CHAPTER 6

Exercise Solutions

EXERCISE 6.1

(a) To compute R^2, we need *SSE* and *SST*. We are given *SSE*. We can find *SST* from the equation

$$\hat{\sigma}_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{N-1}} = \sqrt{\frac{SST}{N-1}} = 13.45222$$

Solving this equation for *SST* yields

$$SST = \hat{\sigma}_y^2 \times (N-1) = (13.45222)^2 \times 39 = 7057.5267$$

Thus,

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{979.830}{7057.5267} = 0.8612$$

(b) The *F*-statistic for testing $H_0: \beta_2 = \beta_3 = 0$ is defined as

$$F = \frac{(SST - SSE)/(K-1)}{SSE/(N-K)} = \frac{(7057.5267 - 979.830)/2}{979.830/(40-3)} = 114.75$$

At $\alpha = 0.05$, the critical value is $F_{(0.95,\,2,\,37)} = 3.25$. Since the calculated *F* is greater than the critical *F*, we reject H_0. There is evidence from the data to suggest that $\beta_2 \neq 0$ and/or $\beta_3 \neq 0$.

EXERCISE 6.2

The model from Exercise 6.1 is $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + e_i$. The *SSE* from estimating this model is 979.830. The model after augmenting with the squares and the cubes of predictions $\hat{y}_i^2$ and $\hat{y}_i^3$ is $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \gamma_1 \hat{y}_i^2 + \gamma_2 \hat{y}_i^3 + e_i$. The *SSE* from estimating this model is 696.5375. To use the RESET test, we set the null hypothesis $H_0:\ \gamma_1 = \gamma_2 = 0$. The *F*-value for testing this hypothesis is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(979.830 - 696.5375)/2}{696.5373/(40-5)} = 7.1175$$

The critical value for significance level $\alpha = 0.05$ is $F_{(0.95,2,35)} = 3.267$. Since the calculated *F* is greater than the critical *F* we reject H_0 and conclude that the model is misspecified.

EXERCISE 6.3

(a) Let the total variation, unexplained variation and explained variation be denoted by *SST*, *SSE* and *SSR*, respectively. Then, we have

$$SSE = \sum \hat{e}_i^2 = \left(N-K\right)\times\hat{\sigma}^2 = (20-3)\times 2.5193 = 42.8281$$

Also,

$$R^2 = 1 - \frac{SSE}{SST} = 0.9466$$

and hence the total variation is

$$SST = \frac{SSE}{1-R^2} = \frac{42.8281}{1-0.9466} = 802.0243$$

and the explained variation is

$$SSR = SST - SSE = 802.0243 - 42.8281 = 759.1962$$

(b) A 95% confidence interval for β_2 is

$$b_2 \pm t_{(0.975,17)}\text{se}(b_2) = 0.69914 \pm 2.110\times\sqrt{0.048526} = (0.2343, 1.1639)$$

A 95% confidence interval for β_3 is

$$b_2 \pm t_{(0.975,17)}\text{se}(b_3) = 1.7769 \pm 2.110\times\sqrt{0.037120} = (1.3704,\ 2.1834)$$

(c) To test H_0: $\beta_2 \geq 1$ against the alternative H_1: $\beta_2 < 1$, we calculate

$$t = \frac{b_2 - \beta_2}{\text{se}\left(b_2\right)} = \frac{0.69914-1}{\sqrt{0.048526}} = -1.3658$$

At a 5% significance level, we reject H_0 if $t < t_{(0.05,17)} = -1.740$. Since $-1.3658 > -1.740$, we fail to reject H_0. There is insufficient evidence to conclude $\beta_2 < 1$.

(d) To test $H_0:\beta_2 = \beta_3 = 0$ against the alternative $H_1:\beta_2 \neq 0$ and/or $\beta_3 \neq 0$, we calculate

$$F = \frac{\text{explained variation}\big/\left(K-1\right)}{\text{unexplained variation}\big/\left(N-K\right)} = \frac{759.1962/2}{42.8281/17} = 151$$

The critical value for a 5% level of significance is $F_{(0.95,2,17)} = 3.59$. Since $151 > 3.59$, we reject H_0 and conclude that the hypothesis $\beta_2 = \beta_3 = 0$ is not compatible with the data.

Exercise 6.3 (continued)

(e) The *t*-statistic for testing $H_0: 2\beta_2 = \beta_3$ against the alternative $H_1: 2\beta_2 \neq \beta_3$ is

$$t = \frac{\left(2b_2 - b_3\right)}{\text{se}\left(2b_2 - b_3\right)}$$

For a 5% significance level we reject H_0 if $t < t_{(0.025,17)} = -2.11$ or $t > t_{(0.975,17)} = 2.11$. The standard error is given by

$$\begin{aligned}\text{se}\left(2b_2 - b_3\right) &= \sqrt{2^2 \times \widehat{\text{var}(b_2)} + \widehat{\text{var}(b_3)} - 2\times 2\times \widehat{\text{cov}(b_2, b_3)}} \\ &= \sqrt{4\times 0.048526 + 0.03712 - 2\times 2\times\left(-0.031223\right)} \\ &= 0.59675\end{aligned}$$

The numerator of the *t*-statistic is

$$2b_2 - b_3 = 2\times 0.69914 - 1.7769 = -0.37862$$

leading to a *t*-value of

$$t = \frac{-0.37862}{0.59675} = -0.634$$

Since $-2.11 < -0.634 < 2.11$, we do not reject H_0. There is no evidence to suggest that $2\beta_2 \neq \beta_3$.

EXERCISE 6.4

In each case we use a two-tail test with a 5% significance level. The critical values are given by $t_{(0.025,60)} = -2.000$ and $t_{(0.975,60)} = 2.000$. The rejection region is $t < -2$ or $t > 2$.

(a) The value of the *t* statistic for testing the null hypothesis $H_0 : \beta_2 = 0$ against the alternative $H_1 : \beta_2 \neq 0$ is

$$t = \frac{b_2}{\text{se}(b_2)} = \frac{3}{\sqrt{4}} = 1.5$$

Since $-2 < 1.5 < 2$, we fail to reject H_0 and conclude that there is no sample evidence to suggest that $\beta_2 \neq 0$.

