Chapter 12

Multicollinearity

Overivew

- Nonsingularity of $X'X$
- Multicollinearity
- Tolerance and Variance Inflation Factor
- Eignevalues and Conditional Indices

12.1 Introduction

- The quality of estimates, as measured by their variances, can be seriously affected if the predictor variables are closely related to each other.
- If the columns of X are linearly dependent, then $(X'X)$ is singular and we cannot estimate $\underline{\beta}$.
- When **near singularity exists**, the variance of estimates can be adversely affected

12.2 Near Singularity

- $X'X$ is considered as near singular if there exists a unit vector $\underline{c}$ such that $\underline{c}'X'X\,\underline{c} = \delta$ is small.

- This will usually result in some of the $\hat{\beta}_i$'s have large variances.

Near Singularity (Continued)

- We may also get some counter intuitive results, especially in signs of the $\widehat{\beta}_i$'s

- Moreover, near singularity can magnify effects of inaccuracies in the elements of X.

- Therefore it is most desirable to detect the presence of near singularity and to identify its causes when it is there.

12.3 Multicollinearity

- **Multicollinearity** is the special case of near singularity where there is a (linear) near relationship between two or more $\underline{x}_{[j]}$'s, where $\underline{x}_{[j]}$'s are the columns of X.
- i.e. the length of $\sum_{j=0}^{p} c_j \underline{x}_{[j]}$ be small with at least two $\underline{x}_{[j]}$'s and corresponding c_j's are not small.
- Since $\sum_{j=0}^{p} c_j \underline{x}_{[j]}$ is affected by the units in which the variables are measured, when assessing the smallness it is desirable to scale X.

Multicollinearity (Continued)

- Instead of considering

$$\underline{y} = X\underline{\beta} + \underline{\epsilon},$$

- we consider

$$\underline{y} = X_{(s)}\underline{\beta}_{(s)} + \underline{\epsilon}$$

where

$$X_{(s)} = XD_{(s)}^{-1} \qquad \underline{\beta}_{(s)} = D_{(s)}\underline{\beta}$$

and

$$D_{(s)} = \text{diag}\left(\|\underline{x}_{[0]}\|, \cdots \cdots, \|\underline{x}_{[p]}\|\right)$$

Multicollinearity (Continued)

- Example

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0.1 & 20 \\ 1 & -0.2 & -10 \\ 1 & 0.1 & 0 \\ 1 & 0 & -10 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix}$$

- Then

$$\left\| \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right\| = \sqrt{4}, \left\| \begin{pmatrix} 0.1 \\ -0.2 \\ 0.1 \\ 0 \end{pmatrix} \right\| = \sqrt{0.06}, \left\| \begin{pmatrix} 20 \\ -10 \\ 0 \\ -10 \end{pmatrix} \right\| = \sqrt{600}$$

Multicollinearity (Continued)

- Example (Continued)

$$D_{(s)} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & \sqrt{0.06} & 0 \\ 0 & 0 & \sqrt{600} \end{pmatrix}$$

and

$$D_{(s)}^{-1} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/\sqrt{0.06} & 0 \\ 0 & 0 & 1/\sqrt{600} \end{pmatrix}$$

Multicollinearity (Continued)

- Example (Continued)

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.1/\sqrt{0.06} & 20/\sqrt{600} \\ 0.5 & -0.2/\sqrt{0.06} & -10/\sqrt{600} \\ 0.5 & 0.1/\sqrt{0.06} & 0 \\ 0.5 & 0 & -10/\sqrt{600} \end{pmatrix} \begin{pmatrix} \beta_{1(s)} \\ \beta_{2(s)} \\ \beta_{3(s)} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix}$$

where $\beta_{0(s)} = 2\beta_0$, $\beta_{1(s)} = \sqrt{0.06}\,\beta_1$ and $\beta_{2(s)} = \sqrt{600}\beta_2$

Multicollinearity (Continued)

- The least squares estimates of $\underline{\beta}_{(s)}$ is given by

$$\underline{\hat{\beta}}_{(s)} = D_{(s)}\underline{\hat{\beta}}$$

and

$$Var\left(\underline{\hat{\beta}}_{(s)}\right) = D_{(s)}Var\left(\underline{\hat{\beta}}\right)D_{(s)}$$

- A consequence of this scaling is that it removes from consideration near singularity caused by a single $\underline{x}_{[j]}$ being of small length.

