PRINCIPLES OF ECONOMETRICS

5TH EDITION

ANSWERS TO ODD-NUMBERED EXERCISES IN CHAPTER 6

EXERCISE 6.1

(a) $R^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{2132.65}{4740.11595} = 0.5501$

(b) $F = \dfrac{(SST - SSE)/(K-1)}{SSE/(N-K)} = \dfrac{(4740.11595 - 2132.65)/2}{2132.65/(50-3)} = 28.73$

At $\alpha = 0.01$, the critical value is $F_{(0.99,2,47)} = 5.09$. We reject H_0. There is evidence from the data to suggest that $\beta_2 \neq 0$ and/or $\beta_3 \neq 0$.

(c) The *F*-value for testing this hypothesis is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(2132.65 - 1072.88)/2}{1072.88/(50-5)} = 22.23$$

The critical value for significance level $\alpha = 0.01$ is $F_{(0.99,2,45)} = 5.11$. Since the calculated *F* is greater than the critical *F* we reject H_0 and conclude that the model is misspecified.

(d) $R^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{401.179}{4740.11595} = 0.9154$

(e) $F = \dfrac{\left(SSE_R - SSE_U\right)/J}{SSE_U/(N-K)} = \dfrac{(401.179 - 388.684)/2}{388.684/(50-6)} = 0.707$

The critical value for significance level $\alpha = 0.05$ is $F_{(0.95,2,44)} = 3.209$. Since the calculated *F* is less than the critical *F* we fail to reject H_0. There is no evidence from RESET to suggest the model is misspecified. Including z_i^2 appears to have remedied the initial problem.

EXERCISE 6.3

(a) From the definition of w_i, we have

$$b_2^* = \sum w_i\left(y_i - \bar{y}\right) = \sum w_i y_i$$

with the second equality holding because $\sum w_i \bar{y} = \bar{y}\sum w_i = 0$. Then,

$$\begin{aligned} b_2^* &= \sum w_i y_i \\ &= \sum w_i\left(\beta_1 + \beta_2 x_i + \beta_3 z_i + e_i\right) \\ &= \beta_1 \sum w_i + \beta_2 \sum w_i x_i + \beta_3 \sum w_i z_i + \sum w_i e_i \\ &= \beta_2 \sum w_i\left(x_i - \bar{x}\right) + \beta_3 \sum w_i z_i + \sum w_i e_i \\ &= \beta_2 + \beta_3 \sum w_i z_i + \sum w_i e_i \end{aligned}$$

with the last equality holding because $\sum w_i\left(x_i - \bar{x}\right) = \sum\left(x_i - \bar{x}\right)^2 \Big/ \sum\left(x_i - \bar{x}\right)^2 = 1$.

(b) From part (a),

$$E\left(b_2^* \mid \mathbf{X}\right) = \beta_2 + \beta_3 \sum w_i \left(z_i - \overline{z}\right) + \sum w_i E\left(e_i \mid \mathbf{X}\right)$$

$$= \beta_2 + \beta_3 \frac{\frac{1}{N}\sum\left(x_i - \overline{x}\right)\left(z_i - \overline{z}\right)}{\frac{1}{N}\sum\left(x_i - \overline{x}\right)^2}$$

$$= \beta_2 + \beta_3 \frac{\widehat{\text{cov}}\left(x, z\right)}{\widehat{\text{var}}\left(x\right)}$$

(c) From part (a), the conditional variance of b_2^* is given by

$$\text{var}\left(b_2^* \mid \mathbf{X}\right) = \text{var}\left[\left(\beta_2 + \beta_3 \sum w_i z_i + \sum w_i e_i\right) \mid \mathbf{X}\right]$$

$$= \sum w_i^2 \, \text{var}\left(e_i \mid \mathbf{X}\right)$$

$$= \sigma^2 \frac{\sum\left(x_i - \overline{x}\right)^2}{\left[\sum\left(x_i - \overline{x}\right)^2\right]^2} = \frac{\sigma^2}{\sum\left(x_i - \overline{x}\right)^2} = \frac{\sigma^2}{N\,\widehat{\text{var}}(x)}$$

(d) From equation (5.13),

$$\text{var}\left(b_2 \mid \mathbf{X}\right) = \frac{\sigma^2}{\left(1 - r_{xz}^2\right)\sum\left(x_i - \overline{x}\right)^2}$$

where r_{xz} is the sample correlation between *x* and *z*. Then,

$$\text{var}\left(b_2 \mid \mathbf{X}\right) = \frac{\sigma^2}{\left(1 - r_{xz}^2\right)N\,\widehat{\text{var}}(x)} = \frac{\text{var}\left(b_2^* \mid \mathbf{X}\right)}{\left(1 - r_{xz}^2\right)}$$

Because $\left(1 - r_{xz}^2\right) \le 1$, $\dfrac{\text{var}\left(b_2^* \mid \mathbf{X}\right)}{\left(1 - r_{xz}^2\right)} \ge \text{var}\left(b_2^* \mid \mathbf{X}\right)$

and hence $\text{var}\left(b_2^* \mid \mathbf{X}\right) \le \text{var}\left(b_2 \mid \mathbf{X}\right)$. Although b_2^* suffers from omitted variable bias, it does have a lower variance than b_2.

EXERCISE 6.5

The following assumptions are sufficient

$$E\left[EXPER \mid EDUC, AGE\right] = E\left[EXPER \mid AGE\right] = \delta_1 + \delta_2 AGE + \delta_3 AGE^2$$

$$E\left[EXPER^2 \mid EDUC, AGE\right] = E\left[EXPER^2 \mid AGE\right] = \alpha_1 + \alpha_2 AGE + \alpha_3 AGE^2$$

$$E\left[e \mid EDUC, AGE\right] = 0$$

EXERCISE 6.7

The point and interval predictions for *SALES* from Example 6.15 are $\widehat{SALES}_0 = 76.974$ and (67.533, 86.415), respectively.

The point estimate for $E(SALES \mid PRICE = 6, ADVERT = 1.9)$ is

$$\hat{E}(SALES \mid PRICE = 6, ADVERT = 1.9) = 76.974$$

The standard error is $\text{se}_{E(SALES)} = \sqrt{0.84215} = 0.9177$, leading to the following interval estimate for expected sales

$$76.974 \pm 1.9939 \times 0.9177 = (75.144, 78.804)$$

This interval is much narrower than the one for the prediction interval in Example 6.15. Because it is concerned with estimating average sales from all cities rather than predicting sales for a specific city, the error variance for an observation from a specific city is not included.