(b) For testing H_0: $\beta_1 + 2\beta_2 = 5$ against the alternative H_1: $\beta_1 + 2\beta_2 \neq 5$, we use the statistic

$$t = \frac{\left(b_1 + 2b_2\right) - 5}{\text{se}\left(b_1 + 2b_2\right)}$$

For the numerator of the *t*-value, we have

$$b_1 + 2b_2 - 5 = 2 + 2 \times 3 - 5 = 3$$

The denominator is given by

$$\text{se}(b_1 + b_2) = \sqrt{\widehat{\text{var}(b_1 + 2b_2)}} = \sqrt{\widehat{\text{var}(b_1)} + 4 \times \widehat{\text{var}(b_2)} + 4 \times \widehat{\text{cov}(b_1, b_2)}}$$
$$= \sqrt{3 + 4 \times 4 - 4 \times 2} = \sqrt{11} = 3.3166$$

Therefore,

$$t = \frac{3}{3.3166} = 0.9045$$

Since $-2 < 0.9045 < 2$, we fail to reject H_0. There is no sample evidence to suggest that $\beta_1 + 2\beta_2 \neq 5$.

Exercise 6.4 (continued)

(c) For testing $H_0 : \beta_1 - \beta_2 + \beta_3 = 4$ against the alternative $H_1 : \beta_1 - \beta_2 + \beta_3 \neq 4$, we use the statistic

$$t = \frac{(b_1 - b_2 + b_3) - 4}{\text{se}(b_1 - b_2 + b_3)}$$

Now,

$$(b_1 - b_2 + b_3) - 4 = 2 - 3 - 1 - 4 = -6$$

and

$$\begin{aligned}
\text{se}(b_1 - b_2 + b_3) &= \sqrt{\widehat{\text{var}}(b_1 - b_2 + b_3)} \\
&= \sqrt{\widehat{\text{var}}(b_1) + \widehat{\text{var}}(b_2) + \widehat{\text{var}}(b_3) - 2\widehat{\text{cov}}(b_1, b_2) + 2\widehat{\text{cov}}(b_1, b_3) - 2\widehat{\text{cov}}(b_2, b_3)} \\
&= \sqrt{3 + 4 + 3 + 2 \times 2 + 2 \times 1 - 0} \\
&= 4
\end{aligned}$$

Thus,

$$t = \frac{-6}{4} = -1.5$$

Since $-2 < -1.5 < 2$, we fail to reject H_0 and conclude that there is insufficient sample evidence to suggest that $\beta_1 - \beta_2 + \beta_3 = 4$ is incorrect.

EXERCISE 6.5

Consider, for example, the model

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + e_i$$

If we augment the model with the predictions $\hat{y}_i$ the model becomes

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \gamma \hat{y}_i + e_i$$

However, $\hat{y}_i = b_1 + b_2 x_i + b_3 z_i$ is perfectly collinear with x_i and w_i. This perfect collinearity means that least-squares estimation of the augmented model will fail.

EXERCISE 6.6

(a) Least squares estimation of $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + e_i$ gives $b_3 = 0.4979$, $\text{se}(b_3) = 0.1174$ and $t = 0.4979/0.1174 = 4.24$. This result suggests that b_3 is significantly different from zero and therefore w_i should be included in the model. Additionally, the RESET test based on the equation $y_i = \beta_1 + \beta_2 x_i + e_i$ gives F-values of 17.98 and 8.72 which are much higher than the 5% critical values of $F_{(0.95,1,32)} = 4.15$ and $F_{(0.95,2,31)} = 3.30$, respectively. Thus, the model omitting w_i is inadequate.

(b) Let b_2^* be the least squares estimator for β_2 in the model that omits w_i. The omitted-variable bias is given by

$$E(b_2^*) - \beta_2 = \beta_3 \frac{\widehat{\text{cov}(x,w)}}{\widehat{\text{var}(x)}}$$

Now, $\widehat{\text{cov}(x,w)} > 0$ because $r_{xw} > 0$. Thus, the omitted variable bias will be positive. This result is consistent with what we observe. The estimated coefficient for β_2 changes from -0.9985 to 4.1072 when w_i is omitted from the equation.

(c) The high correlation between x_i and w_i suggests the existence of collinearity. The observed outcomes that are likely to be a consequence of the collinearity are the sensitivity of the estimates to omitting w_i (the large omitted variable bias) and the insignificance of b_2 when both variables are included in the equation.

EXERCISE 6.7

(a) The coefficients of ln(*Y*), ln(*K*) and ln(*PF*) are 0.6792, 0.3503 and 0.3219, respectively. Since the model is in log-log form the coefficients are elasticities. The estimate 0.6792 is the percentage change in *VC* when *Y* changes by 1%, with the other variables held constant. Similarly, 0.3503 is the percentage change in *VC* when *K* changes by 1%, and 0.3219 is the percentage change in *VC* when *PF* changes by 1%, keeping the other variables constant in each case.

(b) An increase in any one of the explanatory variables should lead to an increase in variable cost, with the exception of ln(*STAGE*). For a given level of output (passenger-miles) and a given level of capital stock, longer flights should be cheaper than shorter ones. Thus, positive signs are expected for all variables except ln(*STAGE*), whose coefficient should be negative. All coefficients have the expected signs with the exception of $\ln(PM)$.

(c) The coefficient of $\ln(PM)$ has a *p*-value of 0.4966 which is higher than 0.05, indicating that this coefficient is not significantly different from zero. The *p*-values of the other coefficients are all 0.0000, indicating that they are significant.

(d) Augmenting the equation with the squares of the predictions, and squares and cubes of the predictions, yields the RESET test *F*-values of 3.3803 and 1.8601 with corresponding *p*-values of 0.0671 and 0.1577, respectively. These two *p*-values are higher than the conventional 0.05 level of significance indicating that the model is adequate.

(e) From the middle panel of Table 6.6 the *F*-value for testing $H_0: \beta_2 + \beta_3 = 1$ is 6.1048 with a *p*-value of 0.014. This *p*-value is less than the significance level of 0.05. We reject H_0 and conclude that constant returns to scale do not exist.

