Please Click https://canvas.ust.hk and SFQ on the left panel To Fill out End-of-Term Course Survey.

Thanks for your attention!

Chapter 7. Correlated Errors and Collinear Data

Outline

7.1. Correlated Errors: autocorrelation and example

7.1 Correlated Errors: autocorrelation and example

Autocorrelation

One of the standard assumptions in the regression model is that the error terms ε_i and ε_j, associated with the i-th and j-th observations, are **uncorrelated**. Correlation in the error terms suggests that there is additional information in the data that has not been exploited in the current model. When the observations have a **natural sequential order**, the correlation is referred to as **autocorrelation**.

Autocorrelation may occur for several reasons. Adjacent residuals tend to be similar in both temporal and spatial dimensions. Successive residuals in **economic time series** tend to be positively correlated. Large positive errors are followed by other positive errors, and large negative errors are followed by other negative errors. Observations sampled from adjacent experimental plots or areas tend to have residuals that are correlated since they are affected by similar external conditions.

7.1 Correlated Errors: autocorrelation and example

Autocorrelation

The symptoms of autocorrelation may also appear as the result of a variable having been **omitted** from the right-hand side of the regression equation. If successive values of the omitted variable are correlated, the errors from the estimated model will appear to be correlated. When the variable is added to the equation, the apparent problem of autocorrelation disappears. The presence of autocorrelation has several effects on the analysis. These are summarized as follows:

- Least squares estimates of the regression coefficients are unbiased but are not efficient in the sense that they no longer have minimum variance.
- The estimate of σ^2 and the standard errors of the regression coefficients may be seriously understated; that is, from the data the estimated standard errors would be much smaller than they actually are, giving a spurious impression of accuracy
- The confidence intervals and the various tests of significance commonly employed would no longer be strictly valid

The presence of autocorrelation can be a problem of serious concern for the preceding reasons and should not be ignored.

7.1 Correlated Errors: autocorrelation and example

Autocorrelation

We distinguish between two types of autocorrelation and describe methods for dealing with each. The first type is only autocorrelation in appearance. It is due to the **omission of a variable** that should be in the model. Once this variable is uncovered, the autocorrelation problem is resolved. The second type of autocorrelation may be referred to as pure autocorrelation. The methods of correcting for pure autocorrelation involve a transformation of the data.

7.1 Correlated Errors: autocorrelation and example

Example: consumer expenditure and money stock

Table 7.1 gives quarterly data from 1952 to 1956 on consumer expenditure (Y) and the stock of money (X), both measured in billions of current dollars for the United States.

Table 7.1 Consumer Expenditure and Money Stock

Year	Quarter	Consumer Expenditure	Money Stock	Year	Quarter	Consumer Expenditure	Money Stock
1952	1	214.6	159.3	1954	3	238.7	173.9
	2	217.7	161.2		4	243.2	176.1
	3	219.6	162.8	1955	1	249.4	178.0
	4	227.2	164.6		2	254.3	179.1
1953	1	230.9	165.9		3	260.9	180.2
	2	233.3	167.9		4	263.3	181.2
	3	234.1	168.3	1956	1	265.6	181.6
	4	232.3	169.7		2	268.2	182.5
1954	1	233.7	170.5		3	270.4	183.3
	2	236.5	171.6		4	275.6	184.3

7.1 Correlated Errors: autocorrelation and example

Example: consumer expenditure and money stock

A simplified version of the quantity theory of money suggests a model given by

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \tag{7.1}$$

where β_0 and β_1 are constants, ε_t the error term. Economists are interested in estimating β_1 and its standard error; β_1 is called the *multiplier* and has crucial importance as an instrument in fiscal and monetary policy. Since the observations are ordered in time, it is reasonable to expect that autocorrelation may be present. A summary of the regression results is given in Table 7.2

Table 7.2 Results When Consumer Expenditure is Regressed on Money Stock, X

Variable	Coefficient	s.e.	t-Test	p-value
Constant	-154.72	19.850	-7.79	< 0.0001
X	2.30	0.115	20.10	< 0.0001
$n = 20$	$R^2 = 0.957$	$R_a^2 = 0.955$	$\hat{\sigma} = 3.983$	df = 18

7.1 Correlated Errors: autocorrelation and example

Example: consumer expenditure and money stock

The regression coefficients are significant; the standard error of the slope coefficient is 0.115. For a unit change in the money supply the 95% confidence interval for the change in the aggregate consumer expenditure would be $2.30 \pm 2.10 \times 0.115 = (2.06, 2.54)$. The value of R^2 indicates that roughly 96% of the variation in the consumer expenditure can be accounted for by the variation in money stock. The analysis would be complete if the basic regression assumptions were valid. To check on the model assumption, we examine the residuals. If there are indications that autocorrelation is present, the model should be reestimated after eliminating the autocorrelation.

For time series data a useful plot for analysis is the index plot (plot of the standardized residuals versus time). The graph is given in Figure 7.1. The pattern of residuals is revealing and is characteristic of situations where the errors are correlated. Residuals of the same sign occur in clusters or bunches. The characteristic pattern would be that several successive residuals are positive, the next several are negative, and so on. From Figure 7.1 we see that the first seven residuals are positive, the next seven negative, and the last six positive. This pattern suggests that the error terms in the model are correlated and some additional analysis is required.

Figure 7.1

7.1 Correlated Errors: autocorrelation and example

Example: consumer expenditure and money stock

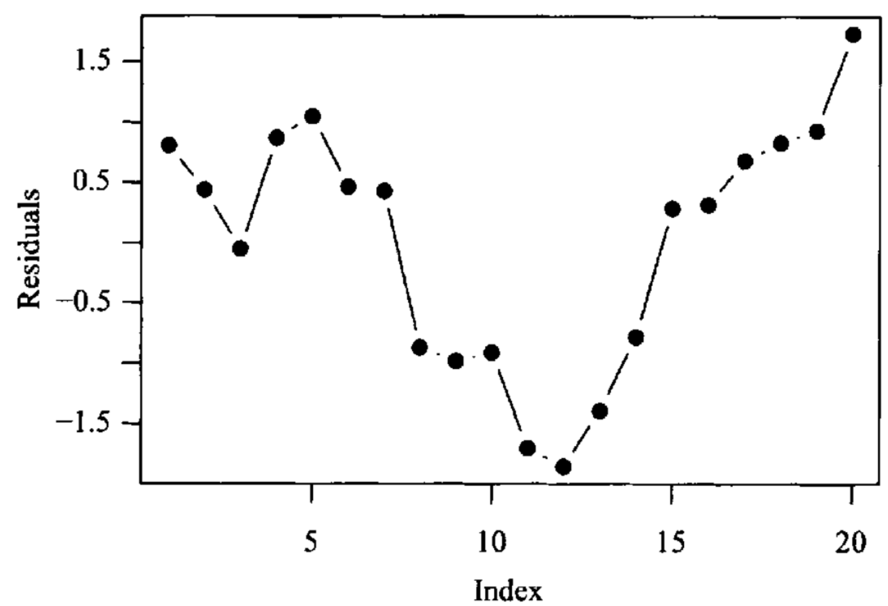


Figure 7.1 Index plot of the standardized residuals.

Run is basically a sequence of one symbol such as + or -.

This visual impression can be formally confirmed by counting the number of runs in a plot of the signs of the residuals, the residuals taken in the order of the observations. These types of plots are called **sequence plots**. In our present example the sequence plot of the signs of the residuals is

$$+ + - + + + + - - - - - - - + + + + + +$$

and it indicates five runs.

7.1 Correlated Errors: autocorrelation and example

Example: consumer expenditure and money stock

With n_1 residuals positive and n_2 residuals negative, under the hypothesis of randomness the expected number of runs μ and its variance σ^2 would be

$$\mu = \frac{2n_1n_2}{n_1+n_2} + 1$$
$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

In our case $n_1 = 12, n_2 = 8$, giving the expected number of runs to be 10.6 and a standard deviation of 2.08. The observed number of runs is 5. The p-value is 0.0072, indicating a significant departure from randomness. This formal **runs test** procedure merely confirms the conclusion arrived at visually that there is a pattern in the residuals.

Run Test Statistic $= \frac{\#\text{ of runs} - \mu}{\sigma}$. Under H_0**: the residuals are uncorrelated random**, this test statistic follows a distribution $N(0,1)$

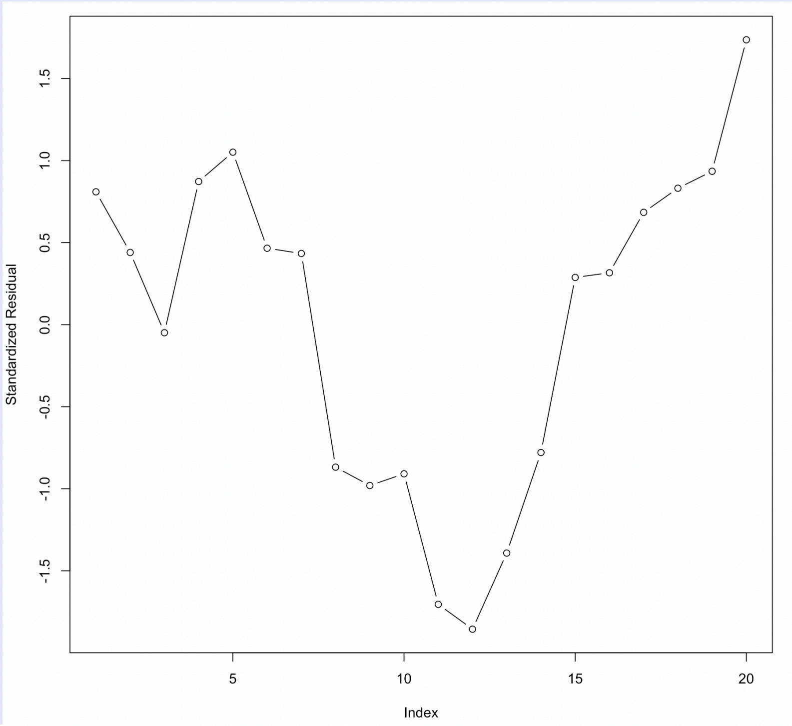
Many computer packages now have the **runs test** as an available option. This approximate runs test for confirmation can therefore be easily executed. The runs test as we have described it should not, however, be used for small values of n_1 and n_2 (**less than 10**). For small values of n_1 and n_2 one needs exact tables of probability to judge significance.

7.1 Correlated Errors: autocorrelation and example

Use R for last example

```
> ########################################  Example of Consumer Expenditure and Stock
> CoE_Stk<-read.table('data/P211.txt',header=TRUE)   ## read the data
> mod1<-lm(Expenditure~Stock,data=CoE_Stk)
> summary(mod1)

Call:
lm(formula = Expenditure ~ Stock, data = CoE_Stk)

Residuals:
   Min     1Q Median     3Q    Max 
-7.176 -3.396  1.396  2.928  6.361 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -154.7192    19.8500  -7.794 3.54e-07 ***
Stock          2.3004     0.1146  20.080 8.99e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.983 on 18 degrees of freedom
Multiple R-squared:  0.9573,    Adjusted R-squared:  0.9549 
F-statistic: 403.2 on 1 and 18 DF,  p-value: 8.988e-14
```

```
###  Standardized residual versus Index plot
pii=hatvalues(mod1)
plot(1:20,mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),xlab="Index",ylab="Standardized Residual",type="b")
```



7.1 Correlated Errors: autocorrelation and example

Use R for last example

```
> ###  compute the statistics of runs test
> n1<-length(which(mod1$residuals >0))    ## number of positive residuals
> n2<-length(which(mod1$residuals <0))    ## number of negative residuals
> mu<-2*n1*n2/(n1+n2)+1
> ssq<-2*n1*n2*(2*n1*n2-n1-n2)/(n1+n2)^2*1/(n1+n2-1)
> num_runs<-5
> run_stat<-(num_runs-mu)/sqrt(ssq)
> run_stat
[1] -2.686458
```

```
> ### use built-in Runs test
> library(snpar)   # or use the library library(randtests)
> runs.test(mod1$residuals,threshold=0)    ## by default the threshold is set as the median of the given data

	Runs Test

data:  mod1$residuals
statistic = -2.6865, runs = 5, n1 = 12, n2 = 8, n = 20, p-value = 0.007221
alternative hypothesis: nonrandomness
```

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Durbin-Watson Statistic

The Durbin-Watson statistic is the basis of a popular test of autocorrelation in regression analysis. The test is based on the assumption that successive errors are correlated, namely

$$\varepsilon_t = \rho\varepsilon_{t-1} + \omega_t, \quad |\rho| < 1 \tag{7.2}$$

where ρ is the correlation coefficient between ε_t and ε_{t-1}, and ω_t is normally independently distributed with zero mean and constant variance. In this case, the errors are said to have **first-order autoregressive structure** or **first-order autocorrelation**. In most situations the error ε_t may have a much more complex correlation structure. The first-order dependency structure, given in (7.2), is taken as a simple approximation to the actual error structure.