Multicollinearity (Continued)

- For $\underline{d} = D_{(s)}\,\underline{c}$,

$$\underline{c}'X'X\underline{c} = \underline{c}'D_{(s)}D_{(s)}^{-1}X'XD_{(s)}^{-1}D_{(s)}\underline{c} = \underline{d}'X_{(s)}'X_{(s)}\underline{d}$$

- It can be shown that

$$\underline{d}'X_{(s)}'X_{(s)}\underline{d} \geq \lambda_{\min} \left\|\underline{d}\right\|^2$$

where $\lambda_{\min}$ is the smallest characteristic root of $X_{(s)}'X_{(s)}$.

- Therefore, if multicollinearity is present [($\underline{c}'X'X\underline{c}$) is small with $\left\|\underline{d}\right\|$ not too small], $\lambda_{\min}$ will be small.

Multicollinearity (Continued)

- Conversely, if we have a small eigenvalue of $X'_{(s)}X_{(s)}$, and if $\underline{\gamma}_0$ is the corresponding eigenvector,

- then $\underline{\gamma}_0{}'X'_{(s)}X_{(s)}\underline{\gamma}_0$ is small and it may easily be shown that multicollinearity would be present.

- Since the eigenvectors are mutually orthogonal, each small eigenvalue represents a different near relationship

12.4 Tolerance and Variance Inflation Factor

- Let R_j^2 be the R^2 obtained by regressing x_j against all other x's.
- Hence R_j^2 can be used to assess the degree to which one predictor variable is related to all other predictor variables.

- The **tolerance** $\mathbf{TOL}_j$ is defined as

$$\mathrm{TOL}_j = 1 - R_j^2.$$

- TOL_j is close to one if x_j is not closely related to other predictor variables.

Variance Inflation Factor

- The **variance inflation factor $\mathbf{VIF}_j$** is given by

$$\mathrm{VIF}_j = \mathrm{TOL}_j^{-1}.$$

- A value of VIF_j, close to one indicates no relationship between x_j and other predictors, while a large value indicates presence of multicollinearity.

- How large is large?

12.5 Eigenvalues and Condition Indices

- Since the sum of eigenvalues is equal to the trace, and each diagonal element of $X'_{(s)}X_{(s)}$ is 1,

  therefore

$$\sum_{j=0}^{\lambda} \lambda_j = tr\left(X'_{(s)}X_{(s)}\right) = p + 1$$

where λ_j's are the eigenvalues of $X'_{(s)}X_{(s)}$.

Condition index η_j

$$\eta_j = \sqrt{\lambda_{\max}/\lambda_j}$$

where $\lambda_{\max} = \max_{0 \le j \le p} \lambda_j$.

- It is suggested that an eigenvalue with $\eta_j > 30$ be flagged for further investigation.

12.6 Example

- If we wish to determine which linear combinations of columns of X are causing the multicollinearity, we study the variance of the coefficients of x_j's.
- How this can be done is illustrated in the following example.

Example 1

- Consider a data set with 5 predictor variables, x_1, x_2, x_3, x_4, and x_5.
- The data set "ch12ex1.txt" can be found in the IVLE

Example (continued)

- SAS program

```
proc reg;
  model y = x1 x2 x3 x4 x5 /tol vif collin;
run;
```

- Partial SAS Printout

Parameter Estimates

Variable	DF	Tolerance	Variance Inflation
Intercept	1	.	0
x1	1	0.00213	469.48738
x2	1	0.28215	3.54427
x3	1	0.00189	528.22431
x4	1	0.00876	114.10379
x5	1	0.28273	3.53694

Example (continued)

- Partial SAS Printout

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	-----Proportion of Variation----- Intercept	x1	x2
1	5.88361	1.00000	0.00035009	0.00000226	0.00031808
2	0.09618	7.82142	0.00101	0.00012175	0.03473
3	0.01368	20.73540	0.73465	0.00018846	0.04731
4	0.00628	30.60737	0.00004340	0.00000439	0.87707
5	0.00021047	167.19513	0.26365	0.07563	0.03759
6	0.00003890	388.90457	0.00029382	0.92405	0.00298

