Answers to Selected Exercises

For

Principles of Econometrics, Fourth Edition

R. CARTER HILL
Louisiana State University

WILLIAM E. GRIFFITHS
University of Melbourne

GUAY C. LIM
University of Melbourne

JOHN WILEY & SONS, INC
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

CONTENTS

Answers for Selected Exercises in:

Probability Primer		1
Chapter 2	The Simple Linear Regression Model	3
Chapter 3	Interval Estimation and Hypothesis Testing	12
Chapter 4	Prediction, Goodness of Fit and Modeling Issues	16
Chapter 5	The Multiple Regression Model	22
Chapter 6	Further Inference in the Multiple Regression Model	29
Chapter 7	Using Indicator Variables	36
Chapter 8	Heteroskedasticity	44
Chapter 9	Regression with Time Series Data: Stationary Variables	51
Chapter 10	Random Regressors and Moment Based Estimation	58
Chapter 11	Simultaneous Equations Models	60
Chapter 15	Panel Data Models	64
Chapter 16	Qualitative and Limited Dependent Variable Models	66
Appendix A	Mathematical Tools	69
Appendix B	Probability Concepts	72
Appendix C	Review of Statistical Inference	76

29 August, 2011

PROBABILITY PRIMER

Exercise Answers

EXERCISE P.1

(a) X is a random variable because attendance is not known prior to the outdoor concert.

(b) 1100

(c) 3500

(d) $6,000,000$

EXERCISE P.3

0.0478

EXERCISE P.5

(a) 0.5.

(b) 0.25

EXERCISE P.7

(a)

$f(c)$
0.15
0.40
0.45

(b) 1.3

(c) 0.51

(d) $f(0,0) = 0.05 \neq f_C(0)\,f_B(0) = 0.15 \times 0.15 = 0.0225$

(e)

A	$f(a)$
5000	0.15
6000	0.50
7000	0.35

(f) 1.0

EXERCISE P.11

(a) 0.0289

(b) 0.3176

(c) 0.8658

(d) 0.444

(e) 1.319

EXERCISE P.13

(a) 0.1056

(b) 0.0062

(c) (a) 0.1587 (b) 0.1265

EXERCISE P.15

(a) 9

(b) 1.5

(c) 0

(d) 109

(e) −66

(f) −0.6055

EXERCISE P.17

(a) $4a + b(x_1 + x_2 + x_3 + x_4)$

(b) 14

(c) 34

(d) $f(4) + f(5) + f(6)$

(e) $f(0, y) + f(1, y) + f(2, y)$

(f) 36

CHAPTER 2

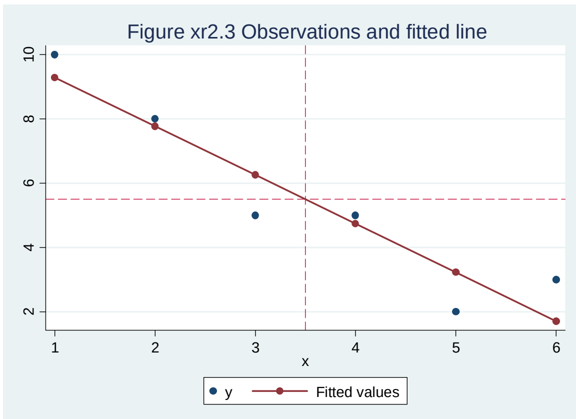
Exercise Answers

EXERCISE 2.3

(a) The line drawn for part (a) will depend on each student's subjective choice about the position of the line. For this reason, it has been omitted.

(b) $b_2 = -1.514286$

$b_1 = 10.8$



(c) $\bar{y} = 5.5$

$\bar{x} = 3.5$

$\hat{y} = 5.5$

Exercise 2.3 (Continued)

(d)

$\hat{e}_i$
0.714286
0.228571
−1.257143
0.257143
−1.228571
1.285714

$\sum \hat{e}_i = 0.$

(e) $\sum x_i \hat{e}_i = 0$

EXERCISE 2.6

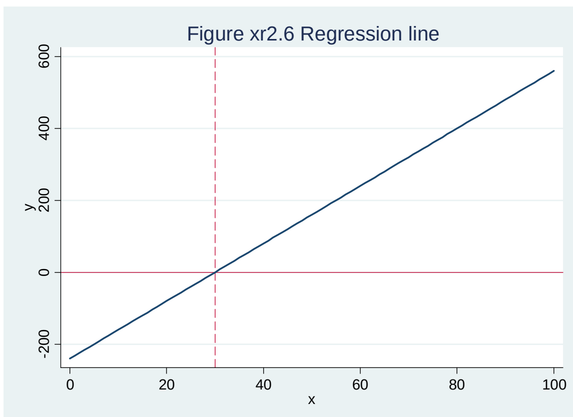
(a) The intercept estimate $b_1 = -240$ is an estimate of the number of sodas sold when the temperature is 0 degrees Fahrenheit. Clearly, it is impossible to sell –240 sodas and so this estimate should not be accepted as a sensible one.