The Durbin-Watson statistic is defined as

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2},$$

where e_i is the i-th ordinary least squares (OLS) residual. The statistics d is used for testing the null hypothesis $H_0 : \rho = 0$ against an alternative $H_1 : \rho > 0$. Note that when $\rho = 0$ in (7.2), the ε's are uncorrelated.

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Durbin-Watson Statistic

Since ρ is unknown, we estimate the parameter ρ by $\hat{\rho}$, where

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}. \tag{7.3}$$

An approximate relationship between d and $\hat{\rho}$ is

$$d \doteq 2(1-\hat{\rho}),$$

($\doteq$ means approximately equal to) showing that d has a range of 0 to 4. Since $\hat{\rho}$ is an estimate of ρ, it is clear that d is close to 2 when $\rho = 0$ and near to zero when $\rho = 1$. The closer the sample value of d to 2, the firmer the evidence that there is no autocorrelation present in the error. Evidence of autocorrelation is indicated by the deviation of d from 2. The formal test for positive autocorrelation operates as follows: Calculate the sample statistic d. Then, if

1. $d < d_L$, reject H_0.
2. $d > d_U$, do not reject H_0.
3. $d_L < d < d_U$, the test is inconclusive.

The values of (d_L, d_U) for different percentage points have been tabulated by Durbin and Watson (1951).

Table

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Durbin-Watson Statistic

Table A.6 Distribution of Durbin-Watson Statistic d: The 5% Significance Points of d_L and d_U (p is the Number of Predictor Variables)

	$p=1$		$p=2$		$p=3$		$p=4$		$p=5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Source: Durbin and Watson (1951).

Table A.7 Distribution of Durbin-Watson Statistic d: The 1% Significance Points of d_L and d_U (p is the Number of Predictor Variables)

	$p=1$		$p=2$		$p=3$		$p=4$		$p=5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Source: Durbin and Watson (1951).

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Durbin-Watson Statistic

Tests for negative autocorrelation are seldom performed. If, however, a test is desired, then instead of working with d, one works with $4 - d$ and follows the same procedure as for the testing of positive autocorrelation.

In our Money Stock and Consumer Expenditure data, the value of d is 0.328. From Table A.6, with $n = 20, p = 1$ (the number of predictors), and a significance level of 0.05, we have $d_L = 1.20$ and $d_U = 1.41$. Since $d < d_L$, we conclude that the value of d is significant at the 5% level and H_0 is rejected, showing that autocorrelation is present. This essentially reconfirms our earlier conclusion, which was arrived at by looking at the index plot of the residuals.

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Durbin-Watson Statistic

If d had been larger than $d_U = 1.41$, autocorrelation would not be a problem and no further analysis is needed. When $d_L < d < d_U$, additional analysis of the equation is optional. We suggest that in cases where the Durbin-Watson statistic lies in the inconclusive region, reestimate the equation using the methods described below to see if any major changes occur.

As pointed out earlier, the presence of correlated errors distorts estimates of standard errors, confidence intervals, and statistical tests, and therefore we should reestimate the equation. When autocorrelated errors are indicated, we work with transformed variables.

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Removal of autocorrelation by Transformation

When the residual plots and Durbin-Watson statistic indicate the presence of correlated errors, the estimated regression equation should be refitted taking the autocorrelation into account. One method for adjusting the model is the use of a transformation that involves the unknown autocorrelation parameter, ρ. The introduction of ρ causes the model to be nonlinear. The direct application of least squares is not possible. However, there are a number of procedures that may be used to circumvent the nonlinearity.

From model (7.1), ε_t and ε_{t-1} can be expressed as

$$\begin{aligned} \varepsilon_t &= y_t - \beta_0 - \beta_1 x_t, \\ \varepsilon_{t-1} &= y_{t-1} - \beta_0 - \beta_1 x_{t-1}. \end{aligned}$$

Substituting these in (7.2), we obtain

$$y_t - \beta_0 - \beta_1 x_t = \rho(y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + \omega_t.$$

Rearranging terms in the above equation, we get

$$\begin{array}{ccccccc} y_t - \rho y_{t-1} & = & \beta_0(1-\rho) & + & \beta_1(x_t - \rho x_{t-1}) & + & \omega_t, \\ y_t^* & = & \beta_0^* & + & \beta_1^* \quad x_t^* & + & \omega_t, \end{array} \tag{7.4}$$

where

$$\begin{aligned} y_t^* &= y_t - \rho y_{t-1}, \\ x_t^* &= x_t - \rho x_{t-1}, \\ \beta_0^* &= \beta_0(1-\rho), \\ \beta_1^* &= \beta_1. \end{aligned}$$

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Removal of autocorrelation by Transformation

Since the ω's are uncorrelated, (7.4) represents a linear model with uncorrelated errors. This suggests that we run an ordinary least squares regression using y_t^* as a response variable and x_t^* as a predictor. The estimates of the parameters in the original equations are

$$\hat{\beta}_0 = \frac{\hat{\beta}_0^*}{1-\hat{\rho}} \quad \text{and} \quad \hat{\beta}_1 = \hat{\beta}_1^* \tag{7.5}$$

Therefore, when the errors in model (7.1) have an autoregressive structure as given in (7.2), we can transform both sides of the equation and obtain transformed variables which satisfy the assumption of uncorrelated errors.

The value of ρ is unknown and has to be estimated from the data. Cochrane and Orcutt have proposed an iterative procedure. The procedure operates as follows:

1. Compute the OLS estimates of β_0 and β_1 by fitting model (7.1) to the data.
2. Calculate the residuals and, from the residuals, estimate ρ using (7.3).
3. Fit the equation given in (7.4) using the variables $y_t - \hat{\rho} y_{t-1}$ and $x_t - \hat{\rho} x_{t-1}$ as response and predictor variables, respectively, and obtain $\hat{\beta}_0$ and $\hat{\beta}_1$ using (7.5).
4. Examine the residuals of the newly fitted equation. If the new residuals continue to show autocorrelation, repeat the entire procedure using the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ as estimates of β_0 and β_1 instead of the original least squares estimates. On the other hand, if the new residuals show no autocorrelation, the procedure is terminated and the fitted equation for the original data is

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t.$$

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Removal of autocorrelation by Transformation

As a practical rule we suggest that if the first application of the Cochrane- Orcutt procedure does not yield non-autocorrelated residuals, one should look for alternative methods of removing autocorrelation. We apply the Cochrane-Orcutt procedure to the data given in Table 7.1.

The d value for the original data is 0.328, which is highly significant. The value of $\hat{\rho}$ is 0.751. On fitting the regression equation to the variables $(y_t - 0.751y_{t-1})$ and $(x_t - 0.751x_{t-1})$, we have a d value of 1.43. The value of d_U for $n = 19$ and $p = 1$ is 1.40 at the 5% level. Consequently, $H_0 : \rho = 0$ is not rejected. The fitted equation is

$$\hat{y}_t^* = -53.70 + 2.64x_t^*,$$

which, using (7.5), the fitted equation in terms of the original dataset is

$$\hat{y}_t = -215.31 + 2.64x_t$$

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Removal of autocorrelation by Transformation

How is this obtained?

The estimated standard error for the slope is 0.307, as opposed to the least squares estimate of the original equation, which was $y_t = -154.7 + 2.3x_t$ with a standard error for the slope of 0.115. The newly estimated standard error is larger by a factor of almost 3. The residual plots for the fitted equation of the transformed variables are shown in Figure 7.2. The residual plots show less clustering of the adjacent residuals by sign, and the Cochrane-Orcutt procedure has worked to our advantage.

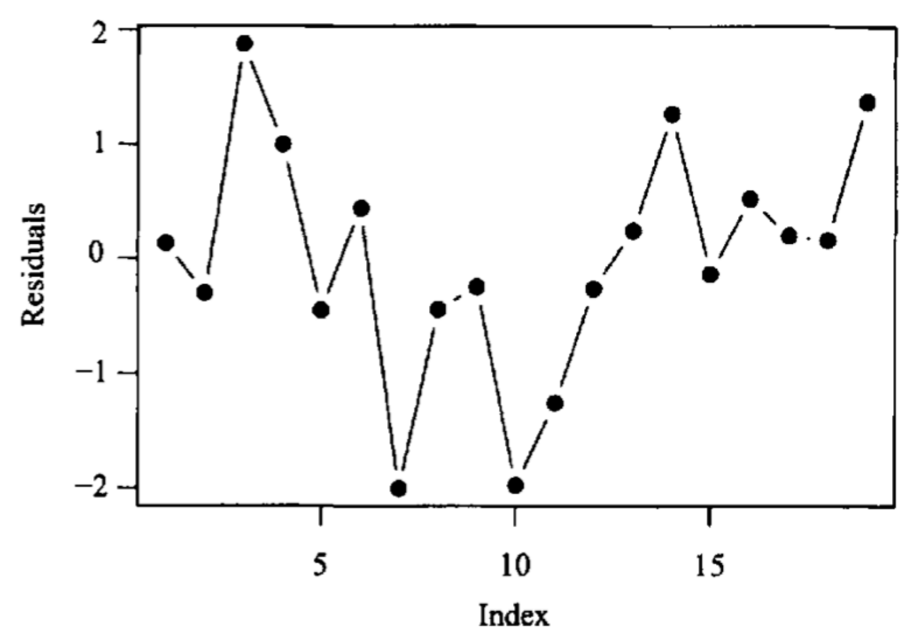


Figure 7.2 Index plot of standardized residuals after one iteration of the Cochrane-Orcutt method.

7.2 Correlated Errors: Durbin-Watson statistic and removal of autocorrelation

Use R for last example

```
> ############################################################  Implement Cochrane and Orcutt method
> CoE_Stk<-read.table('data/P211.txt',header=TRUE)   ## read the data
> mod1<-lm(Expenditure~Stock,data=CoE_Stk)
> hatrho<-sum(mod1$residuals[-1] * mod1$residuals[-20])/sum(mod1$residuals^2)    #### calculate hatrho in the first iteration
> hatrho
[1] 0.7506122
```

```
> ###### transformation data for the 2nd iteration
> CoE_Stk_new<-data.frame(y=CoE_Stk$Expenditure[-1]-CoE_Stk$Expenditure[-20] * hatrho, x<-CoE_Stk$Stock[-1]-CoE_Stk$Stock[-20] * hatrho)
> mod2<-lm(y~x, data=CoE_Stk_new)
> summary(mod2)

Call:
lm(formula = y ~ x, data = CoE_Stk_new)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.3737 -0.7856  0.2747  1.0408  3.9786 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -53.6959    13.6164  -3.943  0.00105 ** 
x             2.6434     0.3069   8.614 1.32e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.263 on 17 degrees of freedom
Multiple R-squared:  0.8136,	Adjusted R-squared:  0.8026 
F-statistic:  74.2 on 1 and 17 DF,  p-value: 1.315e-07
```

```
> hatrho<-sum(mod2$residuals[-1] * mod2$residuals[-19])/sum(mod2$residuals^2)    #### calculate hatrho in the 2nd iteration
> hatrho
[1] 0.2399883
```

7.3. Collinearity: Effects of Collinearity on Inference

7.3 Collinearity: Effects of Collinearity on Inference

Introduction

Interpretation of the multiple regression equation depends implicitly on the assump- tion that the predictor variables are not strongly interrelated. It is usual to interpret a regression coefficient as measuring the change in the response variable when the corresponding predictor variable is increased by one unit and all other predictor variables are held constant. This interpretation may not be valid if there are strong linear relationships among the predictor variables. It is always conceptually possi- ble to increase the value of one variable in an estimated regression equation while holding the others constant. However, there may be no information about the result of such a manipulation in the estimation data. Moreover, it may be impossible to change one variable while holding all others constant in the process being studied. When these conditions exist, simple interpretation of the regression coefficient as a marginal effect is lost.

Recall that Multiple Linear Regression requires $\mathbf{X}'\mathbf{X}$ is invertible

7.3 Collinearity: Effects of Collinearity on Inference

Introduction

When there is a complete absence of linear relationship among the predictor variables, they are said to be **orthogonal**. In most regression applications the predictor variables are not orthogonal. Usually, the lack of orthogonality is not serious enough to affect the analysis. However, in some situations the predictor variables are so strongly interrelated that the regression results are ambiguous. Typically, it is impossible to estimate the unique effects of individual variables in the regression equation. The estimated values of the coefficients are very **sensitive** to slight changes in the data and to the addition or deletion of variables in the equation. The regression coefficients have large sampling errors, which affect both inference and forecasting that is based on the regression model

The condition of severe nonorthogonality (that is, the existence of strong linear relationships among the predictor variables) is also referred to as the problem of **collinear data**, **collinearity**, or **multicollinearity**. The problem can be difficult to detect. It is not a specification error that may be uncovered by exploring regression residual. In fact, collinearity is not a modeling error. It is a condition of deficient data. In any event, it is important to know when collinearity is present and to be aware of its possible consequences. It is recommended that one should be very cautious about any and all substantive conclusions based on a regression analysis in the presence of collinearity

7.3 Collinearity: Effects of Collinearity on Inference

Introduction

1. How does collinearity affect statistical inference and forecasting?
2. How can collinearity be detected?
3. What can be done to resolve the difficulties associated with collinearity?