(f) The *F*-value and the *p*-value for testing $H_0: \beta_4 + \beta_5 + \beta_6 = 1$ can be read from the bottom panel of Table 6.6. The *F* value is very large and the corresponding *p*-value of 0.00000 is below the significance level of 0.05. We reject H_0 and conclude that there is no evidence to suggest that if all input prices increase by the same proportion, variable cost will increase by the same proportion.

Exercise 6.7 (continued)

(g) To test $H_0: \beta_2 + \beta_3 = 1$, the value of the t statistic is

$$t = \frac{b_2 + b_3 - 1}{\text{se}(b_2 + b_3)} = \frac{0.6792 + 0.3503 - 1}{0.01187} = 2.48$$

where the standard error is calculated from

$$\begin{aligned} \text{se}(b_2 + b_3) &= \sqrt{\widehat{\text{var}(b_2 + b_3)}} \\ &= \sqrt{\widehat{\text{var}(b_2)} + \widehat{\text{var}(b_3)} + 2\widehat{\text{cov}(b_2, b_3)}} \\ &= \sqrt{0.002851 + 0.002796 + 2(-0.002753)} \\ &= 0.011874 \end{aligned}$$

We reject H_0 because $2.48 > t_{(0.975,261)} = 1.969$. Note $t^2 = (2.48)^2 = 6.15 \approx F = 6.10$. The difference between t^2 and F is due to rounding error.

To test $H_0: \beta_4 + \beta_5 + \beta_6 = 1$, the value of the t-statistic is

$$t = \frac{b_4 + b_5 + b_6 - 1}{\text{se}(b_4 + b_5 + b_6)} = \frac{0.2754 + 0.3219 - 0.0683 - 1}{0.0542} = -8.69$$

where

$$\text{se}(b_4 + b_5 + b_6) = \sqrt{\widehat{\text{var}(b_4 + b_5 + b_6)}} = \sqrt{0.002938} = 0.0542$$

with

$$\begin{aligned} \widehat{\text{var}(b_4 + b_5 + b_6)} &= \widehat{\text{var}(b_4)} + \widehat{\text{var}(b_5)} + \widehat{\text{var}(b_6)} + 2\widehat{\text{cov}(b_4, b_5)} + 2\widehat{\text{cov}(b_4, b_6)} + 2\widehat{\text{cov}(b_5, b_6)} \\ &= 0.001919 + 0.001303 + 0.010068 - 2 \times 0.000088 \\ &\qquad - 2 \times 0.002159 - 2 \times 0.002929 \\ &= 0.002938 \end{aligned}$$

We reject H_0 because $-8.69 < t_{(0.025,261)} = -1.969$. Note that $t^2 = (-8.69)^2 = 75.52$ which is approximately equal to $F = 75.43$.

EXERCISE 6.8

There are a number of ways in which the restrictions can be substituted into the model, with each one resulting in a different restricted model. We have chosen to substitute out β_1 and β_3. With this in mind, we rewrite the restrictions as

$$\beta_3 = 1 - 3.8\beta_4$$

$$\beta_1 = 80 - 6\beta_2 - 1.9\beta_3 - 3.61\beta_4$$

Substituting the first restriction into the second yields

$$\beta_1 = 80 - 6\beta_2 - 1.9(1 - 3.8\beta_4) - 3.61\beta_4$$

Substituting this restriction and the first one $\beta_3 = 1 - 3.8\beta_4$ into the equation

$$S_i = \beta_1 + \beta_2 P_i + \beta_3 A_i + \beta_4 A_i^2 + e_i$$

yields

$$S_i = \left(80 - 6\beta_2 - 1.9(1 - 3.8\beta_4) - 3.61\beta_4\right) + \beta_2 P_i + \left(1 - 3.8\beta_4\right) A_i + \beta_4 A_i^2 + e_i$$

Rearranging this equation into a form suitable for estimation yields

$$\left(S_i - A_i - 78.1\right) = \beta_2\left(P_i - 6\right) + \beta_4\left(3.61 - 3.8A_i + A_i^2\right) + e_i$$

EXERCISE 6.9

The results of the tests in parts (a) to (e) appear in the following table. Note that, in all cases, there is insufficient evidence to reject the null hypothesis at the 5% level of significance.

Part	H_0	F-value	df	F_c (5%)	p-value
(a)	$\beta_2 = 0$	0.047	(1,20)	4.35	0.831
(b)	$\beta_2 = \beta_3 = 0$	0.150	(2,20)	3.49	0.862
(c)	$\beta_2 = \beta_4 = 0$	0.127	(2,20)	3.49	0.881
(d)	$\beta_2 = \beta_3 = \beta_4 = 0$	0.181	(3,20)	3.10	0.908
(e)	$\beta_2 + \beta_3 + \beta_4 + \beta_5 = 1$	0.001	(1,20)	4.35	0.980

(f) The auxiliary R^2s and the explanatory-variable correlations that are exhibited in the following table suggest a high degree of collinearity in the model.

		Correlation with Variables		
Variable	Auxiliary R^2	ln(L)	ln(E)	ln(M)
ln(K)	0.969	0.947	0.984	0.959
ln(L)	0.973		0.972	0.986
ln(E)	0.987			0.983
ln(M)	0.984			

To examine the effect of collinearity on the reliability of estimation, we examine the estimated equation, with t values in parentheses,

$$\widehat{\ln(Y)} = 0.035 + 0.056\ln(K) + 0.226\ln(L) + 0.044\ln(E) + 0.670\ln(M)$$
$$(t) \quad (0.800)\,(0.216) \qquad (0.511) \qquad (0.112) \qquad (1.855)$$
$$R^2 = 0.952$$

The very small t-values for all variables except $\ln(M)$, our inability to reject any of the null hypotheses in parts (a) through (e), and the high R^2, are indicative of high collinearity. Collectively, all the variables produce a model with a high level of explanation and a good predictive ability. Furthermore, our economic theory tells us that all the variables are important ones in a production function. However, we have not been able to estimate the effects of the individual explanatory variables with any reasonable degree of precision.

EXERCISE 6.10

(a) The restricted and unrestricted least squares estimates and their standard errors appear in the following table. The two sets of estimates are similar except for the noticeable difference in sign for ln(*PL*). The positive restricted estimate 0.187 is more in line with our *a priori* views about the cross-price elasticity with respect to liquor than the negative estimate –0.583. Most standard errors for the restricted estimates are less than their counterparts for the unrestricted estimates, supporting the theoretical result that restricted least squares estimates have lower variances.