EXERCISE 6.9

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 6.11

There are several 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, leading to the model

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

EXERCISE 6.13

(a) Let $\beta_2 =$ the coefficient of *GUNRATE* in the firearm suicide equation. To answer the question is there evidence that the gun buyback has reduced firearm suicides, we test $H_0 : \beta_2 = 0$ against the alternative $H_1 : \beta_2 < 0$. The *t*-value for this test is $t = -0.223/0.069 = -3.232$. The 1% critical value is $t_{(0.01, 195)} = -2.346$. Because $-3.232 < -2.346$, we reject $H_0 : \beta_2 = 0$ at 10%, 5% and 1% significance levels. We conclude that the gun buyback has been successful in reducing the number of firearm suicides.

Let $\alpha_2 =$ the coefficient of *GUNRATE* in the non-firearm suicide equation. To answer the question has there been substitution away from firearms to other means of suicide, we test $H_0:\alpha_2 = 0$ against the alternative $H_1:\alpha_2 > 0$. The *t*-value for this test is $t = 0.553/0.144 = 3.840$. The 1% critical value is $t_{(0.99,195)} = 2.346$. Because $3.840 > 2.346$, we reject $H_0:\alpha_2 = 0$ at 10%, 5% and 1% significance levels. Although the gun buyback has been successful in reducing the number of firearm suicides, it has led to an increase in the number of non-firearm suicides. There has been substitution away from firearms to other means of suicide.

Let $\gamma_5 =$ the coefficient of *YEAR* in the equation for the overall suicide rate. To answer the question is there a separate trend in the suicide rate, we test $H_0:\gamma_5 = 0$ against the alternative $H_1:\gamma_5 \neq 0$. The *t*-value for this test is $t = 0.056/0.393 = 0.142$. The 10% critical values are $\pm t_{(0.95,195)} = \pm 1.653$. Because $-1.653 < 0.142 < 1.653$, we fail to reject $H_0:\gamma_5 = 0$ at 10%, 5% and 1% significance levels. There is no evidence to suggest there is a separate overall trend in the suicide rate.

Tests for the coefficients of *YEAR* in the other two equations suggest there has been a decreasing trend in firearm suicides and an increasing trend in non-firearm suicides, not related to the gun buyback, but there is no evidence of any overall trend in the suicide rate.

(b) Let $\gamma_3 =$ the coefficient of *URATE* in the equation for the overall suicide rate. To answer the question is there evidence that greater unemployment increases the suicide rate, we test $H_0:\gamma_3 = 0$ against the alternative $H_1:\gamma_3 > 0$. The *t*-value for this test is $t = 1.147/1.206 = 0.951$. The 10% critical value is $t_{(0.90,195)} = 1.286$. Because $0.951 < 1.286$, we fail to reject $H_0:\gamma_3 = 0$ at 10%, 5% and 1% significance levels. There is no evidence to suggest that greater unemployment increases the suicide rate.

A similar conclusion is reached from the firearm equation. In the non-firearm equation, testing $H_0:\alpha_3 = 0$ against the alternative $H_1:\alpha_3 > 0$, $t = 1.710$; we reject $H_0:\alpha_3 = 0$ at a 5% significance level, but not at a 1% significance level. There is some evidence that greater unemployment increases the non-firearm suicide rate.

(c) For the firearm equation, we have $H_0:\beta_3 = 0, \beta_4 = 0$ versus $H_1:\beta_3 \neq 0, \text{and/or } \beta_4 \neq 0$. The *F*-value is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(50641 - 29745)/2}{29745/(200-5)} = 68.494$$

For the non-firearm equation, we have $H_0:\alpha_3 = 0, \alpha_4 = 0$ versus $H_1:\alpha_3 \neq 0, \text{and/or } \alpha_4 \neq 0$. The *F*-value is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(131097 - 129122)/2}{129122/(200-5)} = 1.491$$

For the overall equation, we have $H_0:\gamma_3 = 0, \gamma_4 = 0$ versus $H_1:\gamma_3 \neq 0, \text{and/or } \gamma_4 \neq 0$. The *F*-value is

$$F = \frac{\left(SSE_R - SSE_U\right)/J}{SSE_U/(N-K)} = \frac{\left(175562 - 151890\right)/2}{151890/(200-5)} = 15.195$$

The 1%, 5% and 10% critical F-values are, respectively, $F_{(0.99,\,2,\,195)} = 4.716$, $F_{(0.95,\,2,\,195)} = 3.042$, and $F_{(0.90,\,2,\,195)} = 2.330$. All three significance levels lead to the same conclusion. *URATE* and *CITY* contribute to the firearm equation and the overall equation (15.195 and 68.494 are greater than 4.716), but not to the non-firearm equation (1.491 is less than 2.330).

EXERCISE 6.15

(a) We test the null hypothesis $H_0: \beta_3 = 0,\ \beta_4 = 0$ against the alternative that one of the coefficients is not zero. The F-test statistic is

$$F = \frac{\left(SSE_R - SSE_U\right)/J}{SSE_U/\left(N-K\right)} = \frac{\left(1513.6 - 1358.7\right)/2}{1358.7/\left(52-4\right)} = 2.74$$

The 10% critical value is 2.417. We reject the null hypothesis that both coefficients are zero at the 10% level of significance and conclude that at least one of the competing brands prices should be included in the equation.

(b) When *PRB* and *PRC* are omitted from the equation, the coefficient of *PRA* will change by 3.63944. The intercept will change by 17.7507.

(c) Let $m = E(CANS \mid PRA = 0.91, PRB = 0.91, PRC = 0.90)$, then $\hat{m} = 3.4392$

The interval estimate is $\hat{m} \pm t_{(0.975,48)}\mathrm{se}\left(\hat{m}\right) = \left[0.26239708, \quad 6.6160029\right]$

(d) From (b), $\hat{E}(CANS \mid PRA = 0.91) = 1.17979$.

Using (c),

$$\hat{E}(CANS \mid PRA = 0.91) = 22.96 - 47.08 \times 0.91 + 9.30\hat{E}(PRB \mid PRA = 0.91) + 16.51\hat{E}(PRC \mid PRA = 0.91)$$

The estimate from this equation will differ from that in part (c) because in part (c) we set $PRB = 0.91$ and $PRC = 0.90$. These values do not coincide with $\hat{E}(PRB \mid PRA = 0.91)$ and $\hat{E}(PRC \mid PRA = 0.91)$. The values for *PRB* and *PRC* for which the two estimates are the same are $\hat{E}(PRB \mid PRA = 0.91) = 0.84925$ and $\hat{E}(PRC \mid PRA = 0.91) = 0.79737$.