The slope estimate $b_2 = 8$ is an estimate of the increase in sodas sold when temperature increases by 1 Fahrenheit degree. One would expect the number of sodas sold to increase as temperature increases.

(b) $\hat{y} = -240 + 8 \times 80 = 400$

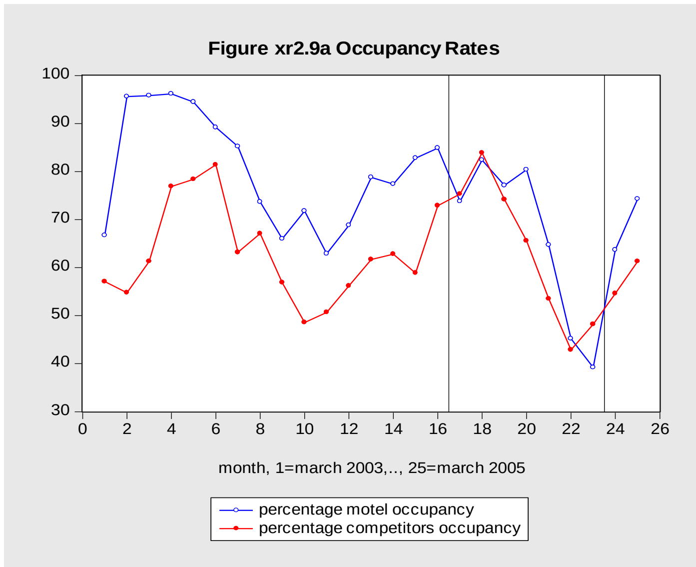
(c) She predicts no sodas will be sold below 30°F.

(d) A graph of the estimated regression line:



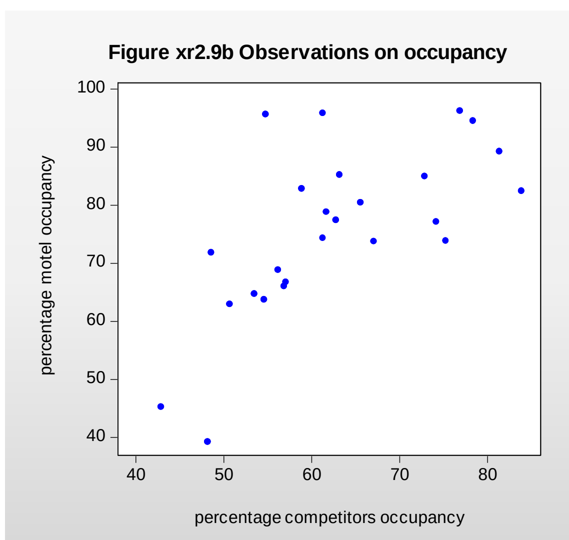
EXERCISE 2.9

(a)



The repair period comprises those months between the two vertical lines. The graphical evidence suggests that the damaged motel had the higher occupancy rate before and after the repair period. During the repair period, the damaged motel and the competitors had similar occupancy rates.

(b) A plot of *MOTEL_PCT* against *COMP_PCT* yields:



There appears to be a positive relationship the two variables. Such a relationship may exist as both the damaged motel and the competitor(s) face the same demand for motel rooms.

Exercise 2.9 (continued)

(c) $\widehat{MOTEL\_PCT} = 21.40 + 0.8646 \times COMP\_PCT$.

The competitors' occupancy rates are positively related to motel occupancy rates, as expected. The regression indicates that for a one percentage point increase in competitor occupancy rate, the damaged motel's occupancy rate is expected to increase by 0.8646 percentage points.