7.3 Collinearity: Effects of Collinearity on Inference

Example

In conjunction with the Civil Rights Act of 1964, the Congress of the United States ordered a survey "concerning the lack of availability of equal educational opportunities for individuals by reason of race, color, religion or national origin in public educational institutions...." Data were collected from a cross section of school districts throughout the country. In addition to reporting summary statistics riables such as level of student achievement and school facilities, regression analysis was used to try to establish factors that are the most important determinants of achievement. The data for this example consist of measurements taken in 1965 for 70 schools selected at random. The data consist of variables that measure student achievements (ACHV), faculty credentials (FAM), the influence of their peer group in the school (PEER), and school facilities (SCHOOL). The objective is to evaluate the effect of school inputs on achievement.

Assume that an acceptable index has been developed to measure those aspects of the school environment that would be expected to affect achievement. The index includes evaluations of the physical plant, teaching materials, special programs, training and motivation of the faculty, and so on. Achievement can be measured by using an index constructed from standardized test scores. There are also other variables that may affect the relationship between school inputs and achievement. Students' performances may be affected by their home environments and the in- fluence of their peer group in the school. These variables must be accounted for in the analysis before the effect of school inputs can be evaluated. We assume that indexes have been constructed for these variables that are satisfactory for our purposes. The data are given in Tables 7 .1 and 7 .2.

Table 7.3, 7.4.

7.3 Collinearity: Effects of Collinearity on Inference

Example

Table 7.3 First 50 Observations of the Equal Educational Opportunity (EEO) Data; Standardized Indexes

Row	ACHV	FAM	PEER	SCHOOL
1	−0.43148	0.60814	0.03509	0.16607
2	0.79969	0.79369	0.47924	0.53356
3	−0.92467	−0.82630	−0.61951	−0.78635
4	−2.19081	−1.25310	−1.21675	−1.04076
5	−2.84818	0.17399	−0.18517	0.14229
6	−0.66233	0.20246	0.12764	0.27311
7	2.63674	0.24184	−0.09022	0.04967
8	2.35847	0.59421	0.21750	0.51876
9	−0.91305	−0.61561	−0.48971	−0.63219
10	0.59445	0.99391	0.62228	0.93368
11	1.21073	1.21721	1.00627	1.17381
12	1.87164	0.41436	0.71103	0.58978
13	−0.10178	0.83782	0.74281	0.72154
14	−2.87949	−0.75512	−0.64411	−0.56986
15	3.92590	−0.37407	−0.13787	−0.21770
16	4.35084	1.40353	1.14085	1.37147
17	1.57922	1.64194	1.29229	1.40269
18	3.95689	−0.31304	−0.07980	−0.21455
19	1.09275	1.28525	1.22441	1.20428
20	−0.62389	−1.51938	−1.27565	−1.36598
21	−0.63654	−0.38224	−0.05353	−0.35560
22	−2.02659	−0.19186	−0.42605	−0.53718
23	−1.46692	1.27649	0.81427	0.91967
24	3.15078	0.52310	0.30720	0.47231
25	−2.18938	−1.59810	−1.01572	−1.48315
26	1.91715	0.77914	0.87771	0.76496
27	−2.71428	−1.04745	−0.77536	−0.91397
28	−6.59852	−1.63217	−1.47709	−1.71347
29	0.65101	0.44328	0.60956	0.32833
30	−0.13772	−0.24972	0.07876	−0.17216
31	−2.43959	−0.33480	−0.39314	−0.37198
32	−3.27802	−0.20680	−0.13936	0.05626
33	−2.48058	−1.99375	−1.69587	−1.87838
34	1.88639	0.66475	0.79670	0.69865
35	5.06459	−0.27977	0.10817	−0.26450
36	1.96335	−0.43990	−0.66022	−0.58490
37	0.26274	−0.05334	−0.02396	−0.16795
38	−2.94593	−2.06699	−1.31832	−1.72082
39	−1.38628	−1.02560	−1.15858	−1.19420
40	−0.20797	0.45847	0.21555	0.31347
41	−1.07820	0.93979	0.63454	0.69907
42	−1.66386	−0.93238	−0.95216	−1.02725
43	0.58117	−0.35988	−0.30693	−0.46232
44	1.37447	−0.00518	0.35985	0.02485
45	−2.82687	−0.18892	−0.07959	0.01704
46	3.86363	0.87271	0.47644	0.57036
47	−2.64141	−2.06993	−1.82915	−2.16738
48	0.05387	0.32143	−0.25961	0.21632
49	0.50763	−1.42382	−0.77620	−1.07473
50	0.64347	−0.07852	−0.21347	−0.11750

Table 7.4 Last 20 Observations of Equal Educational Opportunity (EEO) Data; Standardized Indexes

Row	ACHV	FAM	PEER	SCHOOL
51	2.49414	−0.14925	−0.03192	−0.36598
52	0.61955	0.52666	0.79149	0.71369
53	0.61745	−1.49102	−1.02073	−1.38103
54	−1.00743	−0.94757	−1.28991	−1.24799
55	−0.37469	0.24550	0.83794	0.59596
56	−2.52824	−0.41630	−0.60312	−0.34951
57	0.02372	1.38143	1.54542	1.59429
58	2.51077	1.03806	0.91637	0.97602
59	−4.22716	−0.88639	−0.47652	−0.77693
60	1.96847	1.08655	0.65700	0.89401
61	1.25668	−1.95142	−1.94199	−1.89645
62	−0.16848	2.83384	2.47398	2.79222
63	−0.34158	1.86753	1.55229	1.80057
64	−2.23973	−1.11172	−0.69732	−0.80197
65	3.62654	1.41958	1.11481	1.24558
66	0.97034	0.53940	0.16182	0.33477
67	3.16093	0.22491	0.74800	0.66182
68	−1.90801	1.48244	1.47079	1.54283
69	0.64598	2.05425	1.80369	1.90066
70	−1.75915	1.24058	0.64484	0.87372

7.3 Collinearity: Effects of Collinearity on Inference

Example

Adjustment for the two basic variables (achievement and school) can be accomplished by using the regression model

$$\text{ACHV} = \beta_0 + \beta_1\text{FAM} + \beta_2\text{PEER} + \beta_3\text{SCHOOL} + \varepsilon. \qquad (7.6)$$

The contribution of the school variable can be tested using the t-value for β_3. Recall that the t-value for β_3 tests whether SCHOOL is necessary in the equation when FAM and PEER are already included. Effectively, the model above is being compared to

$$\text{ACHV} - \beta_1\text{FAM} - \beta_2\text{PEER} = \beta_0 + \beta_3\text{SCHOOL} + \varepsilon, \qquad (7.7)$$

that is, the contribution of the school variable is being evaluated after adjustment for FAM and PEER. Another view of the adjustment notion is obtained by noting that the left-hand side of (7.7) is an adjusted achievement index where adjustment is accomplished by subtracting the linear contributions of FAM and PEER. The equation is in the form of a regression of the adjusted achievement score on the SCHOOL variable. This representation is used only for the sake of interpretation. The estimated β's are obtained from the original model given in (7.6). The regression results are summarized in Table 7.5 and a plot of the residuals against the predicted values of ACHV appears as Figure 7.3.

Table 7.5, Figure 7.3

7.3 Collinearity: Effects of Collinearity on Inference

Example

Table 7.5 EEO Data: Regression Results

ANOVA Table

Source	Sum of Squares	df	Mean Square	F-Test
Regression	73.506	3	24.502	5.72
Residuals	282.873	66	4.286	

Coefficients Table

Variable	Coefficient	s.e.	t-Test	p-value
Constant	−0.070	0.251	−0.28	0.7810
FAM	1.101	1.411	0.78	0.4378
PEER	2.322	1.481	1.57	0.1218
SCHOOL	−2.281	2.220	−1.03	0.3080
$n = 70$	$R^2 = 0.206$	$R_a^2 = 0.170$	$\hat{\sigma} = 2.07$	df = 66

$$\text{ACHV} = \beta_0 + \beta_1 \text{FAM} + \beta_2 \text{PEER} + \beta_3 \text{SCHOOL} + \varepsilon$$

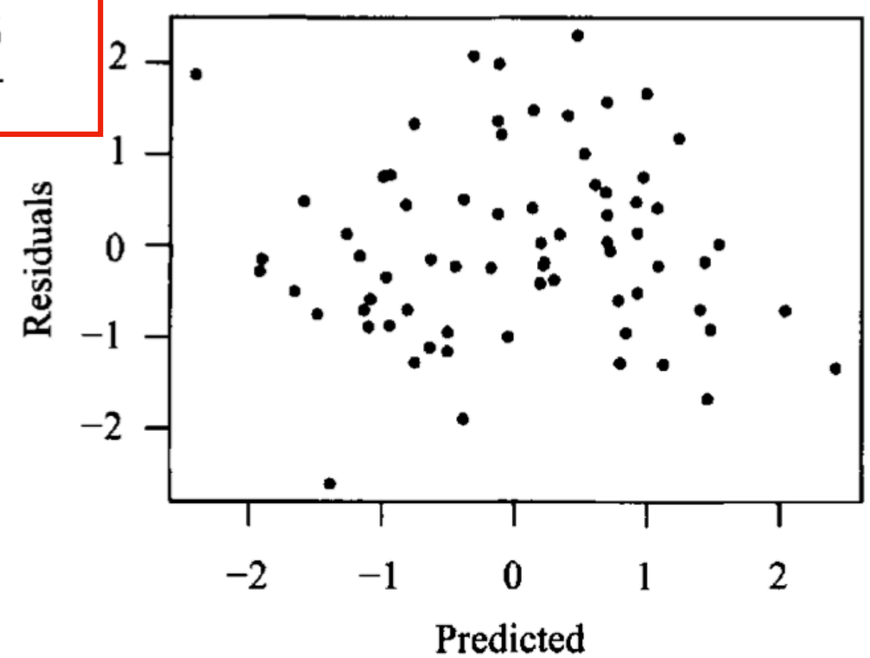


Figure 7.3 Standardized residuals against fitted values of ACHV.

7.3 Collinearity: Effects of Collinearity on Inference

Example

Checking first the residual plot we see that there are no glaring indications of misspecification. The point located in the lower left of the graph has a residual value that is about 2.5 standard deviations from the mean of zero and should possibly be looked at more closely. However, when it is deleted from the sample, the regression results show almost no change. Therefore, the observation has been retained in the analysis.

From Table 7.5 we see that about 20% of the variation in achievement score is accounted for by the three predictors jointly ($R^2 = 0.206$). The F-value is 5.72 based on 3 and 66 degrees of freedom and is significant at better than the 0.01 level. Therefore, even though the total explained variation is estimated at only 20%, it is accepted that FAM, PEER, and SCHOOL are valid predictor variables. However, the individual t-values are all small. In total, the summary statistics say that the three predictors taken together are important but the t-values indicate that none of the variables individually are significant. It follows that anyone predictor may be deleted from the model provided the other two are retained.

7.3 Collinearity: Effects of Collinearity on Inference

Example

These results are typical of a situation where extreme **collinearit**y is present. The predictor variables are so highly correlated that each one may serve as a **proxy** for the others in the regression equation without affecting the total explanatory power. The low t-values confirm that anyone of the predictor variables may be dropped from the equation. Hence the regression analysis has failed to provide any information for evaluating the importance of school inputs on achievement. The culprit is clearly collinearity. The **pairwise correlation coefficients** of the three predictor variables and the corresponding scatter plots (Figure 7.4), all show strong linear relationships among all pairs of predictor variables. All pairwise correlation coefficients are high. In all scatter plots, all the observations lie close to a straight line.

Typical indications of collinearity by regression outputs:

- Case 1: **High** R^2, but **high** p-value for the coefficient estimates
- Case 2: **Significant** F statistic (i.e. model seems valid), but **high** p-value for the coefficient estimates
- Case 3: If the estimated coefficient is contrary to prior expectation (which people are confident to be correct)

Figure 7.4.