(e) The point prediction is $\widehat{CANS} = 3.4392$. An estimate of the variance of the prediction error f is $\widehat{\mathrm{var}}\left(f\right) = 30.80265$.The prediction interval is

$$\widehat{CANS} \pm t_{(0.975,48)}\mathrm{se}\left(f\right) = 3.4392 \pm 2.0106348\left(5.5500135\right) = \left[-7.7198501, \quad 14.59825\right]$$

We estimate that at the given prices the sales are between zero and 14,598 cans. This interval is very much wider than the interval for the average number of cans sold.

(f) The RESET statistic to test the significance of $\widehat{CANS}^2$ is

$$RESET = \frac{(1358.7 - 1198.9)}{1198.9/47} = 6.2645759$$

For degrees of freedom $(1, 47)$ the 10%, 5% and 1% critical values are 2.8154381, 4.0470999, 7.2068389, respectively. Thus, we reject the null hypothesis of no specification error at 10% and 5%, but not at 1%.

EXERCISE 6.17

Values for the various model selection criteria for each of the equations are reported in the table below. All 4 criteria suggest that all the new models with *BATHS* and *BEDROOMS* included are an improvement over the best model in Example 6.16. From the set of new models, no.11, with the extra variables *BATHS* and $BEDROOMS \times SQFT$, is favored by $\bar{R}^2$, AIC and SC, while the RMSE for the predictions in the hold-out sample favors model no.10, with *BATHS* and *BEDROOMS*.

Model	Variables included in addition to those in (xr6.17)	$\bar{R}^2$	AIC	SC	RMSE
2	None	0.7198	–2.609	–2.587	0.2714
9	*BATHS*	0.7377	–2.674	–2.647	0.2615
10	*BATHS*, *BEDROOMS*	0.7405	–2.684	–2.652	**0.2590**
11	$BATHS, BEDROOMS \times SQFT$	**0.7408**	**–2.685**	**–2.653**	0.2633
12	$BATHS, BEDROOMS \times SQFT, BATHS \times SQFT$	0.7405	–2.682	–2.645	0.2648

EXERCISE 6.19

(a) The change in $E(WFOOD \mid TOTEXP, NK)$ from adding an extra child when $TOTEXP = 90$ is

$$\begin{aligned}&\beta_1 + \beta_2 \ln(90) + \beta_3 (NK+1) + \beta_4 \left[(NK+1) \times \ln(90)\right] \\ &\quad - \left\{\beta_1 + \beta_2 \ln(90) + \beta_3 NK + \beta_4 \left[NK \times \ln(90)\right]\right\} \\ &= \beta_3 + \beta_4 \ln(90)\end{aligned}$$

(b) The change in $E(WFOOD \mid TOTEXP, NK)$ from an increase in total expenditure from £80/week to £120/week when $NK = 2$ is

$$\begin{aligned}&\beta_1 + \beta_2 \ln(120) + 2\beta_3 + \beta_4 \left[2\ln(120)\right] \\ &\quad - \left\{\beta_1 + \beta_2 \ln(80) + 2\beta_3 + \beta_4 \left[2\ln(80)\right]\right\} \\ &= \beta_2 \ln(1.5) + 2\beta_4 \ln(1.5)\end{aligned}$$

(c) Direct substitution yields

$$E(WFOOD \mid TOTEXP = 90, NK = 2) = \beta_1 + \beta_2 \ln(90) + 2\beta_3 + 2\beta_4 \ln(90)$$

(d) For $H_0^{(1)}$, the relationship between the F and χ^2 test values is $F = \chi^2$. The relationship between the critical values is $F_{(1-\alpha,1,846)} > \chi^2_{(1-\alpha,1)}$. The relationship between the t and F tests for $H_0^{(1)}$ is $F = t^2$ for the test values and $F_{(1-\alpha,1,846)} = t^2_{(1-\alpha/2,846)}$ for the critical values.

For $H_0^{(2)}$, the relationship between the F and χ^2 test values is $F = \chi^2/2$. The relationship between the critical values is $F_{(1-\alpha,2,846)} > \chi^2_{(1-\alpha,2)}/2$.
For $H_0^{(3)}$, the relationship between the F and χ^2 test values is $F = \chi^2/3$. The relationship between the critical values is $F_{(1-\alpha,3,846)} > \chi^2_{(1-\alpha,3)}/3$.

(e) The p-values are given in the following table

	p-values for hypotheses					
	$H_0^{(1)}$		$H_0^{(2)}$		$H_0^{(3)}$	
Number of observations	F	χ^2	F	χ^2	F	χ^2
100	0.652	0.651	0.462	0.459	0.386	0.381
400	0.953	0.953	0.053	0.052	0.108	0.106
850	0.094	0.094	0.008	0.008	0.022	0.022

(f) With the exception of going from 100 to 400 observations for $H_0^{(1)}$, increasing the number of observations decreases the level of significance at which the null hypothesis is rejected. This finding is in line with a general result that increasing the number of observations increases the power of a test. Adding more hypotheses does not necessarily increase nor decrease the level of significance at which a null hypothesis is rejected. With 400 observations and a 10% significance level we reject $H_0^{(2)}$, but after adding an extra condition for testing $H_0^{(3)}$ the null hypothesis is no longer rejected at this significance level. There are no dramatic differences between the F-test outcomes and the χ^2-test outcomes.

EXERCISE 6.21

(a) (i) A 95% interval estimate for the saturation level α is given by

$$\hat{\alpha} \pm t_{(0.975,43)}\,\text{se}\left(\hat{\alpha}\right) = 0.81438 \pm 2.0167 \times 0.05105 = (0.7114, 0.9173)$$

(ii) A 95% interval estimate for the inflection point $t_I = -\beta/\delta$ is given by

$$-\hat{\beta}/\hat{\delta} \pm t_{(0.975,43)}\,\text{se}\left(-\hat{\beta}/\hat{\delta}\right) = 1.37767/0.05722 \pm 2.0167 \times 2.4777 = (19.08, 29.07)$$

This interval corresponds to years 1988 to 1998.