(d)

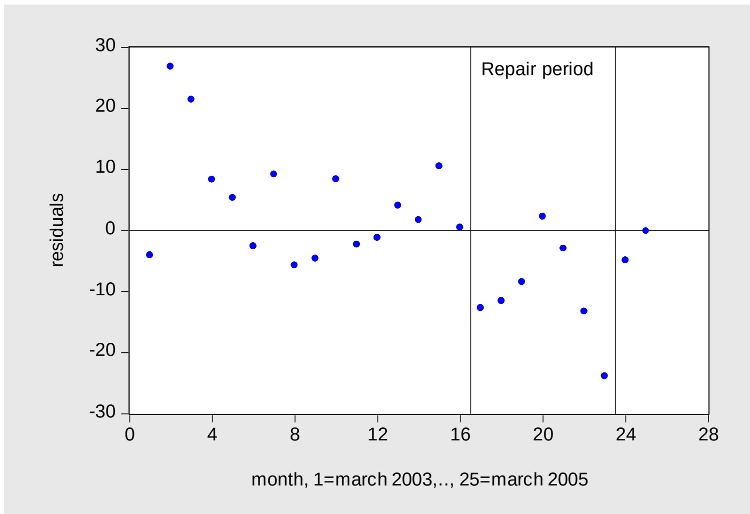


Figure xr2.9(d) Plot of residuals against time

The residuals during the occupancy period are those between the two vertical lines. All except one are negative, indicating that the model has over-predicted the motel's occupancy rate during the repair period.

(e) We would expect the slope coefficient of a linear regression of *MOTEL_PCT* on *RELPRICE* to be negative, as the higher the relative price of the damaged motel's rooms, the lower the demand will be for those rooms, holding other factors constant.

$$\widehat{MOTEL\_PCT} = 166.66 - 122.12 \times RELPRICE$$

(f) The estimated regression is:

$$\widehat{MOTEL\_PCT} = 79.3500 - 13.2357 \times REPAIR$$

In the non-repair period, the damaged motel had an estimated occupancy rate of 79.35%. During the repair period, the estimated occupancy rate was 79.35−13.24 = 66.11%. Thus, it appears the motel did suffer a loss of occupancy and profits during the repair period.

(g) From the earlier regression, we have

$$\overline{MOTEL_0} = b_1 = 79.35\%$$

$$\overline{MOTEL_1} = b_1 + b_2 = 79.35 - 13.24 = 66.11\%$$

Exercise 2.9(g) (continued)

For competitors, the estimated regression is:

$$\widehat{COMP\_PCT} = 62.4889 + 0.8825 \times REPAIR$$

$$\overline{COMP}_0 = b_1 = 62.49\%$$

$$\overline{COMP}_1 = b_1 + b_2 = 62.49 + 0.88 = 63.37\%$$

During the non-repair period, the difference between the average occupancies was:

$$\overline{MOTEL}_0 - \overline{COMP}_0 = 79.35 - 62.49 = 16.86\%$$

During the repair period it was

$$\overline{MOTEL}_1 - \overline{COMP}_1 = 66.11 - 63.37 = 2.74\%$$

This comparison supports the motel's claim for lost profits during the repair period. When there were no repairs, their occupancy rate was 16.86% higher than that of their competitors; during the repairs it was only 2.74% higher.

(h) $$\widehat{MOTEL\_PCT - COMP\_PCT} = 16.8611 - 14.1183 \times REPAIR$$

The intercept estimate in this equation (16.86) is equal to the difference in average occupancies during the non-repair period, $\overline{MOTEL}_0 - \overline{COMP}_0$. The sum of the two coefficient estimates $\left(16.86 + (-14.12) = 2.74\right)$ is equal to the difference in average occupancies during the repair period, $\overline{MOTEL}_1 - \overline{COMP}_1$.

This relationship exists because averaging the difference between two series is the same as taking the difference between the averages of the two series.

EXERCISE 2.12

(a) and (b) $$\widehat{SPRICE} = -30069 + 9181.7\,LIVAREA$$

The coefficient 9181.7 suggests that selling price increases by approximately \$9182 for each additional 100 square foot in living area. The intercept, if taken literally, suggests a house with zero square feet would cost −\$30,069, a meaningless value.

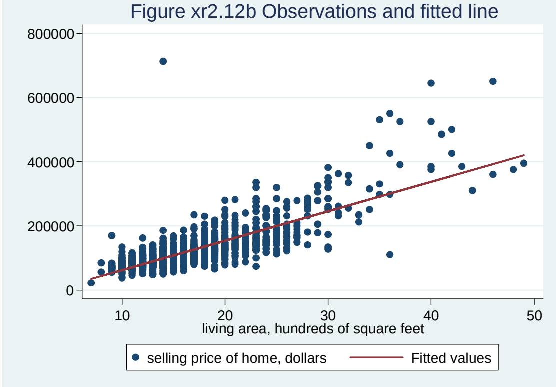


Figure xr2.12b Observations and fitted line

Exercise 2.12 (continued)

