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 $y = \beta_1 + \beta_2 x + \beta_3 z + e$. The *SSE* from estimating this model is 979.830. The model after augmenting with the squares and the cubes of predictions $\hat{y}^2$ and $\hat{y}^3$ is $y = \beta_1 + \beta_2 x + \beta_3 z + \gamma_1 \hat{y}^2 + \gamma_2 \hat{y}^3 + e$. The *SSE* from estimating this model is 696.5375. To use the RESET, 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

$$t = \frac{(2b_2 - b_3)}{\mathrm{se}(2b_2 - b_3)}$$

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

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

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

(a) The value of the *t* statistic for the significance tests is calculated from:

$$t = \frac{b_k}{\text{se}(b_k)}$$

We reject the null hypothesis $H_0:\beta_k = 0$ if $|t| > t_c = 2$. The *t*-values for each of the coefficients are given in the following table. Those which are significantly different from zero at an approximate 5% level are marked *. When *EDUC* and $EDUC^2$ both appear in an equation, their coefficients are not significantly different from zero, with the exception of eqn (B), where $EDUC^2$ is significant. In addition, the interaction term between *EXPER* and *EDUC* is not significant in eqn (A).

Variable		*t*-values[a]				
		Eqn (A)	Eqn (B)	Eqn (C)	Eqn (D)	Eqn (E)
C	β_1	3.97*	6.59*	8.38*	23.82*	9.42*
EDUC	β_2	1.26	0.84	1.04		15.90*
$EDUC^2$	β_3	1.89	2.12*	1.73		
EXPER	β_4	4.58*	6.28*		5.17*	6.11*
$EXPER^2$	β_5	–5.38*	–5.31*		–4.90*	–5.13*
*EXPER*EDUC*	β_6	–1.06				
HRSWK	β_7	8.34*	8.43*	9.87*	10.11*	8.71*

[a] Note: These *t*-values were obtained from the computer output. Some of them do not agree exactly with the *t* ratios obtained using the coefficients and standard errors in Table 6.4. Rounding error discrepancies arise because of rounding in the reporting of values in Table 6.4.

(b) Using the labeling of coefficients in the above table, we see that the restriction imposed on eqn (A) that gives eqn (B) is $\beta_6 = 0$. The *F*-test value for testing $H_0:\beta_6 = 0$ against $H_1:\beta_6 \neq 0$ can be calculated from restricted and unrestricted sums of squared errors as follows:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(222.6674 - 222.4166)/1}{222.4166/993} = 1.120$$

The corresponding *p*-value is 0.290. The critical value at the 5% significance level is $F_{(0.95,1,993)} = 3.851$. Since the *F*-value is less than the critical value (or the *p*-value is greater than 0.05), we fail to reject the null hypothesis and conclude that the interaction term, $EDUC \times EXPER$ is not significant in determining the wage.

The *t*-value for testing $H_0:\beta_6 = 0$ against $H_1:\beta_6 \neq 0$ is –1.058. At the 5% level, its absolute value is less than the critical value, $t_{(0.975,993)} = 1.962$. Thus, the *t*-test gives the same result. The two tests are equivalent because $\sqrt{1.120} = 1.058$ and $\sqrt{3.851} = 1.962$.

Exercise 6.4 (continued)

(c) The restrictions imposed on eqn (A) that give eqn (C) are $\beta_4 = 0,\ \beta_5 = 0$ and $\beta_6 = 0$. Thus, we test

$$H_0: \beta_4 = 0,\ \beta_5 = 0 \text{ and } \beta_6 = 0$$

$$H_1: \text{ At least one of } \beta_4 \text{ or } \beta_5 \text{ or } \beta_6 \text{ is nonzero }.$$

The *F*-value is calculated from:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(233.8317 - 222.4166)/3}{222.4166/993} = 16.988$$

The corresponding *p*-value is 0.0000. The critical value at a 5% significance level is $F_{(0.95,3,993)} = 2.614$. Since the *F*-value is greater than the critical value (or the *p*-value is less than 0.05), we reject the null hypothesis and conclude at least one of β_4 or β_5 or β_6 is nonzero.

By performing this test, we are asking whether experience is relevant for determining the wage level. All three coefficients relate to variables that include *EXPER*. The test outcome suggests that experience is indeed a relevant variable.

(d) The restrictions imposed on eqn (B) that give eqn (D) are $\beta_2 = 0$ and $\beta_3 = 0$. Thus, we test

$$H_0: \beta_2 = 0,\ \beta_3 = 0$$

$$H_1: \text{ At least one of } \beta_2 \text{ or } \beta_3 \text{ is nonzero }.$$

The *F*-value is calculated from:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(280.5061 - 222.6674)/2}{222.6674/994} = 129.1$$

The corresponding *p*-value is 0.0000. The critical value at a 5% significance level is $F_{(0.95,2,994)} = 3.005$. Since the *F*-value is greater than the critical value (or the *p*-value is less than 0.05), we reject the null hypothesis and conclude at least one of β_2 or β_3 is nonzero.

By performing this test, we are asking whether education is relevant for determining the wage level. Both coefficients relate to variables that include *EDUC*. The test outcome suggests that education is indeed a relevant variable.

Exercise 6.4 (continued)

(e) The restrictions imposed on eqn (A) that give eqn (E) are $\beta_3 = 0$ and $\beta_6 = 0$. Thus, we test

$$H_0 : \beta_3 = 0,\ \beta_6 = 0$$

$$H_1 : \text{ At least one of } \beta_3 \text{ or } \beta_6 \text{ is nonzero }.$$

The F-value is calculated from:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(223.6716 - 222.4166)/2}{222.4166/993} = 2.802$$

The corresponding p-value is 0.0612. The critical value at a 5% significance level is $F_{(0.95,2,993)} = 3.005$. Since the F-value is less than the critical value (or the p-value is greater than 0.05), we do not reject the null hypothesis. The assumption $\beta_3 = 0,\ \beta_6 = 0$ is compatible with the data.

By performing this test, we are asking whether it is sufficient to include education as a linear term or whether we should also include it as a quadratic and/or interaction term. The test outcome suggests that including it as a linear term is adequate.

(f) Eqn (E) is the preferred model. All its estimated coefficients are significantly different from zero. It includes both $EXPER$ and $EXPER^2$ which were shown to be jointly significant, and it excludes the interaction term and $EDUC^2$ which, jointly, were not significant.

(g) The AIC for eqn (D):

$$\text{AIC}_\text{D} = \ln\left(\frac{SSE}{N}\right) + \frac{2K}{N} = \ln\left(\frac{280.5061}{1000}\right) + \frac{8}{1000} = -1.263$$

The SC for eqn (A):

$$\text{SC}_\text{A} = \ln\left(\frac{SSE}{N}\right) + \frac{K\ln(N)}{N} = \ln\left(\frac{222.4166}{1000}\right) + \frac{7\times\ln(1000)}{1000} = -1.455$$

Eqn (B) is favored by the AIC criterion. Eqn (E) is favored by the SC criterion.

EXERCISE 6.5

(a) Education and experience will have the same effects on $\ln(WAGE)$ if $\beta_2 = \beta_4$ and $\beta_3 = \beta_5$. The null and alternative hypotheses are:

$$H_0 : \beta_2 = \beta_4 \text{ and } \beta_3 = \beta_5$$

$$H_1 : \beta_2 \neq \beta_4 \text{ or } \beta_3 \neq \beta_5 \text{ or both}$$

(b) The restricted model assuming the null hypothesis is true is

$$\ln(WAGE) = \beta_1 + \beta_4(EDUC + EXPER) + \beta_5(EDUC^2 + EXPER^2) + \beta_6 HRSWK + e$$

(c) The *F*-value is calculated from:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(254.1726 - 222.6674)/2}{222.6674/994} = 70.32$$

The corresponding *p*-value is 0.0000. Also, the critical value at a 5% significance level is $F_{(0.95,2,994)} = 3.005$. Since the *F*-value is greater than the critical value (or the *p*-value is less than 0.05), we reject the null hypothesis and conclude that education and experience have different effects on $\ln(WAGE)$.