(iii) A 95% interval estimate for the EAF share in 1969 when $t = 0$ is given by

$$\frac{\hat{\alpha}}{1+\exp\left(-\hat{\beta}\right)} \pm t_{(0.975,43)}\,\mathrm{se}\left(\frac{\hat{\alpha}}{1+\exp\left(-\hat{\beta}\right)}\right)$$

$$= \frac{0.81438}{1+\exp(1.37767)} \pm 2.0167 \times 0.00671869 = (0.1505, 0.1776)$$

(iv) The predicted shares up to 2050 are given in Table 6.21. They are plotted, along with their 95% bounds, in Figure 6.21. The predictions are approaching the saturation level very slowly. From 2040 to 2050, they increase only a small amount, from 0.7624 to 0.7842, when the saturation level is 0.8144. The 95% bounds become wider the further into the future we are forecasting, reflecting the extra uncertainty from doing so.

Table 6.21 Predicted share of electric arc furnace technology

Year	Share	Year	Share	Year	Share
2016	0.6416	2028	0.7172	2040	0.7624
2017	0.6492	2029	0.7220	2041	0.7651
2018	0.6566	2030	0.7265	2042	0.7677
2019	0.6638	2031	0.7309	2043	0.7701
2020	0.6707	2032	0.7351	2044	0.7725
2021	0.6773	2033	0.7391	2045	0.7747
2022	0.6837	2034	0.7429	2046	0.7768
2023	0.6899	2035	0.7466	2047	0.7788
2024	0.6958	2036	0.7501	2048	0.7807
2025	0.7015	2037	0.7534	2049	0.7825
2026	0.7069	2038	0.7565	2050	0.7842
2027	0.7122	2039	0.7595		

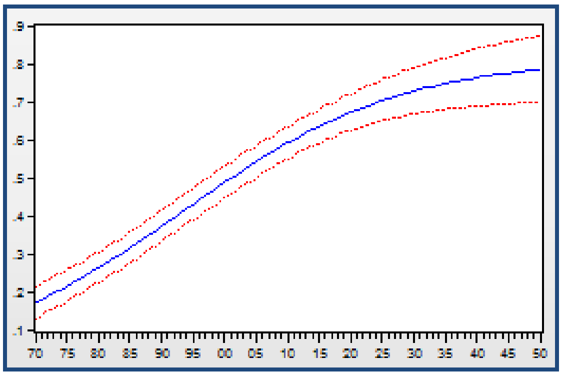

Figure 6.21 Predicted share of electric arc furnace technology and 95% interval bounds

(b) The saturation level is given by α and the point of inflection is given by $t_I = -\beta/\delta$. Having $-\beta/\delta = 25$ is equivalent to $25\delta + \beta = 0$. Thus, we set up the hypotheses

$$H_0 : \alpha = 0.85,\ 25\delta + \beta = 0$$

$$H_1 : \text{at least one of the equalities in } H_0 \text{ does not hold}$$

Values of the $F_{(2,43)}$ and $\chi^2_{(2)}$ test statistics for testing this hypothesis are $F = 4.460$ (*p*-value = 0.0174) and $\chi^2 = 8.92$ (*p*-value = 0.0116). Since both *p*-values are less than 0.05, both test statistics lead to the same conclusion: we reject H_0.

Since $\alpha = 0.85$ and $t_I = 25$ both lie within the respective 95% interval estimates for α and t_I, it is perhaps surprising that the joint hypothesis $H_0 : \alpha = 0.85, t_I = 25$ is rejected. However, the separate interval estimates do not allow for the correlation between the estimates for α and t_I. If these estimates are negatively correlated, then it is unlikely that *both* α and t_I could take on these values which occur in the upper regions of their respective interval estimates.

(c) When the restrictions $\alpha = 0.85$ and $25\delta + \beta = 0$ are imposed, the model becomes

$$y_t = \frac{0.85}{1 + \exp(25\delta - \delta t)} + e_t$$

The *F*-test value is

$$F = \frac{\left(SSE_R - SSE_U\right)/J}{SSE_U/(T-K)} = \frac{\left(0.020462 - 0.017272\right)/2}{0.017272/43} = 3.97$$

which has a corresponding *p*-value of 0.0262. The test outcome is the same. The null hypothesis is rejected. However, the *F*-test value is different. The equivalence of the approaches only holds for equations which are linear in the parameters.

EXERCISE 6.23

(a)

	Coefficient	Std. Error
$\hat{\beta}$	− 2.13711	0.43777
$\hat{\eta}$	0.77897	0.10978
$\hat{\rho}$	− 0.27247	0.74451
$\hat{\delta}$	0.63834	0.09435

(b)

Parameter	Estimate	Standard Error	Interval estimates	
			Lower	Upper
α	0.1180	0.05165	0.0153	0.2207
η	0.7790	0.1098	0.5607	0.9973
ε	1.3745	1.4066	− 1.4226	4.1716
δ	0.6383	0.0944	0.4507	0.8260

(c) For testing $H_0: \eta = 1, \rho = 0$ against $H_1: \eta \neq 1 \text{ or } \rho \neq 0$, we reject H_0 if $F \geq F_{(0.95,2,84)} = 3.105$. Using computer software, the computed value is $F = 2.113$ with *p*-value $\Pr\left(F_{(2,84)} > 2.113\right) = 0.1272$. There is insufficient evidence to suggest a constant-returns-to-scale Cobb-Douglas function is inadequate. Alternatively, carrying out a χ^2-test, we have $\chi^2 = 4.226$, and $\Pr\left(\chi^2_{(2)} > 4.226\right) = 0.1209$

EXERCISE 6.25

(a)

$$\widehat{\ln(PRICE)} = 4.20037 + 0.039603SQFT - 0.018677AGE + 0.00020558AGE^2$$
$$\text{(se)} \quad (0.03009) \quad (0.000974) \quad (0.001436) \quad (0.00002436)$$

(b) Graphing $\hat{E}[\ln(PRICE) \mid SQFT = 22, AGE]$ against *AGE*, we obtain

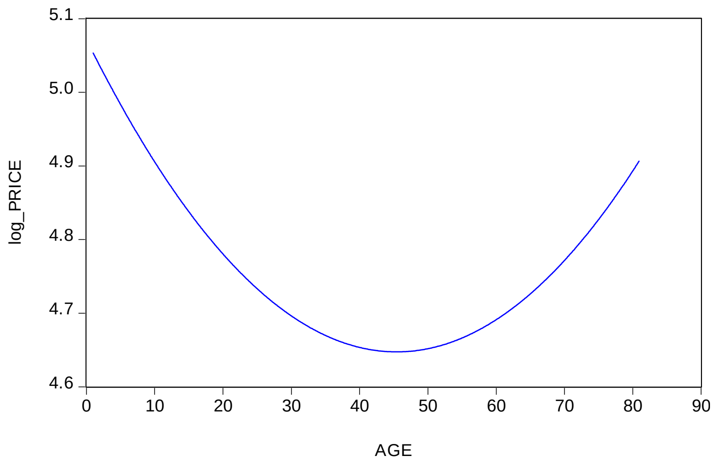

Figure 6.25 Estimated log-price for houses of different ages

(c) Simplifying and combining into the one joint null hypothesis, we have

$$H_0: \begin{cases} 75\beta_3 + 6375\beta_4 = 0 \\ 8\beta_2 + 25\beta_3 + 875\beta_4 = 0 \end{cases}$$

$$H_1: \text{ at least one equality in } H_0 \text{ does not hold}$$

The value of the *F*-statistic for testing H_0 is $F = 1.710$ with *p*-value = 0.1814. Also, the 5% critical value is $F_{(0.95,\,2,\,896)} = 3.006$. Thus, we fail to reject H_0. There is no evidence to suggest the joint hypothesis about expected log prices is incorrect.