(c) The estimated quadratic equation for all houses in the sample is

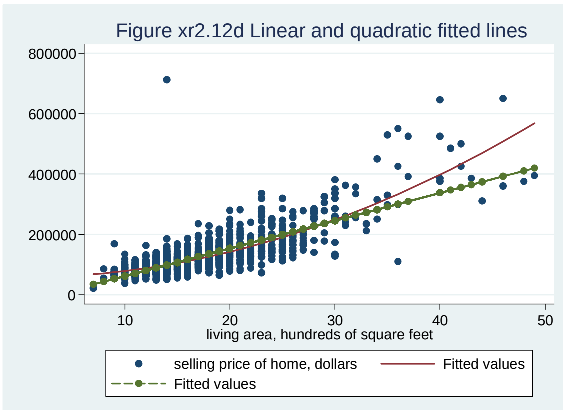
$$\widehat{SPRICE} = 57728 + 212.611 LIVAREA^2$$

The marginal effect of an additional 100 square feet for a home with 1500 square feet of living space is:

$$\widehat{\text{slope}} = \frac{d\left(\widehat{SPRICE}\right)}{dLIVAREA} = 2(212.611)LIVAREA = 2(212.611)(15) = 6378.33$$

That is, adding 100 square feet of living space to a house of 1500 square feet is estimated to increase its expected price by approximately $6378.

(d)



The quadratic model appears to fit the data better; it is better at capturing the proportionally higher prices for large houses.

SSE of linear model, (b): $SSE = \sum \hat{e}_i^2 = 2.23 \times 10^{12}$

SSE of quadratic model, (c): $SSE = \sum \hat{e}_i^2 = 2.03 \times 10^{12}$

The *SSE* of the quadratic model is smaller, indicating that it is a better fit.

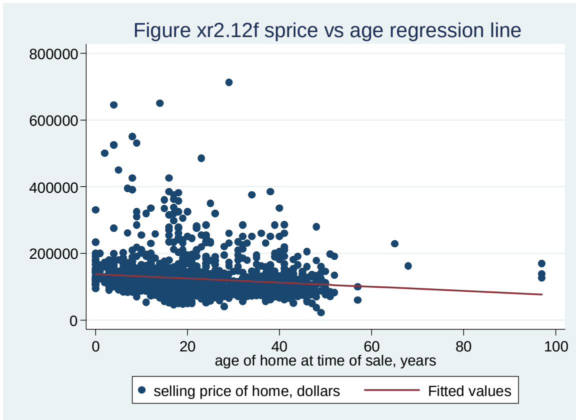
(e) Large lots: $\widehat{SPRICE} = 113279 + 193.83 LIVAREA^2$

Small lots: $\widehat{SPRICE} = 62172 + 186.86 LIVAREA^2$

The intercept can be interpreted as the expected price of the land – the selling price for a house with no living area. The coefficient of *LIVAREA* has to be interpreted in the context of the marginal effect of an extra 100 square feet of living area, which is $2\beta_2 LIVAREA$. Thus, we estimate that the mean price of large lots is $113,279 and the mean price of small lots is $62,172. The marginal effect of living area on price is $\$387.66 \times LIVAREA$ for houses on large lots and $\$373.72 \times LIVAREA$ for houses on small lots.

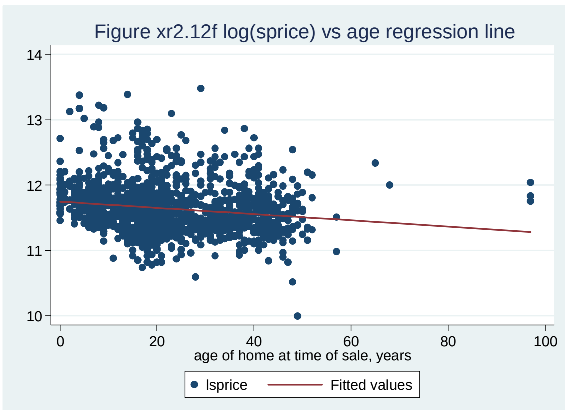
Exercise 2.12 (continued)

(f) The following figure contains the scatter diagram of *PRICE* and *AGE* as well as the estimated equation $\widehat{SPRICE} = 137404 - 627.16AGE$. We estimate that the expected selling price is \$627 less for each additional year of age. The estimated intercept, if taken literally, suggests a house with zero age (i.e., a new house) would cost \$137,404.