7.3 Collinearity: Effects of Collinearity on Inference

Example

Indication of collinearity by correlation coefficients between predictor variables

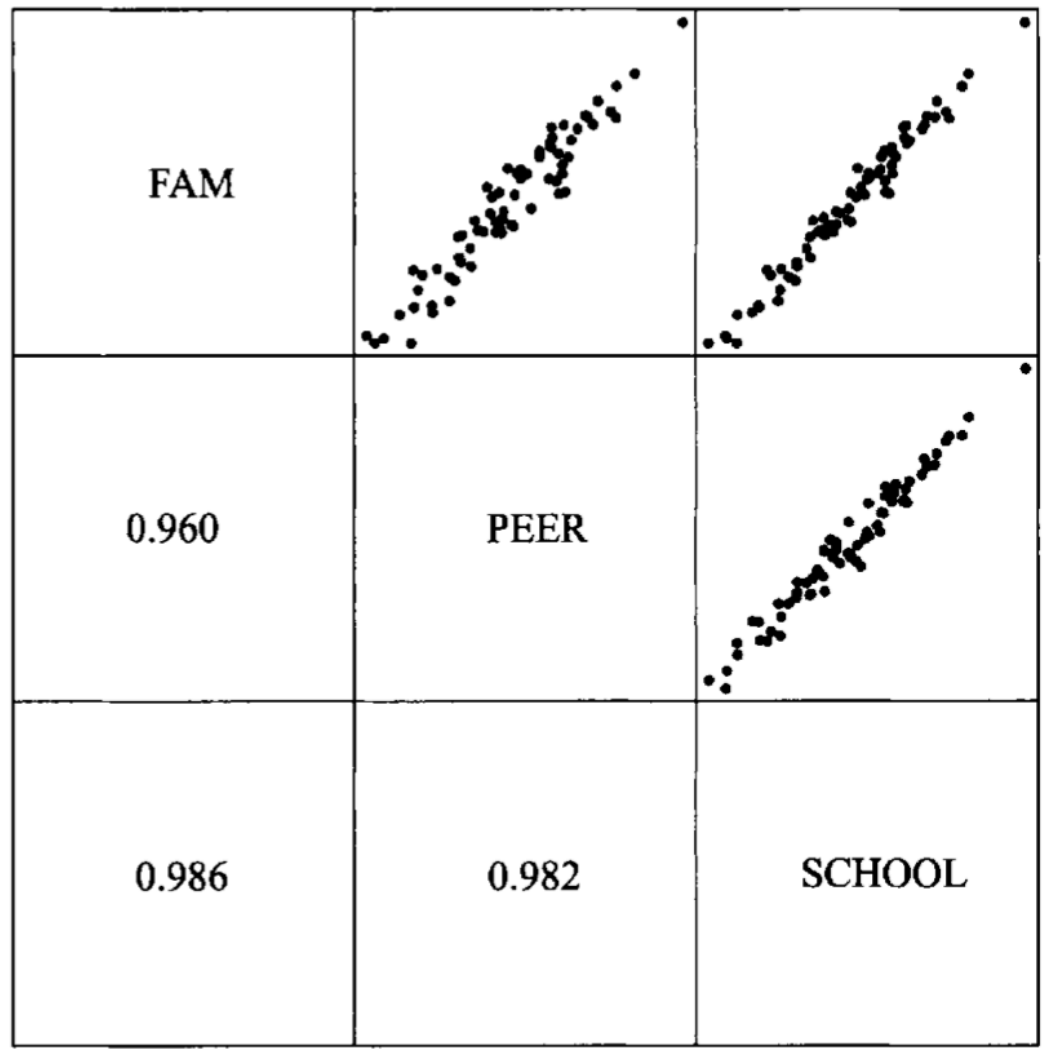


Figure 7.4 Pairwise scatter plots of the three predictor variables FAM, PEER, and SCHOOL and the corresponding pairwise correlation coefficients.

7.3 Collinearity: Effects of Collinearity on Inference

Example

Collinearity in this instance could have been expected. It is the nature of these three variables that each is determined by and helps to determine the others. It is not unreasonable to conclude that there are not three variables but in fact only one. Unfortunately, that conclusion does not help to answer the original question about the effects of school facilities on achievement. There remain two possibilities. First, collinearity may be present because the sample data are deficient, but can proved with additional observations. Second, collinearity may be present because the interrelationships among the variables are an inherent characteristic of the process under investigation. Both situations are discussed in the following paragraphs.

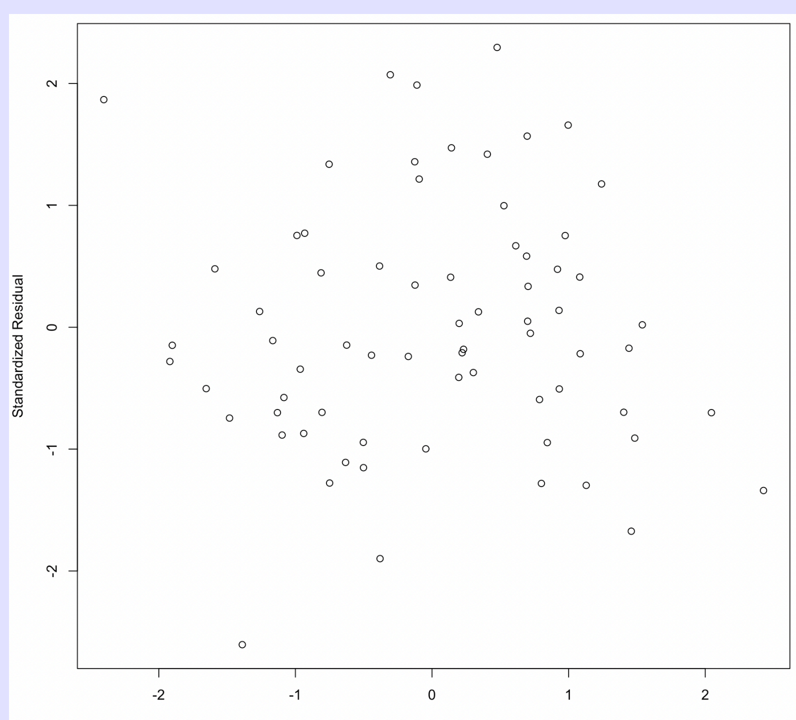
In the first case the sample should have been selected to ensure that the correlations between the predictor variables were not large. For example, in the scatter plot ofFAM versus SCHOOL (the graph in the top right corner in Figure 7.4), there are no schools in the sample with values in the upper-left or lower-right regions of the graph. Hence there is no information in the sample on achievement when the value of FAM is high and SCHOOL is low, or FAM is low and SCHOOL is high. But it is only with data collected under these two conditions that the individual effects of FAM and SCHOOL on ACHV can be determined. For example, assume that there were some observations in the upper-left quadrant of the graph. Then it would at least be possible to compare average ACHV for low and high values of SCHOOL when FAM is held constant.

7.3 Collinearity: Effects of Collinearity on Inference

Use R for last example

```
> ##########################################################################  Example on Equal Education Opportunity Data
> EEO_dat<-read.table('data/P236.txt',header=TRUE)   ## read the data
> mod1<-lm(ACHV~.,data=EEO_dat)
> summary(mod1)

Call:
lm(formula = ACHV ~ ., data = EEO_dat)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2096 -1.3934 -0.2947  1.1415  4.5881

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.06996    0.25064  -0.279    0.781
FAM          1.10126    1.41056   0.781    0.438
PEER         2.32206    1.48129   1.568    0.122
SCHOOL      -2.28100    2.22045  -1.027    0.308

Residual standard error: 2.07 on 66 degrees of freedom
Multiple R-squared:  0.2063,    Adjusted R-squared:  0.1702
F-statistic: 5.717 on 3 and 66 DF,  p-value: 0.001535
```

```
pii=hatvalues(mod1)
plot(mod1$fitted.values,mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),xlab="Fitted Value",ylab="Standardized Residual",type="p")
```



7.3 Collinearity: Effects of Collinearity on Inference

Example

Since there are three predictor variables in the model, then there are eight distinct combinations of data that should be included in the sample. Using + to represent a value above the average and − to represent a value below the average, the eight possibilities are represented in Table 7.6.

Table 7.6 Data Combinations for Three Predictor Variables

	Variable		
Combination	FAM	PEER	SCHOOL
1	+	+	+
2	+	+	−
3	+	−	+
4	−	+	+
5	+	−	−
6	−	+	−
7	−	−	+
8	−	−	−

The large correlations that were found in the analysis suggest that only combinations 1 and 8 are represented in the data. If the sample turned out this way by chance, the prescription for resolving the collinearity problem is to collect additional data on some of the other combinations. For example, data based on combinations 1 and 2 alone could be used to evaluate the effect of SCHOOL on ACHV holding FAM and PEER at a constant level, both above average. If these were the only combinations represented in the data, the analysis would consist of the simple regression of ACHV against SCHOOL. The results would give only a partial answer, namely, an evaluation of the school-achievement relationship when FAM and PEER are both above average.

7.3 Collinearity: Effects of Collinearity on Inference

Example

The prescription for additional data as a way to resolve collinearity is not a panacea. It is often not possible to collect more data because of constraints on bud- gets, time, and staff. It is always better to be aware of impending data deficiencies beforehand. Whenever possible, the data should be collected according to design. Unfortunately, prior design is not always feasible. In surveys, or observational studies such as the one being discussed, the values of the predictor variables are usually not known until the sampling unit is selected for the sample and some costly and time-consuming measurements are developed. Following this procedure, it is fairly difficult to ensure that a balanced sample will be obtained.

The second reason that collinearity may appear is because the relationships among the variables are an inherent characteristic of the process being sampled. If FAM, PEER, and SCHOOL exist in the population only as data combinations 1 and 8 of Table 7.6, it is not possible to estimate the individual effects of these variables on achievement. The only recourse for continued analysis of these effects would be to search for underlying causes that may explain the interrelationships of the predictor variables. Through this process, one may discover other variables that are more basic determinants affecting equal opportunity in education and achievement.

7.4. Collinearity: Effects of Collinearity on Forecasting

7.4. Collinearity: Effects of Collinearity on Forecasting

Example

We shall examine the effects of collinearity in forecasting when the forecasts are based on a multiple regression equation. A historical data set with observations indexed by time is used to estimate the regression coefficients. Forecasts of the response variable are produced by using future values of the predictor variables in the estimated regression equation. The future values of the predictor variables must be known or forecasted from other data and models. We shall not treat the uncertainty in the forecasted predictor variables. In our discussion it is assumed that the future values of the predictor variables are given.

We have chosen an example based on aggregate data concerning import activity in the French economy. The variables are imports (IMPORT), domestic production (DOPROD), stock formation (STOCK), and domestic consumption (CONSUM), all measured in billions of French francs for the years 1949-1966. The data are given in Table 7.7. The model being considered is

$$\text{IMPORT} = \beta_0 + \beta_1\text{DOPROD} + \beta_2\text{STOCK} + \beta_3\text{CONSUM} + \varepsilon. \tag{7.8}$$

7.4. Collinearity: Effects of Collinearity on Forecasting

Example

Table 7.7 Data on French Economy

YEAR	IMPORT	DOPROD	STOCK	CONSUM
49	15.9	149.3	4.2	108.1
50	16.4	161.2	4.1	114.8
51	19.0	171.5	3.1	123.2
52	19.1	175.5	3.1	126.9
53	18.8	180.8	1.1	132.1
54	20.4	190.7	2.2	137.7
55	22.7	202.1	2.1	146.0
56	26.5	212.4	5.6	154.1
57	28.1	226.1	5.0	162.3
58	27.6	231.9	5.1	164.3
59	26.3	239.0	0.7	167.6
60	31.1	258.0	5.6	176.8
61	33.3	269.8	3.9	186.6
62	37.0	288.4	3.1	199.7
63	43.3	304.5	4.6	213.9
64	49.0	323.4	7.0	223.8
65	50.3	336.8	1.2	232.0
66	56.6	353.9	4.5	242.9

The regression results are presented in Table 7.8. The index plot of residuals (Figure 7.5) shows a distinctive pattern, suggesting that the model is not well specified. Even though collinearity appears to be present ($R^2 = 0.973$ and all t- values are small), it should not be pursued further in this model. Collinearity should only be attacked after the model specification is satisfactory. The difficulty with the model is that the European Common Market began operations in 1960, causing changes in import-export relationships. Since our objective here is to study the effects of collinearity, we shall not complicate the model by attempting to capture the behavior after 1959. We shall assume that it is now 1960 and look only at the 11 years 1949-1959. The regression results for those data are summarized in Table 7.9. The residual plot is now satisfactory (Figure 7.6)

7.4. Collinearity: Effects of Collinearity on Forecasting

Example

Table 7.8 Import data (1949–1966): Regression Results

ANOVA Table

Source	Sum of Squares	df	Mean Square	F-Test
Regression	2576.92	3	858.974	168
Residuals	71.39	14	5.099	