(d) Simplifying and combining into one joint null hypothesis, we have.

$$H_0: \begin{cases} \beta_3 + 100\beta_4 = 0 \\ \beta_1 + 22\beta_2 + 50\beta_3 + 2500\beta_4 = 4.60517 \end{cases}$$

$$H_1: \text{ at least one equality in } H_0 \text{ does not hold}$$

The value of the F-statistic for testing H_0 is $F = 3.109$ with p-value = 0.0451. Also, the 5% critical value is $F_{(0.95, 2, 896)} = 3.006$. Thus, we reject H_0. The data do not support the joint hypothesis about expected log prices.

(e) Estimates with *BATHS* and $SQFT \times BEDROOMS$ added are

Variable	Coefficient	Std. Error	*t*-value	*p*-value
C	3.944206	0.042032	93.83751	0.0000
SQFT	0.041407	0.003037	13.63533	0.0000
AGE	-0.016961	0.001404	-12.07745	0.0000
AGE^2	0.000195	2.35E-05	8.286746	0.0000
BATHS	0.175518	0.021777	8.059791	0.0000
*BEDROOMS*SQFT*	-0.002007	0.000585	-3.433094	0.0006

(f) Adding *BATHS* and $SQFT \times BEDROOMS$ will have improved the predictive ability of the model if $\beta_5 \neq 0$ and/or $\beta_6 \neq 0$. Thus, we test $H_0 : \beta_5 = 0, \beta_6 = 0$ against $H_1 : \beta_5 \neq 0$ and/or $\beta_6 \neq 0$. The calculated F-value for this test is $F = 37.358$ with p-value = 0.0000. Also, the 5% critical value is $F_{(0.95, 2, 894)} = 3.006$. Thus, we reject H_0. Adding *BATHS* and $SQFT \times BEDROOMS$ improves the predictive ability of the model.

(g) Using the "natural predictor", before adding one extra bedroom and bathroom the value of the house is estimated to be

$$\widehat{PRICE}_{BEFORE} = \exp\left(b_1 + 23b_2 + 2b_5 + (3\times 23)b_6\right) = 165.525$$

After adding one extra bedroom and bathroom the value of the house is estimated to be

$$\widehat{PRICE}_{AFTER} = \exp\left(b_1 + 25.6b_2 + 3b_5 + (4\times 25.6)b_6\right) = 205.463$$

The estimated increase in value is \$205,463 – \$165,525 = \$39,938.

(h) Assuming there is no inflation in house prices, and the same equation is relevant after 20 years, in 20 years' time the estimated values without and with the extra bedroom and bathroom are

$$\widehat{PRICE}_{BEFORE} = \exp\left(b_1 + 23b_2 + 20b_3 + 20^2 b_4 + 2b_5 + (3\times 23)b_6\right) = 127,466$$

$$\widehat{PRICE}_{AFTER} = \exp\left(b_1 + 25.6b_2 + 20b_3 + 20^2 b_4 + 3b_5 + (4\times 25.6)b_6\right) = 158,221$$

The estimated increase in value is \$158,221 – \$127,466 = \$30,755. The extra bedroom and bathroom has less value in the older house.

EXERCISE 6.27

(a) The least-squares estimated equation using observations where $EDUC > 7$ $(N = 1178)$ is

$$\widehat{\ln(WAGE)} = 1.1391 + 0.11673EDUC + 0.008321EXPER$$
$$\text{(se)} \qquad (0.0869)\ (0.00547) \qquad (0.001059)$$

The hypotheses for the claim about the marginal effects of education and experience are

$$H_0 : \beta_2 = 0.112, \beta_3 = 0.008 \qquad H_1 : \beta_2 \neq 0.112 \text{ or } \beta_3 \neq 0.008$$

We reject H_0 if the computed value for the *F*-statistic, F_{calc}, is greater than or equal to $F_{(0.95,\,2,\,1175)} = 3.003$. We find $F_{calc} = 0.389 < 3.003$, and hence do not reject H_0. There is no evidence to dispute the claim made about the marginal effects of experience and education.

(b) Augmenting the equation with the squares of the predictions, the RESET *F*-value for testing the adequacy of the model is

$$F = \frac{\left(SSE_R - SSE_U\right)/1}{SSE_U/1174} = \frac{261.8566 - 260.4315}{260.4315/1174} = 6.424$$

Since $6.424 > F_{(0.95,\,1,\,1174)} = 3.849$ (*p*-value = 0.0114), we conclude that the squares of the predictions have significant explanatory power and hence the model is inadequate.

Augmenting the equation with the squares and cubes of the predictions, the RESET *F*-value for testing the adequacy of the model is

$$F = \frac{\left(SSE_R - SSE_U\right)/2}{SSE_U/1173} = \frac{\left(261.8566 - 259.8446\right)/2}{259.8446/1173} = 4.541$$

Since $4.541 > F_{(0.95,\,2,\,1173)} = 3.003$ (*p*-value = 0.0108), we conclude that the squares and cubes of the predictions have significant explanatory power and hence the model is inadequate.

(c) The results from estimating the extended model are

Variable	Coefficient	Std. Error	*t*-Value	*p*-value
C	0.065994	0.424611	0.155421	0.8765
EDUC	0.226881	0.053758	4.220395	0.0000
EXPER	0.040795	0.007593	5.372761	0.0000
EDUC^2	-0.003287	0.001724	-1.906540	0.0568
EXPER^2	-0.000481	7.79E-05	-6.181598	0.0000
EDUC*EXPER	-0.000673	0.000419	-1.605578	0.1086

The hypotheses and test results for each part of the question are given in the following table.