EXERCISE 6.6

Consider, for example, the model

$$y = \beta_1 + \beta_2 x + \beta_3 z + e$$

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

$$y = \beta_1 + \beta_2 x + \beta_3 z + \gamma \hat{y} + e$$

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

EXERCISE 6.7

(a) Least squares estimation of $y = \beta_1 + \beta_2 x + \beta_3 w + e$ 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 should be included in the model. Additionally, the RESET based on the equation $y = \beta_1 + \beta_2 x + e$ 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 is inadequate.

(b) Let b_2^* be the least squares estimator for β_2 in the model that omits w. 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 is omitted from the equation.

(c) The high correlation between x and w 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 (the large omitted variable bias) and the insignificance of b_2 when both variables are included in the equation.

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

$$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT + \beta_4 ADVERT^2 + e$$

yields

$$SALES = \left(80 - 6\beta_2 - 1.9(1 - 3.8\beta_4) - 3.61\beta_4\right) + \beta_2 PRICE$$
$$+ \left(1 - 3.8\beta_4\right) ADVERT + \beta_4 ADVERT^2 + e$$

Rearranging this equation into a form suitable for estimation yields

$$\left(SALES - ADVERT - 78.1\right) = \beta_2\left(PRICE - 6\right) + \beta_4\left(3.61 - 3.8ADVERT + ADVERT^2\right) + e$$

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^2 s 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

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

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

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

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 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 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 we conclude that the log-log model better reflects the demand for beer.

EXERCISE 6.13

(a) The estimated model is

$$\hat{Y} = 0.6254 + 0.0302\,t - 0.0794\,RG - 0.0005\,RD + 0.3387\,RF \qquad R^2 = 0.6889$$
$$\begin{array}{llllll} \text{(se)} & (0.2582) & (0.0034) & (0.0817) & (0.0918) & (0.1654) \\ (t) & (2.422) & (8.785) & (-0.972) & (-0.005) & (2.047) \end{array}$$

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

$$\hat{Y} = 0.6515 + 0.0314\,t + 0.0138\,RG + 0.0138\,RD + 0.0138\,RF$$
$$\begin{array}{llllll} \text{(se)} & (0.2679) & (0.0035) & (0.0567) & (0.0567) & (0.0567) \\ (t) & (2.432) & (8.89) & (0.2443) & (0.2443) & (0.2443) \end{array}$$

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^2, RD^2 and RF^2 could be included in the model. These terms could capture a declining marginal effect of rainfall.

EXERCISE 6.14

(a) The estimated model is

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

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 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:

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

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

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

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 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:

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

$$\begin{array}{lll} -0.0278HA^2 & +7.9379CIT & R^2 = 0.2443 \\ (0.0079) & (1.1012) & \\ (-3.525) & (7.208) & \end{array}$$

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 $HE = 5$, $HE = 10$, $HA = 30$ and $HA = 40$ leads to the following results.

	$\partial HW/\partial HE$		$\partial HW/\partial HA$	
	$HE = 5$	$HE = 10$	$HA = 30$	$HA = 40$
Part (c)	0.053	1.564	1.084	0.482
Part (e)	-0.520	1.168	0.953	0.397

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 estimated model is:

$$\widehat{SPRICE} = 11154.3 + 10680.0 LIVAREA - 11.334 AGE - 15552.4 BEDS - 7019.30 BATHS$$
$$\text{(se)} \quad (6555.1) \quad (273.1) \quad (80.502) \quad (1970.0) \quad (2903.82)$$

All coefficients are significantly different from zero with the exception of that for *AGE*. The negative signs on *BEDS* and *BATHS* might be puzzling. Recall, however, that their coefficients measure the effects on price of adding more bedrooms or more bathrooms, while keeping *LIVAREA* constant. Taking space from elsewhere to add bedrooms or bathrooms might reduce the price.

(b) An estimate of the expected difference in prices is:

$$\begin{aligned} SPRICE_{AGE=5} - SPRICE_{AGE=15} &= b_3 \times 5 - b_3 \times 15 \\ &= b_3(5-15) \\ &= (-11.334) \times (-10) \\ &= 113.34 \end{aligned}$$

Holding other variables constant, on average the price of a 5-year old house is 90.67 dollars more than the price of a 15-year old house.

A 95% interval is given by:

$$\begin{aligned} &(SPRICE_{AGE=5} - SPRICE_{AGE=15}) \pm t_{(0.975,1495)} \times \text{se}(-10b_3) \\ &= 113.34 \pm 1.962 \times 10 \times 80.502 = (-1466{,}1692) \end{aligned}$$

With 95% confidence, we estimate that the average price difference between houses that are 5 and 15 years old lies between –$1466 and $1692. This interval is a relatively narrow one, but it is uninformative in the sense that the difference could be negative or positive.

(c) Given that the living area is measured in hundreds of square feet, the expected increase in price is estimated as:

$$\begin{aligned} SPRICE_{LIVAREA=25} - SPRICE_{LIVAREA=20} &= b_2 \times 25 - b_2 \times 20 \\ &= b_{23}(25-20) \\ &= (10680) \times (5) \\ &= 53400 \end{aligned}$$

Holding other variables constant, we estimate that extending the living area by 500 square feet will increase the price of the house by $53400.

The null and alternative hypotheses are $H_0{:}\, 5\beta_2 \le 25000$ and $H_1{:}\, 5\beta_2 > 25000$, that we write alternatively as $H_0{:}\, \beta_2 \le 5000$ and $H_1{:}\, \beta_2 > 5000$. (Note: In the first printing of the text, the wording of the question suggested the alternative hypothesis should be $H_1{:}\, \beta_2 \ge 5000$. Since a null hypothesis should always include an equality, we have change the hypotheses accordingly.)

Exercise 6.15(c) (continued)

At a 5% significance level we reject H_0 if $t > t_{(0.95,1495)} = 1.646$. The calculated *t*-value is

$$t = \frac{b_2 - 5000}{\text{se}(b_2)} = 20.798$$

The corresponding *p*-value is 0.0000. Since the *t*-value is greater than the critical value of 1.646 (or because the *p*-value is less than 0.05), we reject the null hypothesis and conclude that an increase in the price of the house is more than 25000 dollars.

(d) Adding a bedroom of size 500 square feet will change the expected price by $5\beta_2 + \beta_4$. Thus, an estimate of the price change is

$$5b_2 + b_4 \quad = 5 \times 10680 - 15552.4 = 37848$$

A 95% interval estimate of the price change is

$$(5b_2 + b_4) \quad \pm t_{(0.975,1495)}\text{se}(5b_2 + b_4) = 37848 \pm 1.962 \times 2009.8 = (33905{,}41790)$$

With 95% confidence, we estimate the price increase will be between \$33905 and \$41790.

The standard error can be found from computer software or from

$$\begin{aligned} \text{se}(5b_2 + b_4) \quad &= \sqrt{5^2\text{var}(b_2) + \text{var}(b_4) + 5 \times 2\text{cov}(b_2, b_4)} \\ &= \sqrt{25 \times 74610.43 + 3880922 - 10 \times 170680.2} \\ &= 2009.8 \end{aligned}$$

(e) A RESET with one term yields $F = 117.80$ with *p*-value = 0.0000, and with two terms $F = 73.985$ and *p*-value = 0.0000. 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 reasonable.

EXERCISE 6.16

(a) The estimated regression is:

$$\widehat{SPRICE} = 79755.7 + 2994.65 LIVAREA - 830.38 AGE - 11921.9 BEDS - 4971.06 BATHS$$
$$\text{(se)} \quad (8744.3) \quad (772.30) \qquad (197.78) \qquad (1972.1) \qquad (2797.37)$$

$$+169.09 LIVAREA^2 + 14.2326 AGE^2$$
$$(16.13) \qquad (3.3559)$$

(b) To see if $LIVAREA^2$ and AGE^2 are relevant variables, we test the hypotheses

$$H_0 : \beta_6 = 0, \beta_7 = 0$$

$$H_1 : \beta_6 \neq 0 \text{ and/or } \beta_7 \neq 0$$

The restricted *SSE* is that from Exercise 6.15(a): $SSE_R = 2.1111419 \times 10^{12}$. The unrestricted *SSE* is that from part (a), with $LIVAREA^2$ and AGE^2 included. The *F*-value is calculated as follow:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{\left(2.1111419 \times 10^{12} - 1.9434999 \times 10^{12}\right)/2}{1.9434999 \times 10^{12}/(1500-7)} = 64.4$$

The corresponding *p*-value is 0.0000. The critical value at a 5% significance level is 3.00. Since the *F*-value is larger than the critical value (or because the *p*-value is smaller than 0.05), we reject the null hypothesis and conclude that including $LIVAREA^2$ and AGE^2 has improved the model.