	CONST	ln(*PB*)	ln(*PL*)	ln(*PR*)	ln(*I*)
Unrestricted	–3.243	–1.020	–0.583	0.210	0.923
	(3.743)	(0.239)	(0.560)	(0.080)	(0.416)
Restricted	–4.798	–1.299	0.187	0.167	0.946
	(3.714)	(0.166)	(0.284)	(0.077)	(0.427)

(b) The high auxiliary R^2s and sample correlations between the explanatory variables that appear in the following table suggest that collinearity could be a problem. The relatively large standard error and the wrong sign for $\ln(PL)$ are a likely consequence of this correlation.

		Sample Correlation With		
Variable	Auxiliary R^2	ln(*PL*)	ln(*PR*)	ln(*I*)
ln(*PB*)	0.955	0.967	0.774	0.971
ln(*PL*)	0.955		0.809	0.971
ln(*PR*)	0.694			0.821
ln(*I*)	0.964			

(c) We use the *F*-test to test the restriction $H_0 : \beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$ against the alternative hypothesis $H_1 : \beta_2 + \beta_3 + \beta_4 + \beta_5 \neq 0$. The value of the test statistic is *F* = 2.50, with a *p*-value of 0.127. The critical value is $F_{(0.95,1,25)} = 4.24$. Since $2.50 < 4.24$, we do not reject H_0. The evidence from the data is consistent with the notion that if prices and income go up in the same proportion, demand will not change. This idea is consistent with economic theory.

The *F*-value can be calculated from restricted and unrestricted sums of squared errors as follows

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(0.098901 - 0.08992)/1}{0.08992/25} = 2.50$$

Exercise 6.10 (continued)

(d)(e) The results for parts (d) and (e) appear in the following table. The *t*-values used to construct the interval estimates are $t_{(0.975,\,25)} = 2.060$ for the unrestricted model and $t_{(0.975,\,26)} = 2.056$ for the restricted model. The two 95% prediction intervals are (70.6, 127.9) and (59.6, 116.7). The effect of the nonsample restriction has been to increase both endpoints of the interval by approximately 10 litres.

					ln(Q)		Q	
		$\widehat{\ln(Q)}$	se(f)	t_c	lower	upper	lower	upper
(d)	Restricted	4.5541	0.14446	2.056	4.257	4.851	70.6	127.9
(e)	Unrestricted	4.4239	0.16285	2.060	4.088	4.759	59.6	116.7

EXERCISE 6.11

(a) The estimated Cobb-Douglas production function with standard errors in parentheses is

$$\widehat{\ln(Q)} = 0.129 + 0.559\ln(L) + 0.488\ln(K) \qquad R^2 = 0.688$$
$$\text{(se)} \quad (0.546) \quad (0.816) \qquad (0.704)$$

The magnitudes of the elasticities of production (coefficients of ln(L) and ln(K)) seem reasonable, but their standard errors are very large, implying the estimates are unreliable. The sample correlation between ln(L) and ln(K) is 0.986. It seems that labor and capital are used in a relatively fixed proportion, leading to a collinearity problem which has produced the unreliable estimates.

(b) After imposing constant returns to scale the estimated function is

$$\widehat{\ln(Q)} = 0.020 + 0.398\ln(L) + 0.602\ln(K)$$
$$\text{(se)} \quad (0.053) \quad (0.559) \qquad (0.559)$$

We note that the relative magnitude of the elasticities of production with respect to capital and labor has changed, and the standard errors have declined. However, the standard errors are still relatively large, implying that estimation is still imprecise.

EXERCISE 6.12

The RESET test results for the log-log and the linear demand function are reported in the table below.

Test		*F*-value	df	5% Critical *F*	*p*-value
Log-log	1 term	0.0075	(1,24)	4.260	0.9319
	2 terms	0.3581	(2,23)	3.422	0.7028
Linear	1 term	8.8377	(1,24)	4.260	0.0066
	2 terms	4.7618	(2,23)	3.422	0.0186

Because the RESET test returns *p*-values less than 0.05 (0.0066 and 0.0186 for one and two terms respectively), at a 5% level of significance we conclude that the linear model is not an adequate functional form for the beer data. On the other hand, the log-log model appears to suit the data well with relatively high *p*-values of 0.9319 and 0.7028 for one and two terms respectively. Thus, based on the RESET test we conclude that the log-log model better reflects the demand for beer.

EXERCISE 6.13

(a) The estimated model is

$$\hat{Y}_t = 0.6254 + 0.0302t - 0.0794RG_t - 0.0005RD_t + 0.3387RF_t \qquad R^2 = 0.6889$$
$$\begin{array}{lccccc} (\text{se}) & (0.2582) & (0.0034) & (0.0817) & (0.0918) & (0.1654) \\ (t) & (2.422) & (8.785) & (-0.972) & (-0.005) & (2.047) \end{array}$$

We expect the signs for $\beta_2, \beta_3, \beta_4$ and β_5 to be all positive. We expect the wheat yield to increase as technology improves and additional rainfall in each period should increase yield. The signs of b_2 and b_5 are as expected, but those for b_3 and b_4 are not. However, the t-statistics for testing significance of b_3 and b_4 are very small, indicating that both of them are not significantly different from zero. Interval estimates for β_3 and β_4 would include positive ranges. Thus, although b_3 and b_4 are negative, positive values of β_3 and β_4 are not in conflict with the data.

(b) We want to test $H_0: \beta_3 = \beta_4,\ \beta_3 = \beta_5$ against the alternative $H_1: \beta_3, \beta_4$ and β_5 are not all equal. The value of the F test statistic is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)} = \frac{(4.863664 - 4.303504)/1}{4.303504/(48-5)} = 2.7985$$

The corresponding p-value is 0.072. Also, the critical value for a 5% significance level is $F_{(0.95,2,43)} = 3.214$. Since the F-value is less than the critical value (and the p-value is greater than 0.05), we do not reject H_0. The data do not reject the notion that the response of yield is the same irrespective of whether the rain falls during germination, development or flowering.