Coefficients Table

Variable	Coefficient	s.e.	t-Test	p-value
Constant	−19.725	4.125	−4.78	0.0003
DOPROD	0.032	0.187	0.17	0.8656
STOCK	0.414	0.322	1.29	0.2195
CONSUM	0.243	0.285	0.85	0.4093
$n = 18$	$R^2 = 0.973$	$R_a^2 = 0.967$	$\hat{\sigma} = 2.258$	df = 14

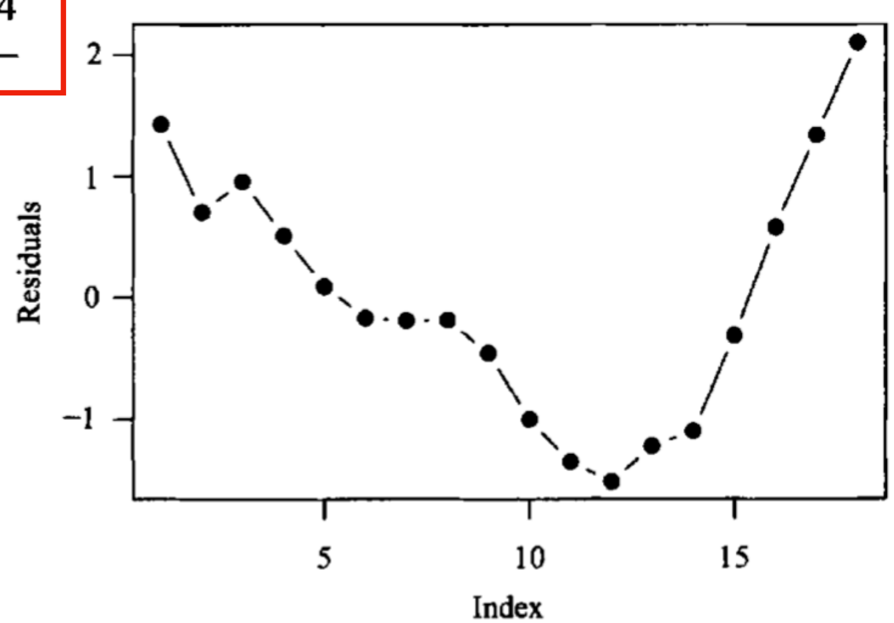


Figure 7.5 Import data (1949–1966): Index plot of the standardized residuals.

7.4. Collinearity: Effects of Collinearity on Forecasting

Example

Table 7.9 Import data (1949–1959): Regression Results

ANOVA Table

Source	Sum of Squares	df	Mean Square	F-Test
Regression	204.776	3	68.2587	286
Residuals	1.673	7	0.2390	

Coefficients Table

Variable	Coefficient	s.e.	t-Test	p-value
Constant	−10.128	1.212	−8.36	< 0.0001
DOPROD	−0.051	0.070	−0.73	0.4883
STOCK	0.587	0.095	6.20	0.0004
CONSUM	0.287	0.102	2.81	0.0263
$n = 11$	$R^2 = 0.992$	$R_a^2 = 0.988$	$\hat{\sigma} = 0.4889$	df = 7

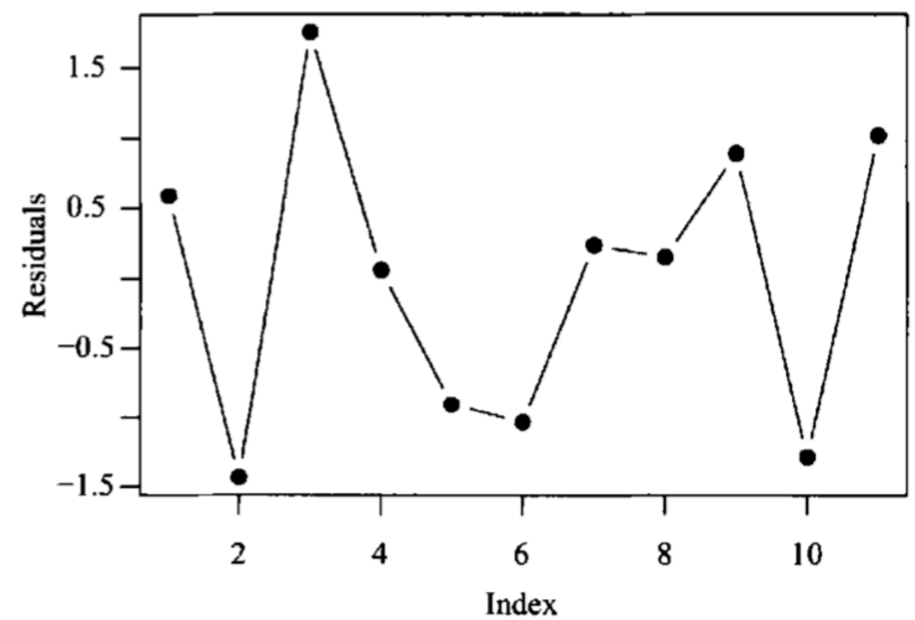


Figure 7.6 Import data (1949–1959): Index plot of the standardized residuals

7.4. Collinearity: Effects of Collinearity on Forecasting

Example

The value of $R^2 = 0.99$ is high. However, the coefficient of DOPROD is negative and not statistically significant, which is contrary to prior expectation. We believe that if STOCK and CONSUM were held constant, an increase in DOPROD would cause an increase in IMPORT, probably for raw materials or manufacturing equipment. Collinearity is a possibility here and in fact this is the case. The simple correlation between CONSUM and DOPROD is 0.997. Upon further investigation it turns out that CONSUM has been about two-thirds of DOPROD throughout the 11-year period. The estimated relationship between the two quantities is

$$\text{CONSUM} = 6.259 + 0.686\text{DOP}$$

Even in the presence of such severe collinearity the regression equation may produce some good forecasts. From Table 7.9, the forecasting equation

$$\text{IMPORT} = -10.13 - 0.051\text{DOPROD} + 0.587\text{STOCK} + 0.287\text{CONSUM}$$

Recall that the fit to the historical data is very good and the residual variation appears to be purely random. To forecast we must be confident that the character and strength of the overall relationship will hold into future periods. This matter of confidence is a problem in all forecasting models whether or not collinearity is present. For the purpose of this example we assume that the overall relationship does hold into future periods. Implicit in this assumption is the relationship between DOPROD and CONSUM. The forecast will be accurate as long as the future values of DOPROD, STOCK, and CONSUM have the relationship that CONSUM is approximately equal to $0.7 \times$ DOPROD.

7.4. Collinearity: Effects of Collinearity on Forecasting

Example

For example, let us forecast the change in IMPORT next year corresponding to an increase in DOPROD of 10 units while holding STOCK and CONSUM at their current levels. The resulting forecast is

$$\text{IMPORT}_{1960} = \text{IMPORT}_{1959} - 0.051(10),$$

which means that IMPORT will decrease by -0.51 units. However, if the relationship between DOPROD and CONSUM is kept intact, CONSUM will increase by $10(2/3) = 6.67$ units and the forecasted result is

$$\text{IMPORT}_{1960} = \text{IMPORT}_{1959} - 0.51 + 0.287 \times 6.67 = \text{IMPORT}_{1959} + 1.5.$$

IMPORT actually increases by 1.5 units, a more satisfying result and probably a better forecast. The case where DOPROD increases alone corresponds to a change in the basic structure of the data that were used to estimate the model parameters and cannot be expected to produce meaningful forecasts.

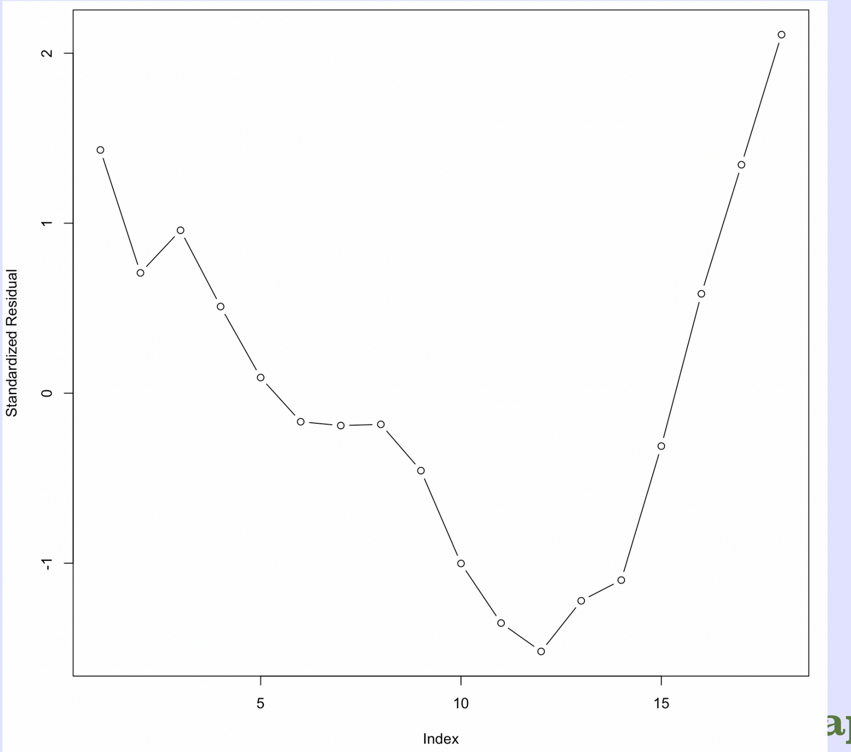
In summary, the two examples demonstrate that multicollinear data can seriously limit the use of regression analysis for inference and forecasting. Extreme care is required when attempting to interpret regression results when collinearity is suspected.

7.4. Collinearity: Effects of Collinearity on Forecasting

Use R for last example

```
> ############################################################################  Example on French Economy
> FREcon_dat<-read.table('data/P241.txt',header=TRUE)   ## read the data
> mod1<-lm(IMPORT~.-YEAR, data=FREcon_dat)
> summary(mod1)

Call:
lm(formula = IMPORT ~ . - YEAR, data = FREcon_dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7208 -1.8354 -0.3479  1.2973  4.1008 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -19.7251     4.1253  -4.782 0.000293 ***
DOPROD        0.0322     0.1869   0.172 0.865650    
STOCK         0.4142     0.3223   1.285 0.219545    
CONSUM        0.2427     0.2854   0.851 0.409268    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.258 on 14 degrees of freedom
Multiple R-squared:  0.973,	Adjusted R-squared:  0.9673 
F-statistic: 168.4 on 3 and 14 DF,  p-value: 3.212e-11
```

```
pii=hatvalues(mod1)
plot(1:dim(FREcon_dat)[1],mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),xlab="Index",ylab="Standardized Residual",type="b")
```

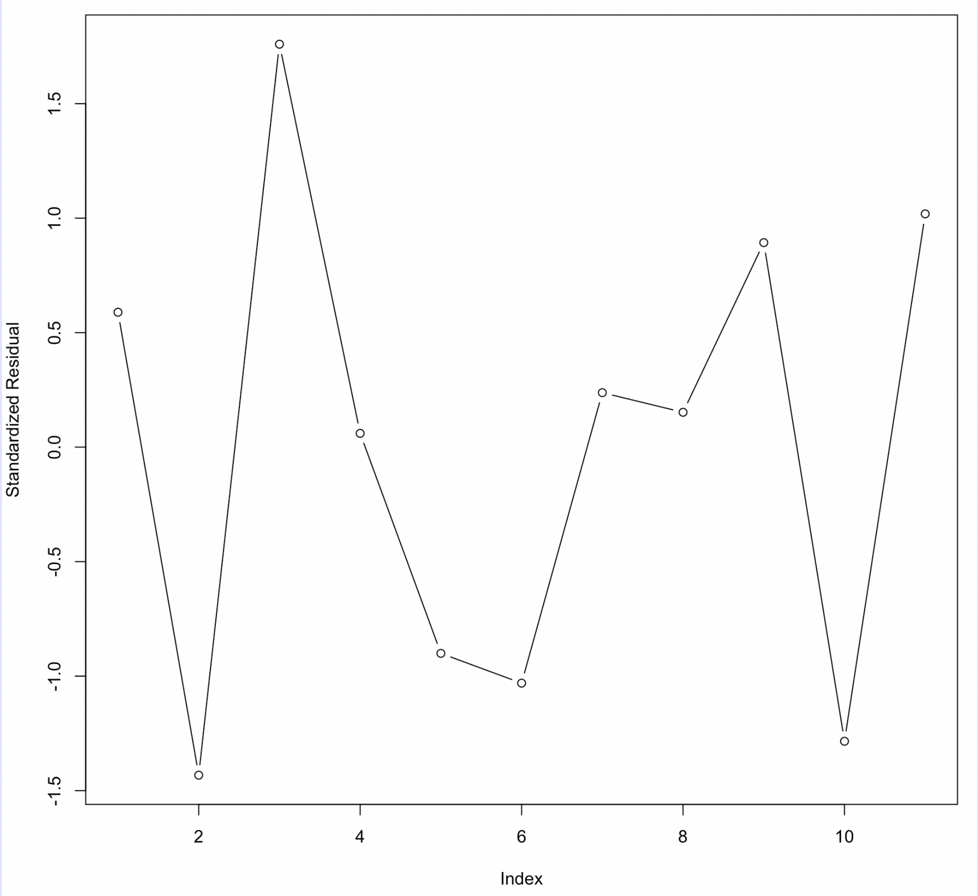


7.4. Collinearity: Effects of Collinearity on Forecasting

Use R for last example

```
> ####### use only part of the data
> FREcon_dat_new <- FREcon_dat[1:11,]
> mod1<-lm(IMPORT~.-YEAR, data=FREcon_dat_new)
> summary(mod1)

Call:
lm(formula = IMPORT ~ . - YEAR, data = FREcon_dat_new)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.52367 -0.38953  0.05424  0.22644  0.78313 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -10.12799    1.21216  -8.355  6.9e-05 ***
DOPROD       -0.05140    0.07028  -0.731 0.488344    
STOCK         0.58695    0.09462   6.203 0.000444 ***
CONSUM        0.28685    0.10221   2.807 0.026277 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4889 on 7 degrees of freedom
Multiple R-squared:  0.9919,	Adjusted R-squared:  0.9884 
F-statistic: 285.6 on 3 and 7 DF,  p-value: 1.112e-07
```

```
pii=hatvalues(mod1)
plot(1:dim(FREcon_dat_new)[1],mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),xlab="Index",ylab="Standardized Residual",type="b")
```