Part	Null hypothesis	F_{calc}	$F_{(0.95,\,2,\,1172)}$	*p*-value	Test decision
(i)	$\beta_2 + 20\beta_4 + 5\beta_6 = 0.112$ $\beta_3 + 10\beta_5 + 10\beta_6 = 0.008$	16.039	3.003	0.0000	Reject H_0
(ii)	$\beta_2 + 20\beta_4 + 5\beta_6 = 0.112$ $\beta_3 + 10\beta_5 + 10\beta_6 = 0.008$	0.580	3.003	0.5599	Fail to reject H_0
(iii)	$\beta_2 + 20\beta_4 + 5\beta_6 = 0.112$ $\beta_3 + 10\beta_5 + 10\beta_6 = 0.008$	13.638	3.003	0.0000	Reject H_0

(d) With the squares of the predictions, the RESET *F*-value is

$$F = \frac{(SSE_R - SSE_U)/1}{SSE_U/1171} = \frac{252.7137 - 252.6801}{252.6801/1171} = 0.156$$

Since $0.156 < F_{(0.95,1,1171)} = 3.849$ (*p*-value = 0.6930), we conclude that the squares of the predictions do not have significant explanatory power; the test does not suggest the model is inadequate.

With the squares and cubes of the predictions, the RESET *F*-value is

$$F = \frac{(SSE_R - SSE_U)/2}{SSE_U/1170} = \frac{(252.7137 - 251.8972)/2}{251.8972/1170} = 1.896$$

Since $1.896 < F_{(0.95,\,2,\,1170)} = 3.003$ (*p*-value = 0.1506), we conclude that the squares and cubes of the predictions do not have significant explanatory power; the test does not suggest the model is inadequate.

(e) The claim is not true for all levels of education and experience. The medians for *EDUC* and *EXPER* are 14 and 23, respectively, values that approximately correspond to those tested in part (ii). Thus, the claim is reasonable for those at the median levels of education and experience, but it does not hold for values of *EDUC* and *EXPER* which are much greater or less than the medians.

EXERCISE 6.29

(a) The estimated equation is

$$\widehat{\ln(SHARE)} = 2.9569 - 0.015074\,TAX$$
$$\text{(se)} \qquad (0.0405) \quad (0.001038)$$

We find that an increase in the marginal tax rate for top earners by one percentage point is associated with a decrease in the income share of the top earners by 1.5 percent. It is unlikely we could interpret this estimate as causal. *SHARE* will depend not only on *TAX*, but also on variables related to the general state of the economy. If these variables are correlated with *TAX*, which is likely, β_2 cannot be given a causal interpretation.

(b) The estimated model with the quadratic trend included is

$$\widehat{\ln(SHARE)} = 3.1325 - 0.0079849\,TAX - 0.027217\,YEAR + 0.00031104\,YEAR^2$$
$$\text{(se)} \qquad (0.0411) \quad (0.0015703) \qquad (0.004054) \qquad (0.00004250)$$

After taking out the effect of the quadratic trend, we find that an increase in the marginal tax rate for top earners by one percentage point is associated with a decrease in the income share of the top earners by 0.8 percent. Adding the trend has decreased the estimated effect of the marginal tax rate by a considerable amount.

The correlations between *TAX* and *YEAR* and *TAX* and $YEAR^2$ are given in the following matrix.

	TAX	*YEAR*	*YEAR_SQ*
TAX	1.000000	0.629566	0.462882
YEAR	0.629566	1.000000	0.969008
YEAR_SQ	0.462882	0.969008	1.000000

The correlations between *TAX* and *YEAR* and *TAX* and $YEAR^2$ are 0.630 and 0.463, respectively. These values are moderately high and sufficient for omission of the trend from the equation to result in overestimation of the effect of changes in *TAX* on ln(*SHARE*). A regression of *TAX* on *YEAR* and $YEAR^2$ gives $R^2 = 0.751$, a large enough value to explain the change in magnitude for $\hat{\alpha}_2$.

(c) *SHARE* will be smallest when ln(*SHARE*) is smallest. To find that year, we solve the following equation for *YEAR*

$$\frac{\partial \ln(SHARE)}{\partial YEAR} = \alpha_3 + 2\alpha_4 YEAR = 0 \qquad \text{giving} \qquad YEAR_{MIN} = -\frac{\alpha_3}{2\alpha_4}$$

with point estimate

$$\widehat{YEAR}_{MIN} = -\frac{\hat{\alpha}_3}{2\hat{\alpha}_4} = -\frac{-0.0272166}{2 \times 0.000311042} = 43.751$$

Its standard error is given by $\text{se}\left(\widehat{YEAR}_{MIN}\right) = 1.243246$

A 95% interval estimate for $YEAR_{MIN}$ is

$$\widehat{YEAR}_{MIN} \pm t_{(0.975,\,76)}\,\text{se}\left(\widehat{YEAR}_{MIN}\right) = (41.27,\ 46.23)$$

After rounding to the closest year, the value $YEAR = 43.75$ corresponds to the year 1964 while $YEAR = 41.27$ and $YEAR = 46.23$ correspond to the years 1961 and 1966, respectively. Thus, we estimate that *SHARE* will be smallest in 1964, but, based on a 95% interval estimate, the year when it is smallest could be as early as 1961 or as late as 1966. The smallest actual value for *SHARE* is 7.74 in 1973, a year which does not fall within the interval estimate.

(d) The hypotheses are

$$H_0: \alpha_1 + 50\alpha_2 + 80\alpha_3 + 80^2\alpha_4 = \ln(12)$$
$$H_1: \alpha_1 + 50\alpha_2 + 80\alpha_3 + 80^2\alpha_4 \neq \ln(12)$$

The calculated t-value for this hypothesis is

$$t = \frac{\hat{\alpha}_1 + 50\hat{\alpha}_2 + 80\hat{\alpha}_3 + 80^2\hat{\alpha}_4 - \ln(12)}{\text{se}\left(\hat{\alpha}_1 + 50\hat{\alpha}_2 + 80\hat{\alpha}_3 + 80^2\hat{\alpha}_4\right)} = \frac{0.061677}{0.047484} = 1.299$$

with p-value $\Pr\left(\left|t_{(76)}\right| > 1.299\right) = 0.1979$. Thus, we fail to reject H_0 at all conventional significance levels. The data do not contradict the conjecture about the expected log income share of top earners in the year 2000 for a marginal tax rate of 50%.