(c) (b) An estimate of the expected difference in prices is:

$$\begin{aligned}\widehat{SPRICE}_{AGE=2} - \widehat{SPRICE}_{AGE=10} &= \left(b_3 \times 2 + b_7 \times 2^2\right) - \left(b_3 \times 10 + b_7 \times 10^2\right) \\ &= -8b_3 - 96b_7 \\ &= -8 \times (-830.3785) - 96 \times 14.23261 \\ &= 5276.7\end{aligned}$$

Holding other variables constant, we estimate that the average price difference between a 2-year old house and a 10-year old house is \$5277.

Using $\text{se}\left(-8b_3 - 96b_7\right) = 1291.95$ from computer software, a 95% interval is:

$$\left(\widehat{SPRICE}_{AGE=2} - \widehat{SPRICE}_{AGE=10}\right) \pm t_{(0.975,1495)} \times \text{se}\left(-8b_3 - 96b_7\right)$$
$$= 5276.7 \pm 1.962 \times 1291.95 = (2741.9, 7811.5)$$

With 95% confidence, we estimate that the average price difference between houses that are 2 and 10 years old lies between \$2742 and \$7812. This interval is a relatively wide one, but a more realistic one than that obtained using the specification in Exercise 6.15.

Exercise 6.16 (continued)

(c) (c) An estimate of the expected increase in price is

$$\widehat{SPRICE}_{LIVAREA=22} - \widehat{SPRICE}_{LIVAREA=20} = \left(22b_2 + 22^2 b_6\right) - \left(20b_2 + 20^2 b_6\right)$$
$$= 2b_2 + 84b_6$$
$$= 2\times 2994.652 + 84\times 169.0916$$
$$= 20193$$

Holding other variables constant, we estimate that extending the living area by 200 square feet will increase the price of the house by $20,193.

The null and alternative hypotheses are

$$H_0: 2\beta_2 + 84\beta_6 \le 20000$$
$$H_1: 2\beta_2 + 84\beta_6 > 20000$$

(Note: In the first printing of the text, the wording of the question suggested the alternative hypothesis should be $H_1: 2\beta_2 + 84\beta_6 \ge 20000$. Since a null hypothesis should always include an equality, we have change the hypotheses accordingly.)

At a 5% significance level we reject H_0 if $t > t_{(0.95,1493)} = 1.646$. The calculated t-value is

$$t = \frac{(2b_2 + 84b_4) - 20000}{\text{se}(2b_2 + 84b_4)} = \frac{193.00}{534.55} = 0.361$$

The corresponding p-value is 0.3591. Since the t-value is less than the critical value of 1.646 (or because the p-value is greater than 0.05), we fail to reject the null hypothesis and conclude that there is not sufficient evidence to show that the increase in the price of the house will be more than 20,000 dollars.

This test outcome is opposite to the conclusion reached in Exercise 6.15. It shows that test conclusions can be sensitive to the model specification.

(c) (d) Adding a bedroom of size 200 square feet will change the expected price by

$$\left(20\beta_2 + 20^2\beta_6 + \beta_4(BEDS+1)\right) - \left(18\beta_2 + 18^2\beta_6 + \beta_4 BEDS\right) = 2\beta_2 + 76\beta_6 + \beta_4$$

Thus, an estimate of the price change is

$$2b_2 + 76b_6 + b_4 = 2\times 2994.652 + 76\times 169.0916 - 11921.92 = 6918.3$$

A 95% interval estimate of the price change is

$$\left(2b_2 + 76b_6 + b_4\right) \pm t_{(0.975,1493)}\text{se}\left(2b_2 + 76b_6 + b_4\right)$$
$$= 6918.3 \pm 1.962\times 1802.468$$
$$= (3382, 10455)$$

With 95% confidence, the estimated price increase is between $3382 and $10,455.

Exercise 6.16 (continued)

(c) (e) A RESET with one term yields $F = 9.90$ with *p*-value = 0.0017; with two terms it yields $F = 32.56$ with *p*-value = 0.0000. Both *p*-values are smaller than a significance level of 0.05, leading us to conclude that the model with $LIVAREA^2$ and AGE^2 included is not adequate, despite being an improvement over the model in Exercise 6.15.

EXERCISE 6.17

(a) The estimated regression is

$$\underset{(\text{se})}{\ln(SPRICE)} = \underset{(0.0505)}{10.7453} + \underset{(0.004477)}{0.082609LIVAREA} - \underset{(0.00009629)}{0.00050364LIVAREA^2} - \underset{(0.0011799)}{0.0079785AGE}$$

$$+\underset{(0.00002001)}{0.00014110AGE^2} - \underset{(0.011316)}{0.075423BEDS}$$

(b) The null and alternative hypotheses are

$$H_0: \beta_2 = 0, \beta_3 = 0$$

$$H_1: \beta_2 \neq 0 \text{ or } \beta_3 \neq 0 \text{ or both are nonzero}$$

The *F*-value can be calculated as:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(177.9768 - 69.4625)/2}{69.4625/1494} = 1166.96$$

The corresponding *p*-value is 0.0000. Also, the critical value is $F_{(0.95,2,1494)} = 3.002$. Since the *F*-value is greater than the critical value (or because the *p*-value is less than 0.05), we reject the null hypothesis and conclude that living area helps explain the selling price.

(c) The null and alternative hypotheses are

$$H_0: \beta_4 = 0, \beta_5 = 0$$

$$H_1: \beta_4 \neq 0 \text{ or } \beta_5 \neq 0 \text{ or both are nonzero}.$$

The *F*-value can be calculated as:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(71.7908 - 69.4625)/2}{69.4625/1494} = 25.04$$

The corresponding *p*-value is 0.0000. The relevant critical value is 3.002. Since the *F*-value is greater than the critical value (or because the *p*-value is less than 0.05), we reject the null hypothesis and conclude that age of house helps explain the selling price.

Exercise 6.17 (continued)

(d) The predicted price using the natural predictor is:

$$\begin{aligned} SPRICE_n &= \exp(10.74528 + 0.082609 LIVAREA - 0.000503644 LIVAREA^2 \\ &\qquad -0.0079785 AGE + 0.00014110 AGE^2 - 0.075423 BEDS \\ &= \exp(10.74528 + 0.082609 \times 20 - 0.000503644 \times 20^2 \\ &\qquad -0.0079785 \times 5 + 0.00014110 \times 5^2 - 0.075423 \times 3 \\ &= 152264 \end{aligned}$$

The predicted price using the corrected predictor is:

$$SPRICE_c = SPRICE_n \exp(\hat{\sigma}^2/2) = 152264 \times \exp(0.0464941/2) = 155845$$

(e) To find a 95% prediction interval for *SPRICE*, we first find such an interval for ln(*SPRICE*)

$$\begin{aligned} \ln(SPRICE) \pm t_{(0.975,1494)} se(f) &= 11.9337 \pm 1.962 \times 0.216071 \\ &= (11.509533, 12.357203) \end{aligned}$$

which yields the following prediction interval for *SPRICE*

$$\big(\exp(11.509533), \exp(12.357203)\big) = (99661, 232630)$$

With 95% confidence, we predict that the selling price of a house with the specified characteristics will lie between \$99,661 and \$232,630.

The standard error of the forecast error for $\ln(SPRICE)$, $\text{se}(f) = 0.216071$, was found using computer software.