(c) The estimated model under the restriction is

$$\hat{Y}_t = 0.6515 + 0.0314t + 0.0138RG_t + 0.0138RD_t + 0.0138RF_t$$
$$\begin{array}{lccccc} (\text{se}) & (0.2679) & (0.0035) & (0.0567) & (0.0567) & (0.0567) \\ (t) & (2.432) & (8.89) & (0.2443) & (0.2443) & (0.2443) \end{array}$$

With the restrictions imposed the signs of all the estimates are as expected. However, the response estimates for rainfall in all periods are not significantly different from zero. One possibility for improving the model is the inclusion of quadratic effects of rainfall in each period. That is, the squared terms RG_t^2, RD_t^2 and RF_t^2 could be included in the model. These terms could capture a declining marginal effect of rainfall. See Chapter 7.

EXERCISE 6.14

(a) The estimated model is

$$\widehat{HW} = -8.1236 + 2.1933HE + 0.1997HA \qquad R^2 = 0.1655$$
$$\text{(se)} \quad (4.1583) \; (0.1801) \qquad (0.0675)$$
$$(t) \quad (-1.954) \; (12.182) \qquad (2.958)$$

An increase of one year of a husband's education leads to a \$2.19 increase in wages. Also, older husbands earn 20 cents more on average per year of age, other things equal.

(b) A RESET test with one term yields $F = 9.528$ with *p*-value = 0.0021, and with two terms $F = 4.788$ and *p*-value = 0.0086. Both *p*-values are smaller than a significance level of 0.05, leading us to conclude that the linear model suggested in part (a) is not adequate.

(c) The estimated equation is:

$$\widehat{HW} = -45.5675 - 1.4580HE + 0.1511HE^2 + 2.8895HA - 0.0301HA^2 \qquad R^2 = 0.1918$$
$$\text{(se)} \quad (17.5436) \; (1.1228) \qquad (0.0458) \qquad (0.7329) \qquad (0.0081)$$
$$(t) \quad (-2.597) \; (-1.298) \qquad (3.298) \qquad (3.943) \qquad (-3.703)$$

Wages are now quadratic functions of age and education. The effects of changes in education and in age on wages are given by the partial derivatives

$$\frac{\partial \widehat{HW}}{\partial HE} = -1.4580 + 0.3022HE \qquad\qquad \frac{\partial \widehat{HW}}{\partial HA} = 2.8895 - 0.0602HA$$

The first of these two derivatives suggests that the wage rate declines with education up to an education level of $HE_{\min} = 1.458/0.30522 = 4.8$ years, and then increases at an increasing rate. A negative value of $\partial HW/\partial HE$ for low values of *HE* is not realistic. Only 7 of the 753 observations have education levels less than 4.8, so the estimated relationship might not be reliable in this region. The derivative with respect to age suggests the wage rate increases with age, but at a decreasing rate, reaching a maximum at the age $HA_{\max} = 2.8895/0.06022 = 48$ years.

(d) A RESET test with one term yields $F = 0.326$ with *p*-value = 0.568, and with two terms $F = 0.882$ and *p*-value = 0.414. Both *p*-values are much larger than a significance level of 0.05. Thus, there is no evidence from the RESET test to suggest the model in part (c) is inadequate.

Exercise 6.14 (continued)

(e) The estimated model is:

$$\widehat{HW} = -37.0540 - 2.2076HE + 0.1688HE^2 + 2.6213HA$$
$$\text{(se)} \quad (17.0160) \quad (1.0914) \quad (0.0444) \quad (0.7101)$$
$$(t) \quad (-2.178) \quad (-2.023) \quad (3.800) \quad (3.691)$$

$$- 0.0278HA^2 + 7.9379CIT \qquad R^2 = 0.2443$$
$$(0.0079) \quad (1.1012)$$
$$(-3.525) \quad (7.208)$$

The wage rate in large cities is, on average, \$7.94 higher than it is outside those cities.

(f) The *p*-value for b_6, the coefficient associated with *CIT*, is 0.0000. This suggests that b_6 is significantly different from zero and *CIT* should be included in the equation. Note that when *CIT* was excluded from the equation in part (c), its omission was not picked up by RESET. The RESET test does not always pick up misspecifications.

(g) From part (c), we have

$$\frac{\partial \widehat{HW}}{\partial HE} = -1.4580 + 0.3022HE \qquad \frac{\partial \widehat{HW}}{\partial HA} = 2.8895 - 0.0602HA$$

and from part (f)

$$\frac{\partial \widehat{HW}}{\partial HE} = -2.2076 + 0.3376HE \qquad \frac{\partial \widehat{HW}}{\partial HA} = 2.6213 - 0.0556HA$$

Evaluating these expressions for $HE = 6$, $HE = 15$, $HA = 35$ and $HA = 50$ leads to the following results.

	$\partial HW/\partial HE$		$\partial HW/\partial HA$	
	$HE = 6$	$HE = 15$	$HA = 35$	$HA = 50$
Part (c)	0.356	3.076	0.781	–0.123
Part (e)	–0.182	2.855	0.678	–0.156

The omitted variable bias from omission of *CIT* does not appear to be severe. The remaining coefficients have similar signs and magnitudes for both parts (c) and (e), and the marginal effects presented in the above table are similar for both parts with the exception of $\partial HW/\partial HE$ for $HE = 6$ where the sign has changed. The likely reason for the absence of strong omitted variable bias is the low correlations between *CIT* and the included variables *HE* and *HA*. These correlations are given by $\text{corr}(CIT, HE) = 0.2333$ and $\text{corr}(CIT, HA) = 0.0676$.

EXERCISE 6.15

(a) The average price of a 40-year old house of size 3600 square feet is

$$PRICE_{(40,3600)} = \beta_1 + 3600\beta_2 + 40\beta_3$$

The average price of a 5-year old house of size 1800 square feet is

$$PRICE_{(5,1800)} = \beta_1 + 1800\beta_2 + 5\beta_3$$

The conjecture that we set up as the alternative hypothesis is

$$\beta_1 + 3600\beta_2 + 40\beta_3 > 2(\beta_1 + 1800\beta_2 + 5\beta_3)$$

Thus, after simplifying this inequality, the null and alternative hypotheses are

$$H_0 : -\beta_1 + 30\beta_3 \le 0 \qquad H_1 : -\beta_1 + 30\beta_3 > 0$$

The test statistic for testing H_0 is

$$t = \frac{-b_1 + 30b_3}{\text{se}(-b_1 + 30b_3)}$$

where

$$\text{se}(-b_1 + 30b_3) = \sqrt{\widehat{\text{var}(b_1)} + 900\widehat{\text{var}(b_3)} - 60\widehat{\text{cov}(b_1, b_3)}}$$

The values for these quantities and the test results for each house category are as follows.

