.000 0.000 -0.556 0.827
V2         -0.119 -0.249 0.261 0.768 0.515
V3          0.480 -0.760 -0.431 0.000 0.000
V4         -0.859 -0.316 -0.394 0.000 0.000
x[, 5] * 10 -0.129 -0.507 0.768 -0.309 -0.203
```

```
> m3=eigen(S)
> m3$values
[1] 108.271939 43.139674 31.267127 4.598098 2.347868
> m3$vectors
          [,1]          [,2]          [,3]          [,4]          [,5]
```

```

[1,]  0.03762881  0.06230915 -0.03997936 -0.55553173  0.82733777
[2,] -0.11892964  0.24930105  0.26052476  0.76839232  0.51517455
[3,]  0.47967291  0.75967654 -0.43064872  0.02807896 -0.08098582
[4,] -0.85891177  0.31639989 -0.39364417 -0.06867379 -0.04989847
[5,] -0.12893518  0.50670427  0.76818907 -0.30895506 -0.20262977

```

```

> names(m2)
[1] "sdev"      "loadings" "center"   "scale"    "n.obs"    "scores"   "call"
> cor(y,m2$scores)
          Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
V1          0.2124404 -0.2220491 -0.1212938 -0.646333476  0.687825655
V2         -0.3978987 -0.5264862  0.4684006  0.529781248  0.253814076
V3          0.6692133 -0.6690038 -0.3228707  0.008072938 -0.016638235
V4         -0.9469980 -0.2202000 -0.2332332 -0.015603522 -0.008101522
x[, 5] * 10 -0.2376782 -0.5895939  0.7609781 -0.117366575 -0.055004739

```

From the output, the covariance matrix is

```

> print(S,digits=3)
          V1    V2    V3    V4 x[, 5] * 10
V1          3.397 -1.10  4.306 -2.08    0.272
V2         -1.102  9.67 -1.513 10.95   12.031
V3          4.306 -1.51 55.626 -28.94   -0.436
V4         -2.078 10.95 -28.937  89.07    9.573
x[, 5] * 10  0.272 12.03 -0.436  9.57   31.863

```

Obviously, only the last row and last column are different from those of Example 8.3; each with a factor of 10.

(b) The eigenvalue-eigenvector pairs are

```

> m3$values
[1] 108.272  43.140  31.267  4.598  2.348
> m3$vectors
      [,1] [,2] [,3] [,4] [,5]
[1,]  0.0376 0.0623 -0.0400 -0.5555  0.8273
[2,] -0.1189 0.2493  0.2605  0.7684  0.5152
[3,]  0.4797 0.7597 -0.4306  0.0281 -0.0810
[4,] -0.8589 0.3164 -0.3936 -0.0687 -0.0499
[5,] -0.1289 0.5067  0.7682 -0.3090 -0.2026

```

(c) The first two PCs explain about 80% of the variance. The correlations are given in the output. They are close, but different from those of Example 8.3 due to the scale change.