(f) Using the natural predictor, the estimated price of Wanling's house after the extension is

$$\begin{aligned} SPRICE_n &= \exp(10.74528 + 0.082609 \times 25 - 0.000503644 \times 25^2 \\ &\qquad -0.0079785 \times 10 + 0.00014110 \times 10^2 - 0.075423 \times 3 \\ &= 199543 \end{aligned}$$

Exercise 6.17 (continued)

(g) Ignoring the error term, the increase in price of the house is given by

$$
\begin{aligned}
SPRICE_{LIVARE=25} - SPRICE_{LIVAREA=23} &= \text{-}\exp(\beta_1 + 25\beta_2 + 25^2\beta_3 + 10\beta_4 + 10^2\beta_5 + 3\beta_6) \\
&\quad \text{-}\exp(\beta_1 + 23\beta_2 + 23^2\beta_3 + 10\beta_4 + 10^2\beta_5 + 3\beta_6) \\
&= \exp(\beta_1 + 10\beta_4 + 100\beta_5 + 3\beta_6)[\exp(25\beta_2 + 625\beta_3) - \exp(23\beta_2 + 529\beta_3)]
\end{aligned}
$$

Let $\text{g}(\beta) = \exp(\beta_1 + 10\beta_4 + 100\beta_5 + 3\beta_6)[\exp(25\beta_2 + 625\beta_3) - \exp(23\beta_2 + 529\beta_3)]$. Then, the null and alternative hypotheses are

$$H_0\text{:g}(\beta) \le 25000 \qquad\qquad H_1\text{:g}(\beta) > 25000$$

(Note: In the first printing of the text, the wording of the question suggested the alternative hypothesis should be $H_1\text{:g}(\beta) \ge 25000$. Since a null hypothesis should always include an equality, we have change the hypotheses accordingly.)

At a 5% significance level we reject H_0 if $t > t_{(0.95,1494)} = 1.645$. The calculated *t*-value is

$$t = \frac{\text{g}(b) - 25000}{\text{se}[\text{g}(b)]} = \frac{-2990.934}{724.836} = -4.126$$

The corresponding *p*-value is 1.0000. Since the *t*-value is less than the critical value of 1.645 (or because the *p*-value is greater than 0.05), we fail to reject the null hypothesis and conclude that there is not sufficient evidence to show that the increase in the price of the house will be more than $25,000.

The standard error $\text{se}[\text{g}(b)] = 724.836$ was found using computer software that utilized the delta method since $g(b)$ is a nonlinear function.

A comparison of this test result to that from similar tests in Exercises 6.15 and 6.16 illustrates the sensitivity of test results to model specification

(h) A RESET with one term yields $F = 0.968$ with *p*-value = 0.3254; using two terms yields $F = 0.495$ with *p*-value = 0.6094. Both *p*-values are larger than a significance level of 0.05, leading us to conclude that the model suggested in part (a) is a reasonable specification. This conclusion is in contrast to those from similar tests in Exercises 6.15 and 6.16. It appears that the log specification is a better model than the linear and quadratic ones considered earlier.

EXERCISE 6.18

(a) The estimated regression is:

$$\begin{aligned}
\widehat{\ln(SPRICE)} &= \underset{(0.2408)}{10.3149} + \underset{(0.02125)}{0.12680}LIVAREA - \underset{(0.0005148)}{0.0012677}LIVAREA^2 - \underset{(0.007373)}{0.016916}AGE \quad \text{(se)} \\
&+ \underset{(0.00012498)}{0.00029391}AGE^2 + \underset{(0.071877)}{0.062799}BEDS - \underset{(0.005844)}{0.013812}(LIVAREA \times BEDS) \\
&+ \underset{(0.00013163)}{0.00024011}(LIVAREA^2 \times BEDS) + \underset{(0.0021610)}{0.0026419}(AGE \times BEDS) \\
&- \underset{(0.000036997)}{0.000045123}(AGE^2 \times BEDS)
\end{aligned}$$

The estimated relationships for 2, 3 and 4 bedroom houses are as follows:

	$BEDS = 2$	$BEDS = 3$	$BEDS = 4$
C	10.4405	10.5033	10.5661
$LIVAREA$	0.099175	0.085363	0.071550
$LIVAREA^2$	−0.00078751	−0.00054740	−0.00030730
AGE	−0.0116321	−0.0089902	−0.0063483
AGE^2	0.00020366	0.00015854	0.00011342

(b) The null and alternative hypotheses are

$$H_0: \beta_6 = 0,\ \beta_8 = 0,\ \beta_9 = 0,\ \beta_{10} = 0$$

$$H_1: \text{ At least one of } \beta_6,\ \beta_8,\ \beta_9 \text{ and } \beta_{10} \text{ is nonzero}$$

The value of F-statistic is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(69.24671 - 69.02920)/4}{69.02920/1490} = 1.174$$

The corresponding p-value is 0.3205. Also, the critical value is $F_{(0.95,4,1490)} = 2.378$. Since the F-value is less than the critical value (or because the p-value is greater than 0.05), we do not reject the null hypothesis at the 5% level, and conclude that β_6, β_8, β_9 and β_{10} are jointly not significantly different from zero. This results suggests that the number of bedrooms effects the price only through its interaction with the living area.

Exercise 6.18(continued)

(c) The estimated regression is:

$$\widehat{\ln(PRICE)} = \underset{(0.0479)}{10.5518} + \underset{(0.004903)}{0.090116} LIVAREA - \underset{(0.00009426)}{0.00034819} LIVAREA^2 - \underset{(0.0011784)}{0.0080479} AGE$$

(se)

$$+ \underset{(0.00001998)}{0.00014243} AGE^2 - \underset{(0.0005695)}{0.0039957} (LVAREA \times BEDS)$$

The estimated relationships for 2, 3 and 4 bedroom houses are as follows:

	$BEDS = 2$	$BEDS = 3$	$BEDS = 4$
C	10.5518	10.5518	10.5518
LIVAREA	0.082125	0.078129	0.074133
$LIVAREA^2$	–0.00034819	–0.00034819	–0.00034819
AGE	–0.0080479	–0.0080479	–0.0080479
AGE^2	0.00014243	0.00014243	0.00014243

In this case only the coefficient of *LIVAREA* changes with the number of bedrooms.

(d) The AIC and SC values for the two models are:

Model in part (a): $AIC = -3.065$ $SC = -3.030$

Model in part (c) $AIC = -3.068$ $SC = -3.046$

Thus, the model in part (c) is favored by both the AIC and the SC.

EXERCISE 6.19

(a) The predicted time it takes Bill to reach the University if he leaves at 7:15AM is

$$
\begin{aligned}
TIME &= b_1 + b_2 \times 45 + b_3 \times 4 + b_4 \times 1 \\
&= 19.9166 + 0.369227 \times 45 + 1.33532 \times 4 + 2.75483 \\
&= 44.628
\end{aligned}
$$

Using suitable computer software, the standard error of the forecast error can be calculated as $se(f) = 4.0996$. Thus, a 95% interval estimate for the travel time is

$$TIME \quad \pm t_{(0.975,227)}\text{se}(f) = 44.628 \pm 1.970 \times 4.0996 = (36.55,52.71)$$

Rounding this interval to 37 – 53 minutes, a 95% interval estimate for Bill's arrival time is from 7:52AM to 8:08AM.

(b) The predicted time it takes Bill to reach the University if he leaves at 7:45AM is

$$
\begin{aligned}
TIME &= b_1 + b_2 \times 75 + b_3 \times 10 + b_4 \times 2 \\
&= 19.9166 + 0.369227 \times 75 + 1.33532 \times 10 + 2.75483 \times 2 \\
&= 66.471
\end{aligned}
$$

Using suitable computer software, the standard error of the forecast error can be calculated as $\text{se}(f) = 4.1459$. Thus, a 95% interval estimate for the travel time is

$$TIME \quad \pm t_{(0.975,227)}\text{se}(f) = 66.471 \pm 1.970 \times 4.1459 = (58.30,74.64)$$

Rounding this interval to 58– 75 minutes, a 95% interval estimate for Bill's arrival time is from 8:43AM to 9:00AM.

EXERCISE 6.20

(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 $F>F_{(0.90,1,348)}=2.72$. 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 is not rejected at a 10% significance level.

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

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

We reject H_0 when $F>F_{(0.95,2,348)}=3.02$. 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.