	All houses	Town houses	French style
$-b_1 + 30b_3$	19296	–169063	291863
$\widehat{\text{var}(b_1)}$	48855007	130798354	1088235489
$\widehat{\text{var}(b_3)}$	19851	140902	12063976
$\widehat{\text{cov}(b_1, b_3)}$	–497267	–3248879	–15538629
$\text{se}(-b_1 + 30b_3)$	9826	21273	113482
t-value	1.964	–7.947	2.572
df	1077	67	94
5% critical value	1.646	1.668	1.661
p-value	0.0249	1.0000	0.0058
Decision	Reject H_0	Accept H_0	Reject H_0

For the all-house and French style categories, the data support the conjecture stated in the alternative hypothesis, namely, that $PRICE_{(40,3600)} > 2 \times PRICE_{(5,1800)}$. In the case of town houses, whose estimated equation suggests that they quickly depreciate, the alternative hypothesis is not supported.

Exercise 6.15 (continued)

(b) The average prices for the three houses are as follows.

(i) $PRICE_{(0,2000)} = \beta_1 + 2000\beta_2$

(ii) $PRICE_{(20,2200)} = \beta_1 + 2200\beta_2 + 20\beta_3$

(iii) $PRICE_{(40,2400)} = \beta_1 + 2400\beta_2 + 40\beta_3$

Setting $PRICE_{(0,2000)} = PRICE_{(20,2200)}$, gives

$$\beta_1 + 2000\beta_2 = \beta_1 + 2200\beta_2 + 20\beta_3$$

which can be simplified to

$$10\beta_2 + \beta_3 = 0$$

Setting $PRICE_{(0,2000)} = PRICE_{(40,2400)}$, gives

$$\beta_1 + 2000\beta_2 = \beta_1 + 2400\beta_2 + 40\beta_3$$

which can be simplified to

$$10\beta_2 + \beta_3 = 0$$

Thus, all three houses will be equally priced if $H_0: 10\beta_2 + \beta_3 = 0$ holds. The *F*-value for testing this null hypothesis against the alternative $H_1: 10\beta_2 + \beta_3 \neq 0$ is $F = 1.12$. The corresponding *p*-value is 0.29. Thus H_0 is not rejected. There is no evidence to suggest the houses are not equally priced.

Remark: In the first printing of *POE*, the third house was given as 40 years old with 2300 (not 2400) square feet. In this case, the null and alternative hypotheses are $H_0: \beta_2 = \beta_3 = 0$ and $H_1: \beta_2 \neq 0$ and/or $\beta_3 \neq 0$. The test values are $F = 773.6$ and *p*-value = 0.0000. The null hypothesis is rejected.

(c) The application of RESET tests to all houses, town houses and French style homes leads to rejection of the adequacy of the model $PRICE = \beta_1 + \beta_2 SQFT + \beta_3 AGE + e$ in all cases. The model might be improved by the inclusion of more variables such as type of neighborhood, and whether the house has particular attributes such as a view, a pool and a fireplace. Also, the functional form might be inadequate. Log-log or log-linear forms or the inclusion of quadratic terms might improve the model

EXERCISE 6.16

(a) A 1% increase in *GROWTH* leads to a change in *VOTE* of β_2. A 1% increase in *INFLATION* leads to a change in *VOTE* of β_3. Thus the change in *VOTE* from increasing both by 1% is $\beta_2 + \beta_3$. Thus, the null and alternative hypotheses are

$$H_0 : \beta_2 + \beta_3 = 0 \qquad \text{and} \qquad H_1 : \beta_2 + \beta_3 > 0$$

The rejection region for a 5% significance level is $t \geq t_{(0.95,28)} = 1.701$. The calculated value of the test statistic is

$$t = \frac{b_2 + b_3}{\text{se}(b_2 + b_3)} = \frac{0.64876 - 0.18622}{0.48685} = 0.95$$

where the standard error is computed from

$$\begin{aligned}\text{se}(b_2 + b_3) &= \sqrt{\widehat{\text{var}(b_2)} + \widehat{\text{var}(b_2)} + 2\widehat{\text{cov}(b_2, b_3)}} \\ &= \sqrt{0.028043 + 0.186606 + 2 \times 0.011186} \\ &= 0.48685\end{aligned}$$

Because $0.95 < 1.701$, H_0 is not rejected. Alternatively, H_0 is not rejected because its *p*-value of 0.175 is greater than 0.05. There is not enough evidence to suggest increasing *GROWTH* and *INFLATION* by 1% will improve Willie's *VOTE*.

(b) Willie will get reelected if $E(VOTE) = \beta_1 + 4\beta_2 + 5\beta_3 > 50$. Thus, the null and alternative hypotheses are

$$H_0 : \beta_1 + 4\beta_2 + 5\beta_3 \leq 50 \qquad \text{and} \qquad H_1 : \beta_1 + 4\beta_2 + 5\beta_3 > 50$$

The rejection region for a 5% significance level is $t \geq t_{(0.95,28)} = 1.701$. The calculated value of the test statistic is

$$t = \frac{b_1 + 4b_2 + 5b_3 - 50}{\text{se}(b_1 + 4b_2 + 5b_3)} = \frac{52.4436 + 4 \times 0.64876 - 5 \times 0.18622 - 50}{1.5340} = \frac{4.1075}{1.5340} = 2.68$$

where the standard error is computed from

$$\begin{aligned}&\text{se}(b_1 + 4b_2 + 5b_3) \\ &= \sqrt{\widehat{\text{var}(b_1)} + 4^2 \times \widehat{\text{var}(b_2)} + 5^2 \times \widehat{\text{var}(b_3)} + 8 \times \widehat{\text{cov}(b_1, b_2)} + 10 \times \widehat{\text{cov}(b_1, b_3)} + 40 \times \widehat{\text{cov}(b_2, b_3)}} \\ &= \sqrt{2.21523 + 16 \times 0.02804 + 25 \times 0.18661 - 8 \times 0.04551 - 10 \times 0.50594 + 40 \times 0.01119} \\ &= 1.5340\end{aligned}$$

Because $2.68 > 1.701$, H_0 is rejected. Alternatively, H_0 is rejected because its *p*-value of 0.006 is less than 0.05. The evidence suggests Willie will get reelected when $GROWTH = 4$ and $INFLATION = 5$.