7.5. Collinearity: Detection of Collinearity

7.5 Collinearity: Detection of Collinearity

Simple Signs of Collinearity

In the preceding examples some of the ideas for detecting collinearity were already introduced. In this section we review those ideas and introduce additional criteria that indicate collinearity.

Collinearity is associated with **unstable** estimated regression coefficients. This situation results from the presence of strong linear relationships among the predictor variables. **It is not a problem of misspecification**. Therefore, the empirical investigation of problems that result from collinear data set should begin **only after** the model has been satisfactorily specified. However, there may be some indications of collinearity that are encountered during the process of adding, deleting, and transforming variables or data points in search of the good model. Indications of collinearity that appear as instability in the estimated coefficients are as follows:

- Large changes in the estimated coefficients when a variable is added or deleted.
- Large changes in the estimated coefficients when a data point is altered or dropped.

Once the **residual plots** indicate that the model has been satisfactorily specified, collinearity may be present if:

- The algebraic signs of the estimated coefficients do not conform to prior expectations; and/or
- Coefficients of variables that are expected to be important have large standard errors (small t-values).

7.5 Collinearity: Detection of Collinearity

Simple Signs of Collinearity

For the IMPORT data discussed previously, the coefficient of DOPROD was negative and not significant. Both results are contrary to prior expectations. The effects of dropping or adding a variable can be seen in Table 7.10. There we see that the presence or absence of certain variables has a large effect on the other coefficients. For the EEO data (Tables 7.3 and 7.4) the algebraic signs are all correct, but their standard errors are so large that none of the coefficients are statistically significant. It was expected that they would all be important.

Table 7.10 Import Data (1949–1959): Regression Coefficients for All Possible Regressions

	Variable			
Regression	Constant	DOPROD	STOCK	CONSUM
1	−6.558	0.146	–	–
2	19.611	–	0.691	–
3	−8.013	–	–	0.214
4	−8.440	0.145	0.622	–
5	−8.884	−0.109	–	0.372
6	−9.743	–	0.596	0.212
7	−10.128	−0.051	0.587	0.287

7.5 Collinearity: Detection of Collinearity

Simple Signs of Collinearity

The presence of collinearity is also indicated by the size of the correlation coefficients that exist among the predictor variables. A large correlation between a pair of predictor variables indicates a strong linear relationship between those two variables. The correlations for the EEO data (Figure 7.4) are large for all pairs of predictor variables. For the IMPORT data, the correlation coefficient between DOPROD and CONSUM is 0.997.

The source of collinearity may be more subtle than a simple relationship between two variables. A linear relation can involve many of the predictor variables. It may not be possible to detect such a relationship with a simple correlation coefficient. As an example, we shall look at an analysis of the effects of advertising expenditures (A_t), promotion expenditures (P_t), and sales expense (E_t) on the aggregate sales of a firm in year t. The data represent a period of 23 years during which the firm was operating under fairly stable conditions. The data are given in Table 7.11.

The proposed regression model is

$$S_t = \beta_0 + \beta_1 A_t + \beta_2 P_t + \beta_3 E_t + \beta_4 A_{t-1} + \beta_5 P_{t-1} + \varepsilon_t, \qquad (7.9)$$

where A_{t-1} and P_{t-1} are the **lagged** one-year variables.

Table 7.9, 7.10

7.5 Collinearity: Detection of Collinearity

Simple Signs of Collinearity

Table 7.11 Annual Data on Advertising, Promotions, Sales Expenses, and Sales (Millions of Dollars)

Row	S_t	A_t	P_t	E_t	A_{t-1}	P_{t-1}
1	20.11371	1.98786	1.0	0.30	2.01722	0.0
2	15.10439	1.94418	0.0	0.30	1.98786	1.0
3	18.68375	2.19954	0.8	0.35	1.94418	0.0
4	16.05173	2.00107	0.0	0.35	2.19954	0.8
5	21.30101	1.69292	1.3	0.30	2.00107	0.0
6	17.85004	1.74334	0.3	0.32	1.69292	1.3
7	18.87558	2.06907	1.0	0.31	1.74334	0.3
8	21.26599	1.01709	1.0	0.41	2.06907	1.0
9	20.48473	2.01906	0.9	0.45	1.01709	1.0
10	20.54032	1.06139	1.0	0.45	2.01906	0.9
11	26.18441	1.45999	1.5	0.50	1.06139	1.0
12	21.71606	1.87511	0.0	0.60	1.45999	1.5
13	28.69595	2.27109	0.8	0.65	1.87511	0.0
14	25.83720	1.11191	1.0	0.65	2.27109	0.8
15	29.31987	1.77407	1.2	0.65	1.11191	1.0
16	24.19041	0.95878	1.0	0.65	1.77407	1.2
17	26.58966	1.98930	1.0	0.62	0.95878	1.0
18	22.24466	1.97111	0.0	0.60	1.98930	1.0
19	24.79944	2.26603	0.7	0.60	1.97111	0.0
20	21.19105	1.98346	0.1	0.61	2.26603	0.7
21	26.03441	2.10054	1.0	0.60	1.98346	0.1
22	27.39304	1.06815	1.0	0.58	2.10054	1.0

Table 7.12 Regression Results for the Advertising Data

ANOVA Table

Source	Sum of Squares	df	Mean Square	F-Test
Regression	307.572	5	61.514	35.3
Residuals	27.879	16	1.742	

Coefficients Table

Variable	Coefficient	s.e.	t-Test	p-value
Constant	−14.194	18.715	−0.76	0.4592
A_t	5.361	4.028	1.33	0.2019
P_t	8.372	3.586	2.33	0.0329
E_t	22.521	2.142	10.51	< 0.0001
A_{t-1}	3.855	3.578	1.08	0.2973
P_{t-1}	4.125	3.895	1.06	0.3053
$n = 22$	$R^2 = 0.917$	$R_a^2 = 0.891$	$\hat{\sigma} = 1.320$	df = 16

7.5 Collinearity: Detection of Collinearity

Simple Signs of Collinearity

The regression results are given in Table 7.12. The plot of residuals versus fitted values and the index plot of residuals (Figures 7.7 and 7.8), as well as other plots of the residuals versus the predictor variables (not shown), do not suggest any problems of misspecification. Furthermore, the correlation coefficients between the predictor variables are small (Table 7.13). However, if we do a little experimentation to check the stability of the coefficients by dropping the contemporaneous advertising variable A from the model, many things change. The coefficient of P_t drops from 8.37 to 3.70; the coefficients of lagged advertising A_{t-1} and lagged promotions P_{t-1} change signs. But the coefficient of sales expense is stable and R^2 does not change.

Table 7.13

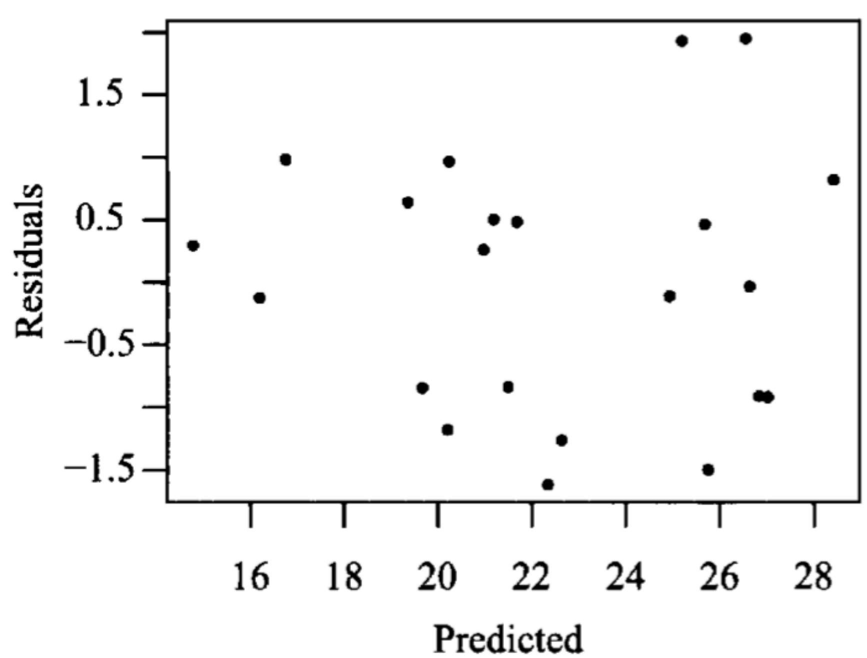


Figure 7.7 Standardized residuals versus fitted values of Sales.

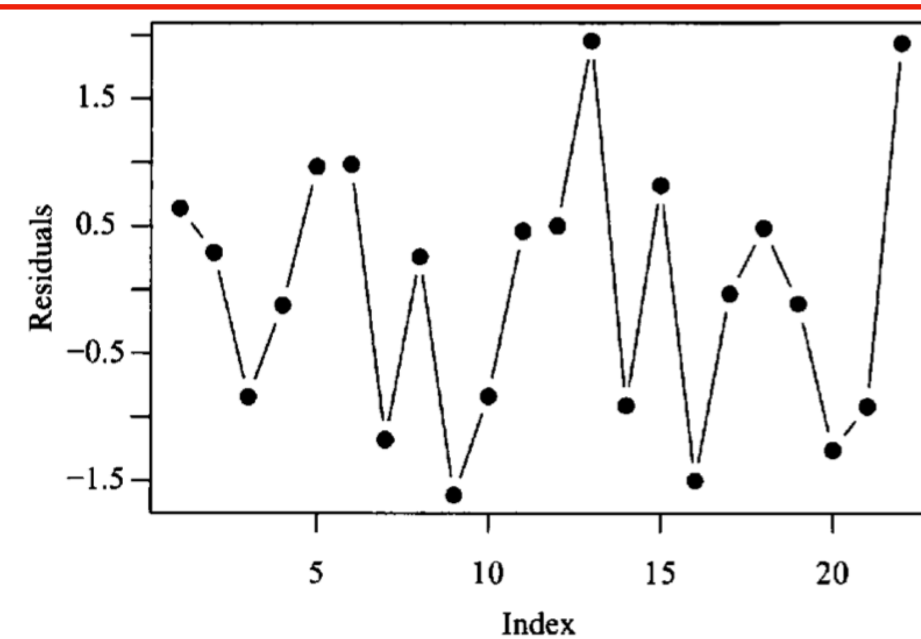


Figure 7.8 Index plot of the standardized residuals.