Exercise 6.20 (continued)

(d) The restricted model for part (a) where $\beta_2 = \beta_3$ is

$$\ln(PROD) = \beta_1 + \beta_2 \ln(AREA \times LABOR) + \beta_4 \ln(FERT) + e$$

The restricted model for part (b) where $\beta_2 + \beta_3 + \beta_4 = 1$ is

$$\ln(PROD) = \beta_1 + \beta_2 \ln(AREA) + \left(1 - \beta_2 - \beta_4\right)\ln(LABOR) + \beta_4 \ln(FERT) + e$$

or,

$$\ln\left(\frac{PROD}{LABOR}\right) = \beta_1 + \beta_2 \ln\left(\frac{AREA}{LABOR}\right) + \beta_4 \ln\left(\frac{FERT}{LABOR}\right) + e$$

The restricted model for part (c) where $\beta_2 = \beta_3$ and $\beta_2 + \beta_3 + \beta_4 = 1$ is

$$\ln\left(\frac{PROD}{FERT}\right) = \beta_1 + \beta_2 \ln\left(\frac{AREA \times LABOR}{FERT^2}\right) + e$$

The estimates and (standard errors) from these restricted models, and the unrestricted model, are given in the following table. Because the unrestricted estimates almost satisfy the restriction $\beta_2 + \beta_3 + \beta_4 = 1$, imposing this restriction changes the unrestricted estimates and their standard errors very little. Imposing the restriction $\beta_2 = \beta_3$ has an impact, changing the estimates for both β_2 and β_3, and reducing their standard errors considerably. Adding $\beta_2 + \beta_3 + \beta_4 = 1$ to this restriction reduces the standard errors even further, leaving the coefficient estimates essentially unchanged.

	Unrestricted	$\beta_2 = \beta_3$	$\beta_2 + \beta_3 + \beta_4 = 1$	$\beta_2 = \beta_3$ $\beta_2 + \beta_3 + \beta_4 = 1$
C	–1.5468 (0.2557)	–1.4095 (0.1011)	–1.5381 (0.2502)	–1.4030 (0.0913)
$\ln(AREA)$	0.3617 (0.0640)	0.3964 (0.0241)	0.3595 (0.0625)	0.3941 (0.0188)
$\ln(LABOR)$	0.4328 (0.0669)	0.3964 (0.0241)	0.4299 (0.0646)	0.3941 (0.0188)
$\ln(FERT)$	0.2095 (0.0383)	0.2109 (0.0382)	0.2106 (0.0377)	0.2118 (0.0376)
SSE	40.5654	40.6052	40.5688	40.6079

EXERCISE 6.21

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.

(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.

(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 with two extra terms.

EXERCISE 6.22

The model for parts (a) and (b) is

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + \beta_4 (AGE \times INCOME) + e$$

(a) The hypotheses are

$$H_0: \beta_2 = \beta_4 = 0 \qquad \text{and} \qquad H_1: \beta_2 \neq 0 \text{ and/or } \beta_4 \neq 0$$

The value of the F statistic under the assumption that H_0 is true is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N - K)} = \frac{(819286 - 580609)/2}{580609/36} = 7.40$$

The 5% critical value for (2, 36) degrees of freedom is F_c = 3.26 and the p-value of the test is 0.002. Thus, we reject H_0 and conclude that age does affect pizza expenditure.

(b) The marginal propensity to spend on pizza is given by

$$\frac{\partial E(PIZZA)}{\partial INCOME} = \beta_3 + \beta_4 AGE$$

Point estimates, standard errors and 95% interval estimates for this quantity, for different ages, are given in the following table.

Age	Point Estimate	Standard Error	Confidence Interval Lower	Confidence Interval Upper
20	4.515	1.520	1.432	7.598
30	3.283	0.905	1.448	4.731
40	2.050	0.465	1.107	2.993
50	0.818	0.710	–0.622	2.258
55	0.202	0.991	–1.808	2.212

The interval estimates were calculated using $t_c = t_{(0.975,36)} = 2.0281$.

The point estimates for the marginal propensity to spend on pizza decline as age increases, as we would expect. However, the confidence intervals are relatively wide indicating that our information on the marginal propensities is not very reliable. Indeed, all the confidence intervals do overlap.

Exercise 6.22 (continued)

(c) This model is given by

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INC + \beta_4 AGE \times INC + \beta_5 AGE^2 \times INC + e$$

The marginal effect of income is now given by

$$\frac{\partial E(PIZZA)}{\partial INCOME} = \beta_3 + \beta_4 AGE + \beta_5 AGE^2$$

If this marginal effect is to increase with age, up to a point, and then decline, then $\beta_5 < 0$. The results are given in the table below. The sign of the estimated coefficient $b_5 = 0.0042$ did not agree with our expectation, but, with a *p*-value of 0.401, it was not significantly different from zero.

Variable	Coefficient	Std. Error	*t*-value	*p*-value
C	109.72	135.57	0.809	0.4238
AGE	–2.0383	3.5419	–0.575	0.5687
INCOME	14.0962	8.8399	1.595	0.1198
$AGE \times INCOME$	–0.4704	0.4139	–1.136	0.2635
$AGE^2 \times INCOME$	0.004205	0.004948	0.850	0.4012

(d) The marginal propensity to spend on pizza, in this case, is given by

$$\frac{\partial E(PIZZA)}{\partial INCOME} = \beta_3 + \beta_4 AGE + \beta_5 AGE^2$$

Point estimates, standard errors and 95% interval estimates for this quantity, for different ages, are given in the following table.

Age	Point Estimate	Standard Error	Confidence Interval Lower	Confidence Interval Upper
20	6.371	2.664	0.963	11.779
30	3.769	1.074	1.589	5.949
40	2.009	0.469	1.056	2.962
50	1.090	0.781	–0.496	2.675
55	0.945	1.325	–1.744	3.634

The interval estimates were calculated using $t_c = t_{(0.975,35)} = 2.0301$.

Exercise 6.22(d) (continued)

As in part (b), the point estimates for the marginal propensity to spend on pizza decline as age increases. There is no "life-cycle effect" where the marginal propensity increases up to a point and then declines. Again, the confidence intervals are relatively wide indicating that our information on the marginal propensities is not very reliable. The range of ages in the sample is 18-55. The quadratic function reaches a minimum at

$$AGE_{\min} = -\frac{0.4704}{2 \times 0.004205} = 55.93$$

Thus, for the range of ages in the sample, the relevant section of the quadratic function is that where the marginal propensity to spend on pizza is declining. It is decreasing at a decreasing rate.

(e) The *p*-values for separate *t* tests of significance for the coefficients of *AGE*, $AGE \times INCOME$, and $AGE^2 \times INCOME$ are 0.5687, 0.2635 and 0.4012, respectively. Thus, each of these coefficients is not significantly different from zero.

To perform a joint test of the significance of all three coefficients, we set up the hypotheses

$$H_0 : \beta_2 = \beta_4 = \beta_5 = 0$$

$$H_1 : \text{ At least one of } \beta_2, \beta_4 \text{ and } \beta_5 \text{ is nonzero}$$

The *F*-value is calculated as follows:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(819285.8 - 568869.2)/3}{568869.2/35} = 5.136$$

The corresponding *p*-value is 0.0048. Also, the critical value at the 5% significance level is $F_{(0.95,3,35)} = 2.874$. Since the *F*-value is greater than the critical value (or because the *p*-value is less than 0.05), we reject the null hypothesis and conclude at least one of β_2, β_4 and β_5 is nonzero. This result suggests that age is indeed an important variable for explaining pizza consumption, despite the fact each of the three coefficients was insignificant when considered separately. Collinearity is the likely reason for this outcome. We investigate it in part (f).

(f) Two ways to check for collinearity are (i) to examine the simple correlations between each pair of variables in the regression, and (ii) to examine the R^2 values from auxiliary regressions where each explanatory variable is regressed on all other explanatory variables in the equation. In the tables below there are 3 simple correlations greater than 0.94 for the regression in part (c) and 5 when $AGE^3 \times INC$ is included. The number of auxiliary regressions with R^2s greater than 0.99 is 3 for the regression in part (c) and 4 when $AGE^3 \times INC$ is included. Thus, collinearity is potentially a problem. Examining the estimates and their standard errors confirms this fact. In both cases there are no *t*-values which are greater than 2 and hence no coefficients are significantly different from zero. None of the coefficients are reliably estimated. In general, including squared and cubed variables can lead to collinearity if there is inadequate variation in a variable.