EXERCISE 6.17

(a) The delay from a train is β_4 and the delay from a red light is β_3. Thus, the null and alternative hypotheses are

$$H_0: 3\beta_3 = \beta_4 \qquad \text{and} \qquad H_1: 3\beta_3 \neq \beta_4$$

The test can be performed with an F or a t statistic, with the critical value for the F-test being $F_{(0.95,1,227)} = 3.883$, and those for the t-test, $t_{(0.025,227)} = -1.970$ and $t_{(0.925,227)} = 1.970$. The rejection regions are $F > 3.883$ for the F-test, and $t < -1.970$ or $t > 1.970$ for the t-test. The calculated value of the t-test statistic is

$$t = \frac{3b_3 - b_4}{\text{se}(3b_3 - b_4)} = \frac{3 \times 1.3353 - 2.7548}{0.5205} = 2.404$$

where the standard error is computed from

$$\begin{aligned}\text{se}(3b_3 - b_4) &= \sqrt{9 \times \widehat{\text{var}}(b_3) + \widehat{\text{var}}(b_4) - 2 \times 3 \times \widehat{\text{cov}}(b_2, b_3)} \\ &= \sqrt{9 \times 0.019311 + 0.092298 + 6 \times 0.00081} \\ &= 0.5205\end{aligned}$$

The calculated value of the F-test statistic is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(3824.793 - 3729.870)/1}{3729.870/227} = 5.78$$

Note that $F = 5.78 = t^2 = 2.404^2$. The null hypothesis is rejected. Using the t-distribution rejection occurs because $2.404 > 1.970$. Using the F-distribution rejection occurs because $5.78 > 3.883$. In both cases the p-value is 0.017. The delay from a train is not equal to three times the delay from a red light.

(b) This test is similar to that in part (a), but it is a one-tail test rather than a two-tail test. The hypotheses are

$$H_0: \beta_4 \geq 3\beta_3 \qquad \text{and} \qquad H_1: \beta_4 < 3\beta_3$$

The rejection region for the t-test is $t < t_{(0.05,227)} = -1.652$, where the t-value is calculated as

$$t = \frac{b_4 - 3b_3}{\text{se}(b_4 - 3b_3)} = \frac{2.7548 - 3 \times 1.3353}{0.5205} = -2.404$$

Since $-2.404 < -1.652$, we reject H_0. The delay from a train is less than three times the delay from a red light.

Exercise 6.17 (continued)

(c) The delay from 3 trains is $3\beta_4$. The extra time gained by leaving 5 minutes earlier is $5+5\beta_2$. Thus, the hypotheses are

$$H_0: 3\beta_4 \le 5+5\beta_2 \qquad \text{and} \qquad H_0: 3\beta_4 > 5+5\beta_2$$

The rejection region for the *t*-test is $t > t_{(0.95,227)} = 1.652$, where the *t*-value is calculated as

$$t = \frac{3b_4 - 5b_2 - 5}{\text{se}(3b_4 - 5b_2)} = \frac{3\times 2.7548 - 5\times 0.36923 - 5}{0.9174} = 1.546$$

and the standard error is computed from

$$\begin{aligned}\text{se}(3b_4 - 5b_2) &= \sqrt{9\times \widehat{\text{var}(b_4)} + 25\times \widehat{\text{var}(b_2)} - 30\times \widehat{\text{cov}(b_2,b_4)}} \\ &= \sqrt{9\times 0.092298 + 25\times 0.000241 + 30\times 0.000165} \\ &= 0.9174\end{aligned}$$

Since $1.546 < 1.652$, we do not reject H_0 at a 5% significance level. Alternatively, we do not reject H_0 because the *p*-value = 0.0617, which is greater than 0.05. There is insufficient evidence to conclude that leaving 5 minutes earlier is not enough time.

(d) The expected time taken when the departure time is 7:15AM, and no red lights or trains are encountered, is $\beta_1 + 45\beta_2$. Thus, the null and alternative hypotheses are

$$H_0: \beta_1 + 45\beta_2 \le 45 \qquad \text{and} \qquad H_1: \beta_1 + 45\beta_2 > 45$$

The rejection region for the *t*-test is $t > t_{(0.95,227)} = 1.652$, where the *t*-value is calculated as

$$t = \frac{b_1 + 45b_2 - 45}{\text{se}(b_1 + 45b_2)} = \frac{19.9166 + 45\times 0.36923 - 45}{1.1377} = -7.44$$

and the standard error is computed from

$$\begin{aligned}\text{se}(b_1 + 45b_2) &= \sqrt{\widehat{\text{var}(b_1)} + 45^2 \times \widehat{\text{var}(b_2)} + 90\times \widehat{\text{cov}(b_1,b_2)}} \\ &= \sqrt{1.574617 + 2025\times 0.00024121 - 90\times 0.00854061} \\ &= 1.1377\end{aligned}$$

Since $-7.44 < 1.652$, we do not reject H_0 at a 5% significance level. Alternatively, we do not reject H_0 because the *p*-value = 1.000, which is greater than 0.05. There is insufficient evidence to conclude that Bill will not get to the University before 8:00AM.

Exercise 6.17 (continued)

(e) The predicted time it takes Bill to reach the University is

$$\widehat{TIME} = b_1 + b_2 \times 30 + b_3 \times 6 + b_4 \times 1 = 41.76$$

Using suitable computer software, the standard error of the forecast error can be calculated as $\text{se}(f) = 4.0704$. Thus, a 95% interval estimate for the travel time is

$$\widehat{TIME} \pm t_{(0.975,227)} \text{se}(f) = 41.76 \pm 1.97 \times 4.0704 = (33.74, 49.78)$$

Rounding this interval to 34 – 50 minutes, a 95% interval estimate for Bill's arrival time is from 7:34AM to 7:50AM.