7.5 Collinearity: Detection of Collinearity

Simple Signs of Collinearity

Table 7.13 Pairwise Correlation Coefficients for the Advertising Data

	A_t	P_t	E_t	A_{t-1}	P_{t-1}
A_t	1.000				
P_t	−0.357	1.000			
E_t	−0.129	0.063	1.000		
A_{t-1}	−0.140	−0.316	−0.166	1.000	
P_{t-1}	−0.496	−0.296	0.208	−0.358	1.000

The evidence suggests that there is some type of relationship involving the contemporaneous and lagged values of the advertising and promotions variables. The regression of A_t on P_t, A_{t-1}, and P_{t-1} returns an R^2 of 0.973. The equation takes the form

$$\hat{A}_t = 4.63 - 0.87P_t - 0.86A_{t-1} - 0.95P_{t-1}$$

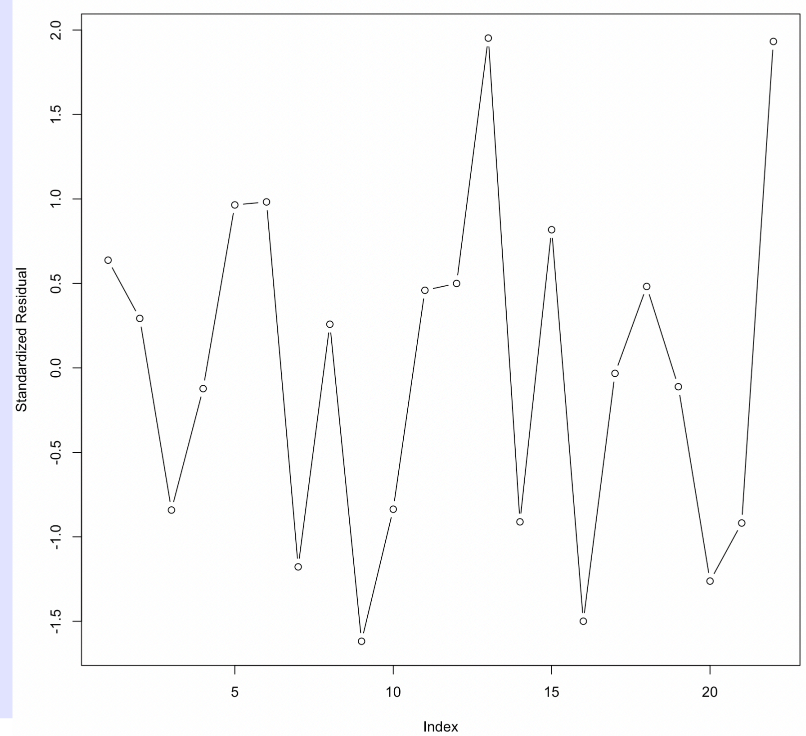
Upon further investigation into the operations of the firm, it was discovered that close control was exercised over the expense budget during those 23 years of stability. In particular, there was an approximate rule imposed on the budget that the sum of A_t, A_{t-1}, P_t, and P_{t-1} was to be held to approximately five units over every two-year period. The relationship $A_t + P_t + A_{t-1} + P_{t-1} \approx 5$ is the cause of the collinearity. The above indicators of the presence of collinearity (e.g., the **computational instability of the regression coefficients**, the regression results do **not conform** to prior expectations, and large values of **pairwise correlation coefficients**) are not sufficient to see a complete picture of collinearity. Another methods for measuring collinearity is the **variance inflation factor**.

7.5 Collinearity: Detection of Collinearity

Use R for last example

```
> ##########################################################################  Example on Advertising data
> Adv_dat<-read.table('data/P248.txt',header=TRUE)   ## read the data
> mod1<-lm(S_t~.,data=Adv_dat)
> summary(mod1)

Call:
lm(formula = S_t ~ ., data = Adv_dat)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8601 -0.9848  0.1323  0.7017  2.2046

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -14.194     18.715  -0.758   0.4592
A_t            5.361      4.028   1.331   0.2019
P_t            8.372      3.586   2.334   0.0329 *
E_t           22.521      2.142  10.512 1.36e-08 ***
A_.t.1.        3.855      3.578   1.077   0.2973
P_.t.1.        4.125      3.895   1.059   0.3053
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.32 on 16 degrees of freedom
Multiple R-squared:  0.9169,    Adjusted R-squared:  0.8909
F-statistic:  35.3 on 5 and 16 DF,  p-value: 4.289e-08
```

```
pii=hatvalues(mod1)
plot(1:dim(Adv_dat)[1],mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),xlab="Index",ylab="Standardized Residual",type="b")
```



7.5 Collinearity: Detection of Collinearity

Use R for last example

```
> ########### Regress A_t onto P_t, A_{t-1}, P_{t-1}
> mod2<-lm(A_t~P_t+A_.t.1.+P_.t.1.,data=Adv_dat)
> summary(mod2)

Call:
lm(formula = A_t ~ P_t + A_.t.1. + P_.t.1., data = Adv_dat)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.134930 -0.052396 -0.004111  0.063809  0.146043 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.63124    0.12937   35.80  < 2e-16 ***
P_t         -0.86953    0.04333  -20.07 9.08e-14 ***
A_.t.1.     -0.86340    0.05024  -17.18 1.30e-12 ***
P_.t.1.     -0.94689    0.04192  -22.59 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0776 on 18 degrees of freedom
Multiple R-squared:  0.9727,    Adjusted R-squared:  0.9681 
F-statistic: 213.6 on 3 and 18 DF,  p-value: 2.951e-14
```

7.5 Collinearity: Detection of Collinearity

Variance Inflation Factor

A thorough investigation of collinearity will involve examining the value of R^2 that results from regressing each of the predictor variables against **all the others**. The relationship between the predictor variables can be judged by examining a quantity called the **variance inflation factor** (VIF). Let R_j^2 be the square of the multiple correlation coefficient that results when the predictor variable X_j is regressed against all the other predictor variables. Then the variance inflation for X_j is

$$\mathrm{VIF}_j = \frac{1}{1-R_j^2}, \quad j = 1, \cdots, p, \tag{7.10}$$

where p is the number of predictor variables. It is clear that if X_j has a strong linear relationship with the other predictor variables, R_j^2 would be close to 1, and VIF_j would be large. Values of variance inflation factors greater than 10 is often taken as a signal that the data have collinearity problem.

In the absence of any linear relationship between the predictor variables (i.e., if the predictor variables are **orthogonal**), R_j^2 would be 0 and VIF_j would be 1. The deviation of the VIF_j value from 1 indicates departure from orthogonality and tendency toward collinearity

The value of VIF_j also measures the amount by which the variance of the j-th regression coefficient is increased due to the linear association of X_j with other predictor variables relative to the variance that would result if X_j were not related to them linearly. This explains the naming of this particular diagnostic.

7.5 Collinearity: Detection of Collinearity

Variance Inflation Factor

As R_j^2 tends toward 1, indicating the presence of a linear relationship in the predictor variables, the VIF for $\hat{\beta}_j$ tends to infinity. It is suggested that a VIF in excess of 10 is an indication that collinearity may be causing problems in estimation.

The precision of an ordinary least squares (OLS) estimated regression coefficient is measured by its variance, which is proportional to σ^2, the variance of the error term in the regression model. The constant of proportionality is the VIF. Thus, the VIFs may be used to obtain an expression for the expected squared distance of the OLS estimators from their true values. Denoting the square of the distance by D^2, it can be shown that, on average,

$E[(\hat{\beta}_j - \beta_j)^2] \propto \sigma^2 \cdot \mathrm{VIF}_j$

$j = 1, \cdots, p$

$D^2 = \sum_{j=1}^p E[(\hat{\beta}_j - \beta_j)^2]$

$$D^2 = \sigma^2 \sum_{j=1}^{p} \mathrm{VIF}_j.$$

This distance is another measure of precision of the least squares estimators. The smaller the distance, the more accurate are the estimates. If the predictor variables were orthogonal, the VIFs would all be 1 and D^2 would be $p\sigma^2$. It follows that the ratio

$$\frac{\sigma^2 \sum_{i=1}^p \mathrm{VIF}_i}{p\sigma^2} = \frac{\sum_{i=1}^p \mathrm{VIF}_i}{p} = \overline{\mathrm{VIF}},$$

which shows that the average of the VIFs measures the squared error in the OLS estimators relative to the size of that error if the data were orthogonal. Hence, $\overline{\mathrm{VIF}}$ may also be used as an index of collinearity.

7.5 Collinearity: Detection of Collinearity

Variance Inflation Factor

Most computer packages now furnish values of VIF_j routinely. Some have built-in messages when high values of VIF_j are observed. In any regression analysis the values of VIF_j should always be examined to avoid the pitfalls resulting from fitting a regression model to collinear data by least squares.

In each of the three examples (EEO, Import, and Advertising) we have seen evidence of collinearity. The VIF_j's and their average values for these data sets are given in Table 7.14. For the EEO data the values of VIF_j range from 30.2 to 83.2, showing that all three variables are strongly intercorrelated and that dropping one of the variables will not eliminate collinearity. The average value of VIF of 50.3 indicates that the squared error in the OLS estimators is 50 times as large as it would be if the predictor variables were orthogonal.

Table 7.14 Variance Inflation Factors for Three Data Sets

EEO		Import		Advertising	
Variable	VIF	Variable	VIF	Variable	VIF
FAM	37.6	DOPROD	469.7	A_t	37.4
PEER	30.2	STOCK	1.0	P_t	33.5
SCHOOL	83.2	CONSUM	469.4	E_t	1.1
				A_{t-1}	26.6
				P_{t-1}	44.1
Average	50.3	Average	313.4	Average	28.5

7.5 Collinearity: Detection of Collinearity

Variance Inflation Factor

For the Import data, the squared error in the OLS estimators is 313 times as large as it would be if the predictor variables were orthogonal. However, the VIF_j's indicate that domestic production and consumption are strongly correlated but are not correlated with the STOCK variable. A regression equation containing either CONSUM or DOPROD along with STOCK will eliminate collinearity.

For the Advertising data, VIF_E (for the variable E) is 1.1, indicating that this variable is not correlated with the remaining predictor variables. The VIF_j's for the other four variables are large, ranging from 26.6 to 44.1. This indicates that there is a strong linear relationship among the four variables, a fact that we have already noted. Here the prescription might be to regress sales S_t against E_t and three of the remaining four variables $(A_t, P_t, A_{t-1}, S_{t-1})$ and examine the resulting VIF_j's to see if collinearity has been eliminated.

We should note, however, that deleting some predictor variables is not always the best way to reduce collinearity and sometimes it does not work at all. For example, we have seen in Hamilton's data (see Chapter 4), that the two predictor variables are mildly collinear but the response variable Y depends on them collectively but not individually. So, when we delete one of the predictor variables, one cannot predict Y using only one of the predictor variables.

7.5 Collinearity: Detection of Collinearity

VIF in R

```
########################################################## Variation Inflation Factor
library(regclass)
EEO_dat<-read.table('data/P236.txt',header=TRUE)   ## read the data
mod1<-lm(ACHV~.,data=EEO_dat)
VIF(mod1)
```

```
> VIF(mod1)
     FAM     PEER   SCHOOL
37.58064 30.21166 83.15544
```

```
> FREcon_dat<-read.table('data/P241.txt',header=TRUE)   ## read the data
> mod1<-lm(IMPORT~.-YEAR, data=FREcon_dat)
> VIF(mod1)
    DOPROD      STOCK     CONSUM
469.742135   1.049877 469.371343
```

```
> Adv_dat<-read.table('data/P248.txt',header=TRUE)   ## read the data
> mod1<-lm(S_t~.,data=Adv_dat)
> VIF(mod1)
      A_t       P_t       E_t   A_.t.1.   P_.t.1.
36.941513 33.473514  1.075962 25.915651 43.520965
```

7.6 Collinearity: Ridge Regression

7.6 Collinearity: Ridge Regression

Ridge regression provides an alternative estimation method that may be used to advantage when the predictor variables are **highly collinear**. There are a number of alternative ways to define and compute ridge estimates. We have chosen to present the method associated with the **ridge trace**. It is a graphical approach and may be viewed as an **exploratory technique**. Ridge analysis using the ridge trace represents a unified approach to problems o f detection and estimation when collinearity is suspected. The estimators produced are biased but tend to have a smaller mean squared error than OLS estimators.

Ridge estimates of the regression coefficients may be obtained by solving a slightly altered form of the **normal equations** (introduced in Chapter 3). Assume that the **standardized form** of the regression model is given by

$\tilde{Y}, \tilde{X}_1, \tilde{X}_2$ are the **standardized** version of Y, X_1, X_2, respectively.
sample mean and sample variance of $\tilde{Y}$ are 0 and 1

$$\tilde{Y} = \theta_1 \tilde{X}_1 + \theta_2 \tilde{X}_2 + \cdots + \theta_p \tilde{X}_p + \varepsilon'. \qquad (7.36)$$

The estimating equations for the ridge regression coefficients are

$$\begin{array}{ccccccccc}
(1+k)\theta_1 & + & r_{12}\,\theta_2 & + & \cdots & + & r_{1p}\,\theta_p & = & r_{1y}, \\
r_{21}\,\theta_1 & + & (1+k)\theta_2 & + & \cdots & + & r_{2p}\,\theta_p & = & r_{2y}, \\
 & \vdots & & \vdots & \vdots & & & \vdots & \\
r_{p1}\,\theta_1 & + & r_{p2}\,\theta_2 & + & \cdots & + & (1+k)\theta_p & = & r_{py},
\end{array} \qquad (7.37)$$

where r_{ij} is the correlation between the ith and jth predictor variables and r_{iy} is the correlation between the ith predictor variable and the response variable $\tilde{Y}$.