Exercise 6.22(f) (continued)

Simple Correlations

	AGE	$AGE \times INC$	$AGE^2 \times INC$	$AGE^3 \times INC$
INC	0.4685	0.9812	0.9436	0.8975
AGE		0.5862	0.6504	0.6887
$AGE \times INC$			0.9893	0.9636
$AGE^2 \times INC$				0.9921

R^2 Values from Auxiliary Regressions

LHS variable	R^2 in part (c)	R^2 in part (f)
INC	0.99796	0.99983
AGE	0.68400	0.82598
$AGE \times INC$	0.99956	0.99999
$AGE^2 \times INC$	0.99859	0.99999
$AGE^3 \times INC$		0.99994

EXERCISE 6.23

Coefficient estimates, standard errors, *t*-values, and *p*-values obtained for this model are given in the following table.

Variable	Coefficient	Std. Error	*t*-value	*p*-value
C	1.13408	0.33982	3.337	0.0009
$EDUC$	0.046418	0.036936	1.257	0.2092
$EDUC^2$	0.0026509	0.0011122	2.383	0.0173
$EXPER$	0.057775	0.009761	5.919	0.0000
$EXPER^2$	–0.0006946	0.0000882	–7.875	0.0000
$EDUC \times EXPER$	–0.0010256	0.0005092	–2.014	0.0442

(a) The percentage change in *WAGE* from an extra year of education is calculated from:

$$\frac{\partial \ln(WAGE)}{\partial EDUC} \times 100 = \left(\beta_2 + 2\beta_3 EDUC + \beta_6 EXPER\right) \times 100$$

The percentage change in *WAGE* from an extra year of experience is calculated from:

$$\frac{\partial \ln(WAGE)}{\partial EXPER} \times 100 = \left(\beta_4 + 2\beta_5 EXPER + \beta_6 EDUC\right) \times 100$$

(i) When $EDUC = 12$ and $EXPER = 2$,

$$\frac{\partial \ln(WAGE)}{\partial EDUC} = 0.046418 + 2 \times 0.0026509 \times 12 - 0.0010256 \times 2 = 0.10799$$

$$\mathrm{se}\left(\frac{\partial \ln(WAGE)}{\partial EDUC}\right) = 0.015433$$

Using $t_{(0.975,\,994)} = 1.9624$, a 95% interval estimate for $100 \times \partial \ln(WAGE) / \partial EDUC$ is

$$10.799 \pm 1.9624 \times 1.5433 = (7.77, 13.83)$$

(ii) When $EDUC = 12$ and $EXPER = 2$,

$$\frac{\partial \ln(WAGE)}{\partial EXPER} = 0.057775 + 2 \times (-0.0006946) \times 2 - 0.0010256 \times 12 = 0.04269$$

$$\mathrm{se}\left(\frac{\partial \ln(WAGE)}{\partial EXPER}\right) = 0.004983$$

A 95% interval estimate for $100 \times \partial \ln(WAGE) / \partial EXPER$ is

$$4.269 \pm 1.9624 \times 0.4983 = (3.29, 5.25)$$

Exercise 6.23(a) (continued)

(iii) When $EDUC = 16$ and $EXPER = 10$,

$$\frac{\partial \ln(WAGE)}{\partial EDUC} = 0.046418 + 2 \times 0.0026509 \times 12 - 0.0010256 \times 2 = 0.12099$$

$$\text{se}\left(\frac{\partial \ln(WAGE)}{\partial EDUC}\right) = 0.010762$$

Using $t_{(0.975,994)} = 1.9624$, a 95% interval estimate for $100 \times \partial \ln(WAGE)/\partial EDUC$ is

$$12.099 \pm 1.9624 \times 1.0762 = (9.99, 14.21)$$

(iv) When $EDUC = 16$ and $EXPER = 10$,

$$\frac{\partial \ln(WAGE)}{\partial EXPER} = 0.057775 + 2 \times (-0.0006946) \times 10 - 0.0010256 \times 16 = 0.027473$$

$$\text{se}\left(\frac{\partial \ln(WAGE)}{\partial EXPER}\right) = 0.003168$$

A 95% interval estimate for $100 \times \partial \ln(WAGE)/\partial EXPER$ is

$$2.7473 \pm 1.9624 \times 0.3168 = (2.13, 3.37)$$

These results suggest that the return to an extra year of education is greater than the return to an extra year of experience. Furthermore, the return to education increases with further education whereas the return to experience decreases with further experience.

(b) The null and alternative hypotheses are:

H_0: $\beta_2 + 24\beta_3 + 2\beta_6 = 0.1$ and $\beta_4 + 4\beta_{53} + 12\beta_6 = 0.04$
H_1: $\beta_2 + 24\beta_3 + 2\beta_6 \neq 0.1$ and/or $\beta_4 + 4\beta_{53} + 12\beta_6 \neq 0.04$

Using econometric software, the *F*-value and the *p*-value are computed as 0.20 and 0.8202, respectively. Since the *p*-value is larger than 0.05, we do not reject the null hypothesis. We conclude that, for 2 years of experience and 12 years of education, the data are compatible with the hypothesis that the return to an extra year of education is 10% and the return to an extra year of experience is 4%.

(c) The null and alternative hypotheses are:

H_0: $\beta_2 + 32\beta_3 + 10\beta_6 = 0.12$ and $\beta_4 + 20\beta_{53} + 16\beta_6 = 0.01$
H_1: $\beta_2 + 32\beta_3 + 10\beta_6 \neq 0.12$ and/or $\beta_4 + 20\beta_{53} + 16\beta_6 \neq 0.01$

Using econometric software, the *F*-value and the *p*-value are computed as 15.29 and 0.000, respectively. Since the *p*-value is smaller than 0.05, we reject the null hypothesis. We conclude that, for 10 years of experience and 16 years of education, the data are not compatible with the hypothesis that the return to an extra year of education is 12% and the return to an extra year of experience is 1%.

Exercise 6.23 (continued)

(d) The null and alternative hypotheses are:

$$H_0: \quad \beta_2 + 24\beta_3 + 2\beta_6 = 0.1,\ \ \beta_4 + 4\beta_5 + 12\beta_6 = 0.04$$
$$\beta_2 + 32\beta_3 + 10\beta_6 = 0.12 \ \text{ adn } \ \beta_4 + 20\beta_5 + 16\beta_6 = 0.01$$

H_1: At least one of the above equations does not hold

Using econometric software, the *F*-value and the *p*-value are computed as 79.08 and 0.000, respectively. Since the *p*-value is smaller than 0.05, we reject the null hypothesis. We conclude that the data are not compatible with the hypothesis that, for 2 years of experience and 12 years of education, the return to an extra year of education is 10% and the return to an extra year of experience is 4%, and for 10 years of experience and 16 years of education, the return to an extra year of education is 12% and the return to an extra year of experience is 1%.

(e) From the joint hypotheses in part (c), we have

$$\beta_2 = 0.12 - 32\beta_3 - 10\beta_6 \qquad \text{and} \qquad \beta_4 = 0.01 - 20\beta_{53} + 16\beta_6$$

Substituting these expressions into the original equation yields

$$\begin{aligned}\ln(WAGE) &= \beta_1 + (0.12 - 32\beta_3 + 10\beta_6)EDUC + \beta_3 EDUC^2 \\ &+(0.01 - 20\beta_5 - 16\beta_6)EXPER + \beta_5 EXPER^2 + \beta_6(EDUC \times EXPER) + e\end{aligned}$$

$$\begin{aligned}\ln(WAGE) - 0.12EDUC - 0.01EXPER &= \beta_1 + \beta_3(EDUC^2 - 32EDUC) \\ &+ \beta_5(EXPER^2 - 20EXPER) \\ &+ \beta_6(EDUC \times EXPER - 10EDUC - 16EXPER) + e\end{aligned}$$

Estimating the above model, and substituting into the restrictions to find estimates for β_2 and β_4 yields

Variable	Coefficient	Std. Error	*t*-value	*p*-value
C	1.447107	0.3113551	4.65	0.000
EDUC	0.0478522	0.0371698	1.29	0.198
$EDUC^2$	0.0025846	0.0010851	2.38	0.017
EXPER	0.032401	0.0064916	4.99	0.000
$EXPER^2$	–0.0002753	0.0000428	–6.44	0.000
$EDUC \times EXPER$	–0.001056	0.0003701	–2.85	0.004

To confirm the result in (c), we can manually calculate the *F*-value.