(e) The hypotheses are

$$H_0: \begin{Bmatrix} \alpha_1 + 50\alpha_2 + 80\alpha_3 + 80^2\alpha_4 = \ln(12) \\ \alpha_1 + 50\alpha_2 + 5\alpha_3 + 5^2\alpha_4 = \ln(12) \end{Bmatrix}$$

$$H_1: \text{at least one equality in } H_0 \text{ does not hold}$$

The calculated F-value for this hypothesis is 1.270 with p-value $\Pr\left(F_{(2,76)} > 1.270\right) = 0.2867$. Thus, we fail to reject H_0 at the 10%, 5% and 1% levels of significance. We conclude that the data are consistent with the claim that a marginal tax rate of 50% leads to the same expected log share of ln(12) for the top income earners in both 1925 and 2000.

(f) Adding *GWTH* to the equation, we obtain

$$\widehat{\ln(SHARE)} = 3.1264 - 0.0081225TAX - 0.027073YEAR + 0.00030983YEAR^2 + 0.0034719GWTH$$
$$\text{(se)} \quad (0.0409) \; (0.0015585) \quad (0.004018) \quad (0.00004212) \quad (0.0022362)$$

Comparing the estimates of the coefficients in this equation with the corresponding ones in part (b), we find that adding *GWTH* has led to very little change in the estimates. They are almost identical. Also, the same test outcomes are obtained when the tests in parts (d) and (e) are repeated using the equation with *GWTH* included. For the test in part (d), we obtain $t = 1.104$ and $\Pr\left(\left|t_{(75)}\right| > 1.104\right) = 0.2730$; for the test in part (e), $F = 0.999$ and $\Pr\left(F_{(2,75)} > 0.999\right) = 0.3731$.

The lack of changes can be explained by the low correlations between *GWTH* and the other right-hand side variables in the equation. These correlations are given in the following matrix. They are all less than 0.1.

	TAX	*YEAR*	*YEAR_SQ*
GWTH	0.063947	0.019622	0.009369

(g) To obtain an expected log income share of ln(12) in 2001, when $GWTH_{2001} = 3$, we require

$$TAX = \frac{\ln(12) - \alpha_1 - 81\alpha_3 - 81^2\alpha_4 - 3\alpha_5}{\alpha_2}$$

A point estimate for this quantity is $\widehat{TAX} = 60.54$.

Using the delta rule to get its standard error, we obtain $\text{se}\left(\widehat{TAX}\right) = 7.514183$.

A 95% interval estimate is given by

$$\widehat{TAX} \pm t_{(0.975,75)}\,\text{se}\left(\widehat{TAX}\right) = (45.57, 75.51)$$

We estimate that a marginal tax rate of 60.5% would be required. However, the interval estimate that ranges from 45.6% to 75.5% shows there is a great deal of uncertainty about this estimate.

EXERCISE 6.31

(a) Estimates for the model $CSUMPTN = \beta_1 + \beta_2 HOURS + \beta_3 GOV + \beta_4 R + \beta_5 INC + e$ are given in the following output.

Dependent Variable: CSUMPTN
Method: Least Squares
Sample: 1971 2007
Included observations: 37

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003980	0.003221	1.235388	0.2257
HOURS	0.424608	0.184279	2.304153	0.0279
GOV	0.164079	0.141829	1.156879	0.2559
R	0.243435	0.074703	3.258690	0.0027
INC	0.608275	0.138554	4.390149	0.0001

The coefficient of β_3, with *p*-value 0.2559 and the intercept with a *p*-value of 0.2257 are not significantly different from zero at a 5% significance level.

(b) The positive estimates for β_2 and β_3 suggest that both hours worked and government consumption are complements for private consumption. Testing $H_0: \beta_2 = 0, \beta_3 = 0$ against the alternative that at least one of β_2 or β_3 is zero yields a test value of $F = 2.662$. Comparing this value with the critical value $F_{(0.95,\,2,\,32)} = 3.295$, or examining its *p*-value of 0.0852, we find we cannot reject $H_0: \beta_2 = 0, \beta_3 = 0$ at a 5% significance level. Thus, it is possible that both hours worked and government consumption are neither complements nor substitutes for private consumption.

(c) Re-estimating the equation $CSUMPTN = \beta_1 + \beta_2 HOURS + \beta_3 GOV + \beta_4 R + \beta_5 INC + e$ with *GOV* omitted yields the following results:

Dependent Variable: CSUMPTN
Method: Least Squares
Sample: 1971 2007
Included observations: 37

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006141	0.002638	2.327820	0.0262
HOURS	0.308006	0.155062	1.986338	0.0554
R	0.259905	0.073709	3.526097	0.0013
INC	0.746998	0.069770	10.70651	0.0000

Omitting the variable *GOV* has led to a decrease in the magnitude of the coefficient estimate for *HOURS*, making it no longer significantly different from zero at a 5% level of significance, and an increase in the magnitude of the coefficient estimate for *INC*. The coefficient estimate for *R* has changed very little.

(d) Estimates for the equation $GOV = \alpha_1 + \alpha_2 HOURS + \alpha_3 R + \alpha_4 INC + v$ are

Dependent Variable: GOV
Method: Least Squares
Sample: 1971 2007
Included observations: 37

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.013171	0.003221	4.088575	0.0003
HOURS	-0.710646	0.189352	-3.753038	0.0007
R	0.100382	0.090009	1.115249	0.2728
INC	0.845469	0.085199	9.923439	0.0000

The *p*-values in this output suggest that *GOV* is highly correlated with *HOURS* and *INC*, but less so with *R*. These results explain why omission of *GOV* from the equation in part (c) has led to relatively large changes in the coefficient estimates for *HOURS* and *INC*, and only a small change in the coefficient estimate for *R*. Also, the signs of the coefficients of *HOURS* and *INC* in part (d) explain the direction of the changes. In part (c) the effect on *CSUMPTN* of a change in *HOURS* includes the negative indirect effect of *HOURS* on *GOV*. Similarly, the effect on *CSUMPTN* of a change in *INC* includes the positive indirect effect of *INC* on *GOV*. The changes can be reconciled explicitly as follows:

$$E\left(CSUMPTN \mid HOURS, R, INC\right) = \beta_1 + \beta_3\alpha_1 + \left(\beta_2 + \beta_3\alpha_2\right) HOURS + \left(\beta_4 + \beta_3\alpha_3\right) R + \left(\beta_5 + \beta_3\alpha_4\right) INC$$

$$= \gamma_1 + \gamma_2 HOURS + \gamma_3 R + \gamma_4 INC$$

Estimates for $\left(\gamma_1, \gamma_2, \gamma_3, \gamma_4\right)$ were given in part (c). We can derive those estimates from those given in parts (a) and (d):

$$\hat{\gamma}_1 = b_1 + b_3 a_1 = 0.003980 + 0.164079 \times 0.013171 = 0.006141$$

$$\hat{\gamma}_2 = b_2 + b_3 a_2 = 0.424608 - 0.164079 \times 0.710646 = 0.308006$$

$$\hat{\gamma}_3 = b_4 + b_3 a_3 = 0.243435 + 0.164079 \times 0.100382 = 0.259906$$

$$\hat{\gamma}_4 = b_5 + b_3 a_4 = 0.608275 + 0.164079 \times 0.845469 = 0.746999$$

