Example

Multicollinearity

Consumption Expenditure in Relation to Income and Wealth

TABLE 10.5 Hypothetical Data on Consumption Expenditure Y, Income X_2, and Wealth X_3

Y, \$	X_2, \$	X_3, \$
70	80	810
65	100	1009
90	120	1273
95	140	1425
110	160	1633
115	180	1876
120	200	2052
140	220	2201
155	240	2435
150	260	2686

```
  Source |       SS       df       MS              Number of obs =      10
---------+------------------------------           F(  2,     7) =   92.40
   Model |  8565.55407     2  4282.77704           Prob > F      =  0.0000
Residual |  324.445926     7   46.349418           R-squared     =  0.9635
---------+------------------------------           Adj R-squared =  0.9531
   Total |        8890     9  987.777778           Root MSE      =   6.808

------------------------------------------------------------------------------
       y |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
      x2 |   .9415373   .8228983     1.14   0.290     -1.004308    2.887383
      x3 |  -.0424345   .0806645    -0.53   0.615     -.2331757    .1483067
   _cons |   24.77473     6.7525     3.67   0.008      8.807609    40.74186
------------------------------------------------------------------------------
```

$$\hat{Y}_i = 24.7747 + 0.9415 X_{2i} - 0.0424 X_{3i}$$

$$(6.7525) \quad (0.8229) \quad (0.0807)$$

$$t = (3.6690) \quad (1.1442) \quad (-0.5261)$$

$$R^2 = 0.9635 \quad \bar{R}^2 = 0.9531 \quad df = 10 - 3 = 7$$

Regression shows that income and wealth together explain about 96 % of the variation in consumption expenditure, and yet **neither of the slope coefficients is individually statistically significant**. Moreover, not only is the wealth variable statistically insignificant but also it has the wrong sign

TABLE 10.6 ANOVA Table for the Consumption–Income–Wealth Example

Source of Variation	SS	df	MSS
Due to regression	8,565.5541	2	4,282.7770
Due to residual	324.4459	7	46.3494

$$H_0 = \beta_2 = \beta_3 = 0$$

$$F = \frac{4282.7770}{46.3494} = 92.4019$$

Reject the null hypothesis

(92.4019 > Critical F-value)

This example shows dramatically what multicollinearity does. The fact that the F test is significant but the t values of X_2 and X_3 are individually insignificant means that the two variables are so highly correlated that it is impossible to isolate the individual impact of either income and wealth on consumption

$$\hat{X}_{3i} = 7.5454 + 10.1909 X_{2i}$$

Source	SS	df	MS
Model	3427202.73	1	3427202.73
Residual	7123.27273	8	890.409091
Total	3434326	9	381591.778

Number of obs	10
F(1, 8)	3849.02
Prob > F	0.0000
R-squared	0.9979
Adj R-squared	0.9977
Root MSE	29.84

x3	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x2	10.19091	.1642623	62.04	0.000	9.81212	10.5697
_cons	7.545455	29.47581	0.26	0.804	-60.42589	75.5168

$$\hat{Y}_i = 24.4545 + 0.5091 X_{2i}$$

Source	SS	df	MS
Model	8552.72727	1	8552.72727
Residual	337.272727	8	42.1590909
Total	8890	9	987.777778

Number of obs	10
F(1, 8)	202.87
Prob > F	0.0000
R-squared	0.9621
Adj R-squared	0.9573
Root MSE	6.493

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x2	.5090909	.0357428	14.24	0.000	.4266678	.591514
_cons	24.45455	6.413817	3.81	0.005	9.664256	39.24483

$$\hat{Y}_i = 24.411 + 0.0498 X_{3i}$$

Source	SS	df	MS
Model	8504.87666	1	8504.87666
Residual	385.123344	8	48.1404181
Total	8890	9	987.777778

Number of obs	10
F(1, 8)	176.67
Prob > F	0.0000
R-squared	0.9567
Adj R-squared	0.9513
Root MSE	6.9383

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x3	.0497638	.003744	13.29	0.000	.0411301	.0583974
_cons	24.41104	6.874097	3.55	0.007	8.559349	40.26274

Regressions show very clearly that in situations of extreme multicollinearity dropping the highly collinear variable will often make the other X variable statistically significant.

This result would suggest that a way out of extreme collinearity is to drop the collinearity variable.