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(260.7611 - 252.9759)/2}{252.9759/994} = 15.295$$

EXERCISE 6.24

(a) β_2 is the direct price elasticity of sales of brand 1 with respect to changes in the price of brand 1. The expected sign of β_2 is negative. Holding other variables constant, a 1% increase in price per can of brand 1 changes brand 1's sales by $\beta_2\%$.

β_3 is the cross price elasticity of sales of brand 1 with respect to changes in the price of brand 2. The expected sign of β_3 is positive. Holding other variables constant, a 1% increase in price per can of brand 2 changes brand 1's sales by $\beta_3\%$.

β_4 is the cross price elasticity of sales of brand 1 with respect to changes in the price of brand 3. The expected sign of β_4 is positive. Holding other variables constant, a 1% increase in price per can of brand 3 changes brand 1's sales by $\beta_4\%$.

(b) The regression results are

Variable	Coefficient	Std. Error	*t*-value	*p*-value
C	7.8894	0.2514	31.376	0.0000
ln(*APR1*)	–4.6246	0.6383	–7.245	0.0000
ln(*APR2*)	0.9904	0.5338	1.855	0.0697
ln(*APR3*)	1.6871	0.7460	2.262	0.0283

All coefficients have the expected signs and all are significantly different from zero at a 5% level of significance with the exception of b_3 which is the coefficient of $\ln(APR2)$.

(c) If $\beta_2+\beta_3+\beta_4=0$, we can rewrite the regression equation as:

$$\begin{aligned}\ln(SAL1) &= \beta_1 + (-\beta_3-\beta_4)\ln(APR1) + \beta_3\ln(APR2) + \beta_4\ln(APR3) + e\\ &= \beta_1 + \beta_3\left[\ln(APR2)-\ln(APR1)\right] + \beta_4\left[\ln(APR2)-\ln(APR1)\right] + e\\ &= \beta_1 + \beta_3\ln\left(\frac{APR2}{APR1}\right) + \beta_4\ln\left(\frac{APR3}{APR1}\right) + e\\ &= \beta_1 - \beta_3\ln\left(\frac{APR1}{APR2}\right) - \beta_4\ln\left(\frac{APR1}{APR3}\right) + e\\ &= \alpha_1 + \alpha_2\ln\left(\frac{APR1}{APR2}\right) + \alpha_3\ln\left(\frac{APR1}{APR3}\right) + e\end{aligned}$$

where we have set $\alpha_1=\beta_1$, $\alpha_2=-\beta_3$ and $\alpha_3=-\beta_4$.

(d) The null and alternative hypotheses are:

$$H_0:\beta_2+\beta_3+\beta_4=0 \qquad H_1:\beta_2+\beta_3+\beta_4\neq 0$$

Using econometric software, we find the *F*-value for this hypothesis to be 3.841, with corresponding *p*-value of 0.0588. Since 0.0588 < 0.10, we do not reject H_0 at a 10% significance level. The data do support the marketing manager's claim.

Exercise 6.24 (continued)

(e) The estimated regression is:

$$\widehat{\ln(SAL1)} = 8.3567 - 1.3177\ln\left(\frac{APR1}{APR2}\right) - 2.7001\ln\left(\frac{APR1}{APR3}\right)$$

$$\text{(se)} \quad (0.0820) \quad (0.5215) \qquad\qquad (0.5534)$$

$a_2 = -1.318$ implies that, holding other variables constant, a 1% increase in the price ratio of brand 1 to brand 2 tuna decreases the sales of brand 1 tuna by 1.318%.

$a_3 = -2.70$ implies that, holding other variables constant, a 1% increase in the price ratio of brand 1 to brand 3 tuna decreases the sales of brand 1 tuna by 2.70%.

The *t*-values for a_2 and a_3 are -2.527 and -4.879, respectively, indicating that both these estimated coefficients are significantly different from zero.

The *F*-test result in part (d) can be confirmed using the sums of squared errors from the restricted and unrestricted models

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N - K)} = \frac{(16.6956 - 15.4585)/1}{15.4585/48} = 3.841$$

(f) Both estimated models in parts (b) and (e) suggest that brand 3 is the stronger competitor to brand 1 because $b_4 > b_3$ and $a_3 < a_2$. A price change in brand 3 has a greater effect on sales of brand 1 than a price change in brand 2.

(g) To confirm that brand 3 is the stronger competitor, we set up an alternative hypothesis that brand 3 is a stronger competitor than brand 2.

For the model in part (a),

$$H_0: \beta_4 \le \beta_3 \text{ against } H_1: \beta_4 > \beta_3$$

The value of the *t*-statistic is

$$t = \frac{b_4 - b_3}{\text{se}(b_4 - b_3)} = \frac{1.6871 - 0.9904}{0.9507} = 0.733$$

The corresponding *p*-value is 0.234. Also, the critical value at a 5% level of significance is $t_{(0.95,48)} = 1.677$. Since $t < 1.677$, we do not reject the null hypothesis. At a 5% level of significance, the evidence is not sufficiently strong to confirm that brand 3 is a stronger competitor than brand 2.

The standard error can be calculated as follows

$$\begin{aligned}\text{se}(b_4 - b_3) &= \sqrt{\widehat{\text{var}}(b_4) + \widehat{\text{var}}(b_3) - 2\times\widehat{\text{cov}}(b_4, b_3)} \\ &= \sqrt{0.556547 + 0.284986 - 2\times(-0.031110)} \\ &= 0.9507\end{aligned}$$

Exercise 6.24(g) (continued)

For the model in part (c),

$$H_0 : \alpha_3 \geq \alpha_2 \quad \text{against} \quad H_1 : \alpha_3 < \alpha_2$$

The value of the *t*-statistic is

$$t = \frac{a_3 - a_2}{\text{se}(a_3 - a_2)} = \frac{-2.7001 - (-1.3177)}{0.9092} = -1.520$$

The corresponding *p*-value is 0.0674. Also, the critical value at a 5% level of significance is $t_{(0.05,49)} = -1.677$. Since $t > -1.677$, we do not reject the null hypothesis. At a 5% level of significance, the evidence is not sufficiently strong to confirm that brand 3 is a stronger competitor than brand 2.

The opposite conclusion is reached if we use a 10% significance level. In this case, $t_{(0.10,49)} = -1.299 > -1.520$, and the evidence is sufficiently strong to confirm that brand 3 is a stronger competitor.

The standard error can be calculated as follows

$$\begin{aligned} \text{se}(a_3 - a_2) &= \sqrt{\widehat{\text{var}}(a_3) + \widehat{\text{var}}(a_2) - 2 \times \widehat{\text{cov}}(a_3, a_2)} \\ &= \sqrt{0.306213 + 0.271995 - 2 \times (-0.124246)} \\ &= 0.9092 \end{aligned}$$

EXERCISE 6.25

(a) To appreciate the relationship between the 3 equations, we begin by rewriting the first equation as follows

$$\begin{aligned} SAL1 &= \beta_1 + \beta_2 APR1 + \beta_3 APR2 + \beta_4 APR3 + e \\ &= \beta_1 + \beta_2\left(\frac{PR1}{100}\right) + \beta_3\left(\frac{PR2}{100}\right) + \beta_4\left(\frac{PR3}{100}\right) + e \\ &= \alpha_1 + \alpha_2 PR1 + \alpha_3 PR2 + \alpha_4 PR3 + e \end{aligned}$$

where $\alpha_1 = \beta_1, \alpha_2 = \beta_2/100, \alpha_3 = \beta_3/100, \alpha_4 = \beta_4/100$. Thus, the coefficients of *PR1*, *PR2*, and *PR3* in the second equation will be 100 times smaller than the coefficients of *APR1*, *APR2*, and *APR3* in the first equation. The intercept coefficient remains unchanged.