EXERCISE 6.18

(a) We are testing the null hypothesis $H_0:\beta_2 = \beta_3$ against the alternative $H_1:\beta_2 \neq \beta_3$. The test can be performed with an F or a t statistic. Using an F-test, we reject H_0 when $F > F_{(0.95,1,348)}$, where $F_{(0.95,1,348)} = 3.868$. The calculated F-value is 0.342. Thus we do not reject H_0 because $0.342 < 3.868$. Also, the p-value of the test is 0.559, confirming non-rejection of H_0. The hypothesis that the land and labor elasticities are equal cannot be rejected at a 5% significance level.

Using a t-test, we reject H_0 when $t > t_{(0.975,348)}$ or $t < t_{(0.025,348)}$ where $t_{(0.975,348)} = 1.967$ and $t_{(0.025,348)} = -1.967$. The calculated t-value is

$$t = \frac{b_2 - b_3}{\text{se}(b_2 - b_3)} = \frac{0.36174 - 0.43285}{0.12165} = -0.585$$

In this case H_0 is not rejected because $-1.967 < -0.585 < 1.967$. The p-value of the test is 0.559. The hypothesis that the land and labor elasticities are equal cannot be rejected at a 5% significance level.

(b) We are testing the null hypothesis $H_0:\beta_2 + \beta_3 + \beta_4 = 1$ against the alternative $H_1:\beta_2 + \beta_3 + \beta_4 \neq 1$, using a 10% significance level. The test can be performed with an F or a t statistic. Using an F-test, we reject H_0 when $F > F_{(0.90,1,348)}$, where $F_{(0.90,1,348)} = 2.72$. The calculated F-value is 0.0295. Thus, we do not reject H_0 because $0.0295 < 2.72$. Also, the p-value of the test is 0.864, confirming non-rejection of H_0. The hypothesis of constant returns to scale cannot be rejected at a 10% significance level.

Using a t-test, we reject H_0 when $t > t_{(0.95,348)}$ or $t < t_{(0.05,348)}$ where $t_{(0.95,348)} = 1.649$ and $t_{(0.05,348)} = -1.649$. The calculated t-value is

$$t = \frac{b_2 + b_3 + b_4 - 1}{\text{se}(b_2 + b_3 + b_4)} = \frac{0.36174 + 0.43285 + 0.209502 - 1}{0.023797} = 0.172$$

In this case H_0 is not rejected because $-1.649 < 0.172 < 1.649$. The p-value of the test is 0.864. The hypothesis of constant returns to scale cannot be rejected at a 10% significance level.

Exercise 6.18 (continued)

(c) In this case the null and alternative hypotheses are

$$H_0:\begin{cases}\beta_2-\beta_3=0\\ \beta_2+\beta_3+\beta_4=1\end{cases} \qquad H_1:\begin{cases}\beta_2-\beta_3\neq 0 \quad \text{and/or}\\ \beta_2+\beta_3+\beta_4\neq 1\end{cases}$$

We reject H_0 when $F > F_{(0.95,2,348)}$, where $F_{(0.95,2,348)} = 3.02$. The calculated *F*-value is 0.183. Thus, we do not reject H_0 because $0.183 < 3.02$. Also, the *p*-value of the test is 0.833, confirming non-rejection of H_0. The joint null hypothesis of constant returns to scale and equality of land and labor elasticities cannot be rejected at a 5% significance level.

(d) The mean of log output when $AREA = 2$, $LABOR = 100$ and $FERT = 175$ is

$$\begin{aligned}E\left[\ln(PROD)\right] &= \beta_1+\beta_2\times\ln(2)+\beta_3\times\ln(100)+\beta_4\times\ln(175)\\ &= \beta_1+0.69315\beta_2+4.60517\beta_3+5.16479\beta_4\end{aligned}$$

Thus, the null and alternative hypotheses are

$$H_0:\beta_1+0.69315\beta_2+4.60517\beta_3+5.16479\beta_4=1.5$$

$$H_1:\beta_1+0.69315\beta_2+4.60517\beta_3+5.16479\beta_4\neq 1.5$$

We reject H_0 when $F > F_{(0.95,1,348)}$, where $F_{(0.95,1,348)} = 3.868$. The calculated *F*-value is 208. Thus, we reject H_0 because $208 > 3.868$. Also, the *p*-value of the test is less than 0.0001, confirming rejection of H_0. The hypothesis that the mean of log output is equal to 1.5 when the inputs are set at the specified levels is rejected.

EXERCISE 6.19

The results are summarized in the following table.

	Full model	*FERT* omitted	*LABOR* omitted	*AREA* omitted
b_2 (*AREA*)	0.3617	0.4567	0.6633	
b_3 (*LABOR*)	0.4328	0.5689		0.7084
b_4 (*FERT*)	0.2095		0.3015	0.2682
RESET(1) p-value	0.5688	0.8771	0.4281	0.1140
RESET(2) p-value	0.2761	0.4598	0.5721	0.0083

(i) With *FERT* omitted the elasticity for *AREA* changes from 0.3617 to 0.4567, and the elasticity for *LABOR* changes from 0.4328 to 0.5689. The RESET F-values (p-values) for 1 and 2 extra terms are 0.024 (0.877) and 0.779 (0.460), respectively. Omitting *FERT* appears to bias the other elasticities upwards, but the omitted variable is not picked up by the RESET test.

(ii) With *LABOR* omitted the elasticity for *AREA* changes from 0.3617 to 0.6633, and the elasticity for *FERT* changes from 0.2095 to 0.3015. The RESET F-values (p-values) for 1 and 2 extra terms are 0.629 (0.428) and 0.559 (0.572), respectively. Omitting *LABOR* also appears to bias the other elasticities upwards, but again the omitted variable is not picked up by the RESET test.

(iii) With *AREA* omitted the elasticity for *FERT* changes from 0.2095 to 0.2682, and the elasticity for *LABOR* changes from 0.4328 to 0.7084. The RESET F-values (p-values) for 1 and 2 extra terms are 2.511 (0.114) and 4.863 (0.008), respectively. Omitting *AREA* appears to bias the other elasticities upwards, particularly that for *LABOR*. In this case the omitted variable misspecification has been picked up by the RESET test with two extra terms.