(e) Re-estimating the models in parts (a) and (c) with the year 2007 omitted yields:

Dependent Variable: CSUMPTN
Method: Least Squares
Sample: 1971 2006
Included observations: 36

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004210	0.003299	1.276270	0.2113
HOURS	0.446854	0.192558	2.320614	0.0271
GOV	0.174939	0.145465	1.202621	0.2382
R	0.237276	0.076774	3.090567	0.0042
INC	0.593240	0.143926	4.121834	0.0003

Dependent Variable: CSUMPTN
Method: Least Squares
Sample: 1971 2006
Included observations: 36

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006360	0.002791	2.278774	0.0295
HOURS	0.316552	0.160290	1.974873	0.0570
R	0.256924	0.075537	3.401289	0.0018
INC	0.743555	0.071856	10.34786	0.0000

Point and 95% interval forecasts for *CSUMPTN* in 2007 are:

From model (a): $\widehat{CSUMPTN}_{2007} = 0.008779$ and the 95% interval forecast is

$$\widehat{CSUMPTN}_{2007} \pm t_{(0.975,\,31)}\text{se}(f) = (-0.01381, 0.03137)$$

From model (c): $\widehat{CSUMPTN}_{2007} = 0.006649$ and the 95% interval forecast is

$$\widehat{CSUMPTN}_{2007} \pm t_{(0.975,\,32)}\text{se}(f) = (-0.01578, 0.02908)$$

(f) The 2007 realized value for consumption growth was $CSUMPTN_{2007} = 0.003606$. The model in part (c) produced the more accurate forecast for 2007.

EXERCISE 6.33 (a)

Dependent Variable: LOG(PRICE)
Method: Least Squares
Sample: 1 1022 IF CABERNET=1
Included observations: 437

cabernet wines

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.846092	0.179905	10.26147	0.0000
SCORE	0.021656	0.001952	11.09663	0.0000
AGE	0.067283	0.008853	7.600391	0.0000
CASES	0.003614	0.001052	3.435795	0.0006

R-squared	0.304617	Mean dependent var	4.018609
Adjusted R-squared	0.299799	S.D. dependent var	0.139785
S.E. of regression	0.116969	Akaike info criterion	-1.444705
Sum squared resid	5.924196	Schwarz criterion	-1.407360

Dependent Variable: LOG(PRICE)
Method: Least Squares
Sample: 1 1022 IF PINOT=1
Included observations: 290

pinot wines

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.920380	0.204410	9.394730	0.0000
SCORE	0.020399	0.002333	8.743736	0.0000
AGE	0.077050	0.009901	7.781790	0.0000
CASES	0.002567	0.011279	0.227570	0.8201

R-squared	0.339668	Mean dependent var	3.891548
Adjusted R-squared	0.332741	S.D. dependent var	0.123993
S.E. of regression	0.101285	Akaike info criterion	-1.728056
Sum squared resid	2.933984	Schwarz criterion	-1.677437

Dependent Variable: LOG(PRICE)
Method: Least Squares
Sample: 1 1022 IF OTHER=1
Included observations: 295

other varieties

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.657799	0.222400	7.454115	0.0000
SCORE	0.023179	0.002520	9.197971	0.0000
AGE	0.078215	0.009092	8.602931	0.0000
CASES	0.004309	0.004144	1.039863	0.2993

R-squared	0.391415	Mean dependent var	3.953677
Adjusted R-squared	0.385141	S.D. dependent var	0.149644
S.E. of regression	0.117340	Akaike info criterion	-1.434012
Sum squared resid	4.006701	Schwarz criterion	-1.384019

Casual inspection suggests separate equations are not needed for the three different varieties. Coefficient estimates for corresponding variables are all similar in magnitude, particularly when considered relative to their standard errors.

(b) $SSE_U = 5.924196 + 2.933984 + 4.006701 = 12.864881$

(c) The total number of coefficients in the 3 equations is 12. If a single equation is estimated for all 3 varieties, the number of parameters is 4. Thus, if we restrict corresponding coefficients for all varieties to be equal, there are $12 - 4 = 8$ restrictions.

(d) Using all the data to estimate one equation gives the following results:

Dependent Variable: LOG(PRICE)
Method: Least Squares
Sample: 1 1022
Included observations: 1022

all varieties

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.768808	0.109036	16.22220	0.0000
SCORE	0.022170	0.001241	17.86691	0.0000
AGE	0.074700	0.004643	16.08986	0.0000
CASES	0.003832	0.000965	3.970711	0.0001

R-squared	0.422624	Mean dependent var	3.963812
Adjusted R-squared	0.420923	S.D. dependent var	0.148129
S.E. of regression	0.112722	Akaike info criterion	-1.523878
Sum squared resid	12.93496	Schwarz criterion	-1.504584

(e) Let the coefficients of the cabernet, pinot and other varieties be denoted by $(\beta_{C1},\beta_{C2},\beta_{C3},\beta_{C4})$, $(\beta_{P1},\beta_{P2},\beta_{P3},\beta_{P4})$, and $(\beta_{O1},\beta_{O2},\beta_{O3},\beta_{O4})$, respectively. The null and alternative hypotheses are

$$H_0: \beta_{C1} = \beta_{P1} = \beta_{O1},\ \beta_{C2} = \beta_{P2} = \beta_{O2},\ \beta_{C3} = \beta_{P3} = \beta_{O3},\ \beta_{C4} = \beta_{P4} = \beta_{O4}$$

H_1: At least one of the inequalities in H_0 does not hold.

The value of the test statistic is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(12.93496 - 12.864881)/8}{12.864881/1010} = 0.688$$

This value is less than the 5% critical value $F_{(0.95,\,8,\,1010)} = 1.948$. Thus, the null hypothesis is not rejected. There is no evidence to suggest that there should be different equations for different varieties.