7.6 Collinearity: Ridge Regression

The solution to (7.37), $\hat{\theta}_1, \cdots, \hat{\theta}_p$, is the set of **estimated ridge regression coefficients**. The ridge estimates may be viewed as resulting from a set of data that has been slightly altered.

Using Matrix Notation $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)' \in \mathbb{R}^{p\times 1}$

$$\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \cdots, \tilde{\mathbf{x}}_p) \in \mathbb{R}^{n\times p}$$

Standardized Predictor Matrix

$$\tilde{\mathbf{Y}} = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{pmatrix} \in \mathbb{R}^{n\times 1}$$

Standardized Response vector

By chapter 3, $\mathbf{R} = \frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n-1}$ and $\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}}{n-1} = \begin{pmatrix} r_{1y} \\ r_{2y} \\ \vdots \\ r_{py} \end{pmatrix}$

by adding a diagonal matrix, it forces the matrix $\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n-1} + k\mathbf{I}_p$ to be invertible

$$\underline{\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n-1} + k\mathbf{I}_p\right)}\boldsymbol{\theta} = \frac{\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}}{n-1}$$

Normal Equations for Ridge Regression

$$\hat{\boldsymbol{\theta}} = \left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n-1} + k\mathbf{I}_p\right)^{-1}\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}}{n-1}$$

Estimated Ridge Regression Coefficient

7.6 Collinearity: Ridge Regression

Penalized Least Square Estimate

The OLS (ordinary least square) estimate aims to minimize the sum of squares:

$$\hat{\boldsymbol{\beta}} \quad \Longleftarrow \quad \min_{\boldsymbol{\beta}} \frac{1}{n-1}\|\tilde{\mathbf{X}}\boldsymbol{\beta} - \tilde{\mathbf{Y}}\|^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\tilde{\mathbf{x}}_i'\boldsymbol{\beta} - \tilde{y}_i)^2$$

The Ridge regression is equivalent to a **penalized** least squares with the penalty $k\|\boldsymbol{\beta}\|^2$, that is

$$\hat{\boldsymbol{\beta}} \quad \Longleftarrow \quad \min_{\boldsymbol{\beta}} \frac{1}{n-1}\|\tilde{\mathbf{X}}\boldsymbol{\beta} - \tilde{\mathbf{Y}}\|^2 + k\|\boldsymbol{\beta}\|^2$$

The solution is exactly the Ridge estimate given in last slide.

7.6 Collinearity: Ridge Regression

The essential parameter that distinguishes ridge regression from OLS is k. Note that when $k = 0$, the θ' s are the OLS estimates. The parameter k may be referred to as the **bias parameter**. As k increases from zero, bias of the estimates increases. On the other hand, the total variance (the sum of the variances of the estimated regression coefficients) is

λ_j is the jth largest eigenvalue of $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$

$$\text{Total Variance}(k) = \sum_{j=1}^{p} \text{Var}(\hat{\theta}_j(k)) = \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2}, \qquad (7.38)$$

which is a decreasing function of k. The formula in (7.38) shows the effect of the ridge parameter on the total variance of the ridge estimates of the regression coefficients. Substituting $k = 0$ in (7.38), we obtain

$$\text{Total Variance}(0) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}, \qquad (7.39)$$

which shows the effect of small eigenvalue on the total variance of the OLS estimates of the regression coefficients.

As k continues to increase without bound, the regression estimates all tend toward **zero**. The idea of ridge regression is to pick a value of k for which the reduction in total variance is not exceeded by the increase in bias.

It has been shown that there is a positive value of k for which the ridge estimates will be stable with respect to small changes in the estimation data. In practice, a value of k is chosen by computing $\hat{\theta}_1, \cdots, \hat{\theta}_p$ for a range of k values between 0 and 1 and plotting the results against k. The resulting graph is known as the **ridge trace** and is used to select an appropriate value for k. Guidelines for choosing k are given in the following example.

7.6 Collinearity: Ridge Regression

Example

A method for detecting collinearity that comes out of ridge analysis deals with the instability in the estimated coefficients resulting from slight changes in the estimation data. The instability may be observed in the **ridge trace**. The ridge trace is a simultaneous graph of the regression coefficients, $\hat{\theta}_1, \cdots, \hat{\theta}_p$, plotted against k for various values of k such as $0.001, 0.002$, and so on. Figure 7.14 is the **ridge trace** for the IMPORT data. The graph is constructed from Table 7.15, which has the ridge estimated coefficients for 29 values of k ranging from 0 to 1. Typically, the values of k are chosen to be concentrated near the low end of the range. If the estimated coefficients show large fluctuations for small values of k, instability has been demonstrated and collinearity is probably at work.

What is evident from the trace or equivalently from Table 7.15 is that the estimated values of the coefficients θ_1 and θ_3 are quite unstable for small values of k. The estimate of θ_1 changes rapidly from an implausible negative value of -0.339 to a stable value of about 0.43. The estimate of θ_3 goes from 1.303 to stabilize at about 0.50. The coefficient of $\tilde{X}_2$ (STOCK), θ_2 is unaffected by the collinearity and remains stable throughout at about 0.21.

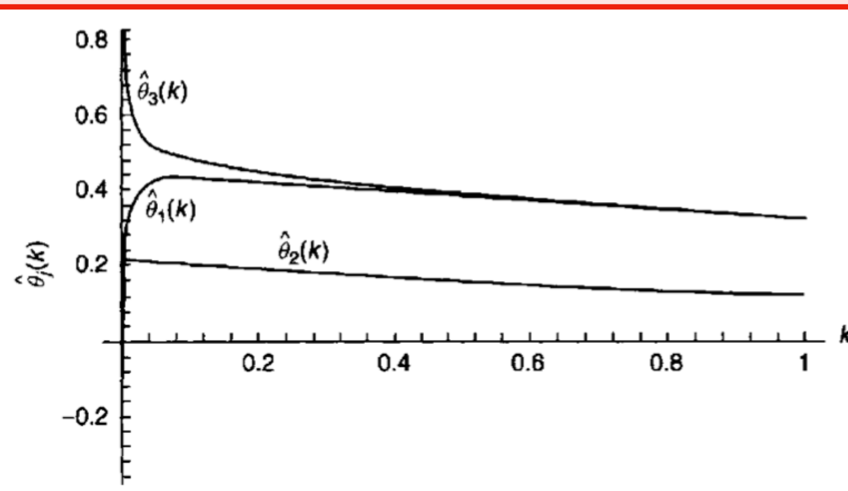


Figure 7.14 Ridge trace: IMPORT data (1949–1959).

7.6 Collinearity: Ridge Regression

Example

Table 7.15 Ridge Estimates $\hat{\theta}_j(k)$, as Functions of the Ridge Parameter k, for the IMPORT Data (1949–1959)

k	$\hat{\theta}_1(k)$	$\hat{\theta}_2(k)$	$\hat{\theta}_3(k)$
0.000	−0.339	0.213	1.303
0.001	−0.117	0.215	1.080
0.003	0.092	0.217	0.870
0.005	0.192	0.217	0.768
0.007	0.251	0.217	0.709
0.009	0.290	0.217	0.669
0.010	0.304	0.217	0.654
0.012	0.328	0.217	0.630
0.014	0.345	0.217	0.611
0.016	0.359	0.217	0.597
0.018	0.370	0.216	0.585
0.020	0.379	0.216	0.575
0.022	0.386	0.216	0.567
0.024	0.392	0.215	0.560
0.026	0.398	0.215	0.553
0.028	0.402	0.215	0.548
0.030	0.406	0.214	0.543
0.040	0.420	0.213	0.525
0.050	0.427	0.211	0.513
0.060	0.432	0.209	0.504
0.070	0.434	0.207	0.497
0.080	0.436	0.206	0.491
0.090	0.436	0.204	0.486
0.100	0.436	0.202	0.481
0.200	0.426	0.186	0.450
0.300	0.411	0.173	0.427
0.400	0.396	0.161	0.408
0.500	0.381	0.151	0.391
0.600	0.367	0.142	0.376
0.700	0.354	0.135	0.361
0.800	0.342	0.128	0.348
0.900	0.330	0.121	0.336
1.000	0.319	0.115	0.325

7.6 Collinearity: Ridge Regression

Example

The next step in the ridge analysis is to select a value of k and to obtain the corresponding estimates of the regression coefficients. If collinearity is a serious problem, the ridge estimators will vary dramatically as k is slowly increased from zero. As k increases, the coefficients will eventually stabilize. Since k is a bias parameter, it is desirable to select the smallest value of k for which stability occurs since the size of k is directly related to the amount of bias introduced. Several methods have been suggested for the choice of k. These methods include:

1. *Fixed Point.* Hoerl, Kennard, and Baldwin (1975) suggest estimating k by

$$k = \frac{p\hat{\sigma}^2(0)}{\sum_{j=1}^{p} [\hat{\theta}_j(0)]^2}, \tag{7.40}$$

where $\hat{\theta}_1(0), \cdots, \hat{\theta}_p(0)$ are the least squares estimates of $\theta_1, \cdots, \theta_p$ when the model in (7.36) is fitted to the data (i.e., when $k = 0$), and $\hat{\sigma}^2(0)$ is the corresponding residual mean square.

2. *Iterative Method.* Hoerl and Kennard (1976) propose the following iterative procedure for selecting k: Start with the initial estimate of k in (7.40) Denote this value by k_0. Then, calculate

$$k_1 = \frac{p\hat{\sigma}^2(0)}{\sum_{j=1}^{p} [\hat{\theta}_j(k_0)]^2}. \tag{7.41}$$

Then use k_1 to calculate k_2 as

$$k_2 = \frac{p\hat{\sigma}^2(0)}{\sum_{j=1}^{p} [\hat{\theta}_j(k_1)]^2}. \tag{7.42}$$

Repeat this process until the difference between two successive estimates of k is negligible.

3. *Ridge Trace.* The behavior of $\hat{\theta}_{\mathbf{j}}(k)$ as a function of k is easily observed from the ridge trace. The value of k selected is the smallest value for which all the coefficients $\hat{\theta}_j(k)$ are stable. In addition, at the selected value of k, the residual sum of squares should remain close to its minimum value. The variance inflation factors,[9] $\mathrm{VIF}_j(k)$, should also get down to less than 10. (Recall that a value of 1 is a characteristic of an orthogonal system and a value less than 10 would indicate a noncollinear or stable system.)

$\mathrm{VIF}_j(k)$ is the jth diagonal element of the matrix $\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n-1} + k\mathbf{I}\right)^{-1} \frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n-1} \left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{n-1} + k\mathbf{I}\right)^{-1}$

7.6 Collinearity: Ridge Regression

Example

For the IMPORT data, the fixed point formula in (7.40) gives

$$k = \frac{3 \times 0.0101}{(-0.339)^2 + (0.213)^2 + (1.303)^2} = 0.0164. \qquad (7.43)$$

The iterative method gives the following sequence: $k_0 = 0.0164$, $k_1 = 0.0161$, and $k_2 = 0.0161$. So, it converges after two iterations to $k = 0.0161$.

The estimated coefficients from the model stated in standardized and original variables units are summarized in Table 7.16 The original coefficient $\hat{\beta}_j$ is obtained from the standardized coefficient $\hat{\theta}_j$ For example, $\hat{\beta}_1$ is calculated by Formulas have been introduced in Chapter 3: before and after centering and scaling

$$\hat{\beta}_1 = (s_y/s_1)\hat{\theta}_1 = (4.5437/29.9995)(0.4196) = 0.0635.$$

Thus, the resulting model in terms for the original variables fitted by the Ridge method using $k = 0.04$ is

$$\begin{aligned} \text{IMPORT} = & -8.5537 + 0.0635\text{DOPROD} \\ & + 0.5859\text{STOCK} + 0.1156\text{CONSUM}. \end{aligned}$$

7.6 Collinearity: Ridge Regression

Example

Table 7.16 OLS and Ridge Estimates of the Regression Coefficients for IMPORT Data (1949–1959)

	OLS ($k = 0$)		Ridge ($k = 0.04$)	
Variable	Standardized Coefficients	Original Coefficients	Standardized Coefficients	Original Coefficients
Constant	0	−10.1300	0	−8.5537
DOPROD	−0.3393	−0.0514	0.4196	0.0635
STOCK	0.2130	0.5869	0.2127	0.5859
CONSUM	1.3027	0.2868	0.5249	0.1156
	$R^2 = 0.992$		$R^2 = 0.988$	