Example (continued)

- Partial SAS Printout

Collinearity Diagnostics

	---------Proportion of Variation--------		
Number	x3	x4	x5
1	0.00000196	0.00000862	0.00030262
2	0.00009996	0.00045541	0.02991
3	0.00015342	0.00016611	0.04180
4	6.065364E-8	4.212776E-7	0.88588
5	0.03840	0.97312	0.00415
6	0.96134	0.02625	0.03796

Example (continued)

- R program

```
> model1=lm(y~x1+x2+x3+x4+x5)
> library(car)
> vif(model1)
       x1         x2          x3          x4         x5
469.48737    3.54427   528.22431  114.103786   3.536941
> 1/vif(model1)
       x1         x2          x3          x4          x5
0.0021299  0.2821454  0.0018931  0.0087639   0.2827301
```

Example (continued)

```
> #Condition indices
> library(perturb)
> colldiag(model1)
  Condition
  Index     Variance Decomposition Proportions
              intercept x1    x2    x3    x4    x5
1    1.000  0.000     0.000 0.000 0.000 0.000 0.000
2    7.821  0.001     0.000 0.035 0.000 0.000 0.030
3   20.735  0.735     0.000 0.047 0.000 0.000 0.042
4   30.607  0.000     0.000 0.877 0.000 0.000 0.886
5  167.195  0.264     0.076 0.038 0.038 0.973 0.004
6  388.905  0.000     0.924 0.003 0.961 0.026 0.038
```

Example (continued)

- TOL's for x_1 (0.00213), x_3 (0.00189) and x_4 (0.00876) are small.

- Hence they may be associated and should be tagged for further study.

- The condition indices corresponding to the two smallest eigenvalues are greater 30.

- The proportions of variance indicate that the following possible linear combinations.

Example (continued)

- The last row with the smallest condition index shows a linear combination of x_1 and x_3 as they have the large proportions of variance (0.9241 and 0.9613).

- The only large variance component in the second last row is associated with x_4 (0.9731).

Example (continued)

- However, the 0.0756 in the $Var(\hat{\beta}_1)$ column accounts for most of the variance of $Var(\hat{\beta}_1)$ not accounted for by the smallest eigenvalue and a similar situation exists in the $Var(\hat{\beta}_3)$ column (0.0384).

- Therefore, there is a linear combination involving x_1, x_3 and x_4, which also contributes to small eigenvalues.

Example (continued)

Further study

- Since the TOL for x_3 is the smallest among the 3 potential collinear predictors, we remove it from the full model.

```
> model2=lm(y~x1+x2+x4+x5)
> vif2=vif(model2)
> vif2
       x1        x2        x4        x5
98.995919  3.543295 99.981578  3.391114
> tol2=1/vif2
> tol2
        x1         x2         x4         x5
0.01010143 0.28222315 0.01000184 0.29488835
```

Example (continued)

```
> colldiag(model2)
```

	Condition Index	Variance Decomposition Proportions				
		intercept	x1	x2	x4	x5
1	1.000	0.001	0.000	0.000	0.000	0.000
2	7.822	0.000	0.001	0.038	0.001	0.033
3	19.368	0.755	0.002	0.044	0.001	0.042
4	27.933	0.000	0.000	0.878	0.000	0.924
5	165.018	0.245	0.997	0.040	0.998	0.001

Example (continued)

Further study

- Since the TOL for x_4 is the smallest among the 2 potential collinear predictors, we remove it from the previous model.

```
> model3=lm(y~x1+x2+x5)
> vif3=vif(model3)
> vif3
      x1       x2       x5
1.010440 3.402556 3.388926
> tol3=1/vif3
> tol3
       x1        x2        x5
0.9896682 0.2938967 0.2950787
```

All VIFs are not large

All TOLs are not small

Example (continued)

```
> colldiag(model3)

    Condition
    Index     Variance Decomposition Proportions
                intercept x1        x2        x5
   1     1.000  0.001     0.002     0.001     0.001
   2     8.409  0.011     0.361     0.048     0.040
   3    17.897  0.988     0.635     0.037     0.035
   4    25.002  0.000     0.002     0.914     0.924
```

The largest condition index is larger than 30