For the third equation, we write

$$SAL1 = \alpha_1 + \alpha_2 PR1 + \alpha_3 PR2 + \alpha_4 PR3 + e$$

$$1000 \times SALES = \alpha_1 + \alpha_2 PR1 + \alpha_3 PR2 + \alpha_4 PR3 + e$$

$$\begin{aligned} SALES &= \frac{\alpha_1}{1000} + \frac{\alpha_2}{1000}PR1 + \frac{\alpha_3}{1000}PR2 + \frac{\alpha_4}{1000}PR3 + \frac{e}{1000} \\ &= \gamma_1 + \gamma_2 PR1 + \gamma_3 PR2 + \gamma_4 PR3 + e^* \end{aligned}$$

where $\gamma_1 = \alpha_1/1000, \gamma_2 = \alpha_2/1000, \gamma_3 = \alpha_3/1000, \gamma_4 = \alpha_4/1000$. Thus, all coefficients in the third equation, including the intercept, will be 1000 times smaller than those in the second equation.

The estimated regressions are:

$$\widehat{SAL1} = 22963.43 - 47084.47 APR1 + 9299.00 PR2 + 16511.29 PR3$$

$$\widehat{SAL1} = 22963.43 - 470.8447 PR1 + 92.9900 PR2 + 165.1129 PR3$$

$$\widehat{SALES} = 22.963 - 0.47084 PR1 + 0.09299 PR2 + 0.16511 PR3$$

The relationships between the estimated coefficients in these three equations agree with the conclusions we reached by algebraically manipulating the equations.

Exercise 6.25 (continued)

(b) To obtain the relationship between the coefficients of the first two equations, we write

$$\begin{aligned}\ln(SAL1) &= \beta_1 + \beta_2 APR1 + \beta_3 APR2 + \beta_4 APR3 + e \\ &= \beta_1 + \beta_2\left(\frac{PR1}{100}\right) + \beta_3\left(\frac{PR2}{100}\right) + \beta_4\left(\frac{PR3}{100}\right) + e \\ &= \alpha_1 + \alpha_2 PR1 + \alpha_3 PR2 + \alpha_4 PR3 + e\end{aligned}$$

where $\alpha_1 = \beta_1, \alpha_2 = \beta_2/100, \alpha_3 = \beta_3/100, \alpha_4 = \beta_4/100$. The relationships between the coefficients are the same as those in part (a). The coefficients of *PR1*, *PR2*, and *PR3* in the second equation will be 100 times smaller than the coefficients of *APR1*, *APR2*, and *APR3* in the first equation. The intercept coefficient remains unchanged.

To obtain the third equation from the second, we write

$$\begin{aligned}\ln(SAL1) &= \alpha_1 + \alpha_2 PR1 + \alpha_3 PR2 + \alpha_4 PR3 + e \\ \ln(SALES \times 1000) &= \alpha_1 + \alpha_2 PR1 + \alpha_3 PR2 + \alpha_4 PR3 + e \\ \ln(SALES) &= \alpha_1 - \ln(1000) + \alpha_2 PR1 + \alpha_3 PR2 + \alpha_4 PR3 + e \\ &= \gamma_1 + \gamma_2 PR1 + \gamma_3 PR2 + \gamma_4 PR3 + e\end{aligned}$$

where $\gamma_1 = \alpha_1 - \ln(1000), \gamma_2 = \alpha_2, \gamma_3 = \alpha_3, \gamma_4 = \alpha_4$. The coefficients of the third equation are identical to those of the second equation, with the exception of the intercept which differs by the amount $\ln(1000) = 6.907755$.

The estimated regressions are:

$$\widehat{\ln(SAL1)} = 10.45595 - 6.2176 APR1 + 1.4174 APR2 + 2.1472 APR3$$

$$\widehat{\ln(SAL1)} = 10.45595 - 0.062176 PR1 + 0.014174 PR2 + 0.021472 PR3$$

$$\widehat{\ln(SALES)} = 3.54819 - 0.062176 PR1 + 0.014174 PR2 + 0.021472 PR3$$

These estimates agree with the relationships established algebraically. Note that

$$a_1 - \ln(1000) = 10.45595 - 6.90776 = 3.54819 = \hat{\gamma}_1$$

Exercise 6.25 (continued)

(c) To obtain the relationship between the coefficients of the first two equations, we write

$$\begin{aligned}
\ln(SAL1) &= \beta_1 + \beta_2 \ln(APR1) + \beta_3 \ln(APR2) + \beta_4 \ln(APR3) + e \\
&= \beta_1 + \beta_2 \ln\left(\frac{PR1}{100}\right) + \beta_3 \ln\left(\frac{PR2}{100}\right) + \beta_4 \ln\left(\frac{PR3}{100}\right) + e \\
&= \beta_1 + \beta_2 \ln(PR1) + \beta_3 \ln(PR2) + \beta_4 \ln(PR3) - \left(\beta_2 + \beta_3 + \beta_4\right)\ln(100) + e \\
&= \alpha_1 + \alpha_2 \ln(PR1) + \alpha_3 \ln(PR2) + \alpha_4 \ln(PR3) + e
\end{aligned}$$

where $\alpha_1 = \beta_1 - \left(\beta_2 + \beta_3 + \beta_4\right)\ln(100),\ \alpha_2 = \beta_2,\ \alpha_3 = \beta_3,\ \alpha_4 = \beta_4$. Thus, all coefficients of the second equation are identical to those of the first equation with the exception of the intercept which differs by the amount $\left(\beta_2 + \beta_3 + \beta_4\right)\ln(100)$.

To obtain the third equation from the second, we write

$$\begin{aligned}
\ln(SAL1) &= \alpha_1 + \alpha_2 \ln(PR1) + \alpha_3 \ln(PR2) + \alpha_4 \ln(PR3) + e \\
\ln(SALES \times 1000) &= \alpha_1 + \alpha_2 \ln(PR1) + \alpha_3 \ln(PR2) + \alpha_4 \ln(PR3) + e \\
\ln(SALES) &= \alpha_1 - \ln(1000) + \alpha_2 \ln(PR1) + \alpha_3 \ln(PR2) + \alpha_4 \ln(PR3) + e \\
&= \gamma_1 + \gamma_2 \ln(PR1) + \gamma_3 \ln(PR2) + \gamma_4 \ln(PR3) + e
\end{aligned}$$

where $\gamma_1 = \alpha_1 - \ln(1000), \gamma_2 = \alpha_2, \gamma_3 = \alpha_3, \gamma_4 = \alpha_4$. This result is the same as that obtained in part (b). The coefficients of the third equation are identical to those of the second equation, with the exception of the intercept which differs by the amount $\ln(1000) = 6.907755$.

In all three cases only the intercept changes. This is a general result. Changing the units of measurement of variables in a log-log model does not change the values of the coefficients which are elasticities.

The estimated regressions are:

$$\widehat{\ln(SAL1)} = 7.88938 - 4.6246\ln(APR1) + 0.9904\ln(APR2) + 1.6871\ln(APR3)$$

$$\widehat{\ln(SAL1)} = 16.85591 - 4.6246\ln(PR1) + 0.9904\ln(PR2) + 1.6871\ln(PR3)$$

$$\widehat{\ln(SALES)} = 9.94816 - 4.6246\ln(PR1) + 0.9904\ln(PR2) + 1.6871\ln(PR3)$$

As expected, the elasticity estimates are the same in all three equations. To reconcile the three different intercepts, first note that

$$a_1 - \ln(1000) = 16.855913 - 6.907755 = 9.948158 = \hat{\gamma}_1$$

Comparing equations 1 and 2, we note that

$$\begin{aligned}
&b_1 - \left(b_2 + b_3 + b_4\right)\ln(100) \\
&\quad = 7.889381 - (-4.624576 + 0.990379 + 1.687140) \times 4.60517 \\
&\quad = 16.85591 = a_1
\end{aligned}$$