The following figure contains the scatter diagram of ln(*PRICE*) and *AGE* as well as the estimated equation $\widehat{\ln(SPRICE)} = 11.746 - 0.00476AGE$. In this estimated model, each extra year of age reduces the selling price by 0.48%. To find an interpretation from the intercept, we set $AGE = 0$, and find an estimate of the price of a new home as

$$\exp\left[\widehat{\ln(SPRICE)}\right] = \exp(11.74597) = \$126,244$$

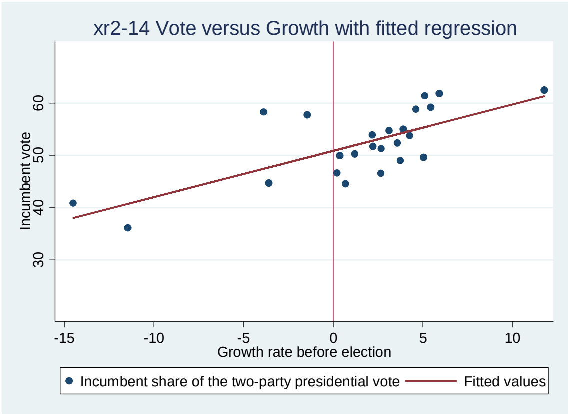


Based on the plots and visual fit of the estimated regression lines, the log-linear model shows much less of problem with under-prediction and so it is preferred.

(g) The estimated equation for all houses is $\widehat{SPRICE} = 115220 + 133797LGELOT$. The estimated expected selling price for a house on a large lot (*LGELOT* = 1) is 115220+133797 = \$249017. The estimated expected selling price for a house not on a large lot (*LGELOT* = 0) is \$115220.

EXERCISE 2.14

(a) and (b)



There appears to be a positive association between *VOTE* and *GROWTH*.

The estimated equation for 1916 to 2008 is

$$\widehat{VOTE} = 50.848 + 0.88595GROWTH$$

The coefficient 0.88595 suggests that for a 1 percentage point increase in the growth rate of *GDP* in the 3 quarters before the election there is an estimated increase in the share of votes of the incumbent party of 0.88595 percentage points.

We estimate, based on the fitted regression intercept, that that the incumbent party's expected vote is 50.848% when the growth rate in *GDP* is zero. This suggests that when there is no real *GDP* growth, the incumbent party will still maintain the majority vote.

(c) The estimated equation for 1916 - 2004 is

$$\widehat{VOTE} = 51.053 + 0.877982GROWTH$$

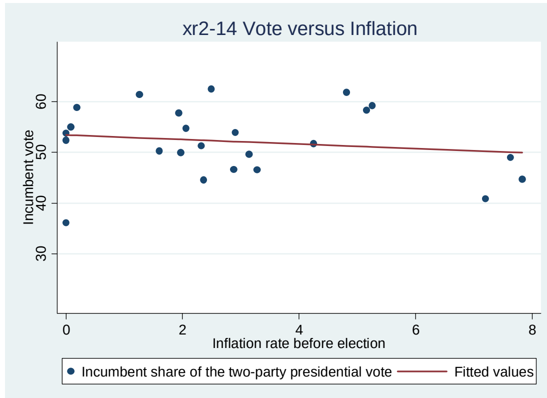
The actual 2008 value for growth is 0.220. Putting this into the estimated equation, we obtain the predicted vote share for the incumbent party:

$$\widehat{VOTE}_{2008} = 51.053 + 0.877982GROWTH_{2008} = 51.053 + 0.877982(0.220) = 51.246$$

This suggests that the incumbent party will maintain the majority vote in 2008. However, the actual vote share for the incumbent party for 2008 was 46.60, which is a long way short of the prediction; the incumbent party did not maintain the majority vote.

Exercise 2.14 (continued)

(d)



There appears to be a negative association between the two variables.

The estimated equation is:

$$\widehat{VOTE} = 53.408 - 0.444312 INFLATION$$

We estimate that a 1 percentage point increase in inflation during the incumbent party's first 15 quarters reduces the share of incumbent party's vote by 0.444 percentage points.

The estimated intercept suggests that when inflation is at 0% for that party's first 15 quarters, the expected share of votes won by the incumbent party is 53.4%; the incumbent party is predicted to maintain the majority vote when inflation, during its first 15 quarters, is at 0%.

CHAPTER 3

Exercise Answers

EXERCISE 3.3

(a) Reject H_0 because $t = 3.78 > t_c = 2.819$.

(b) Reject H_0 because $t = 3.78 > t_c = 2.508$.

(c) Do not reject H_0 because $t = 3.78 > t_c = -1.717$.

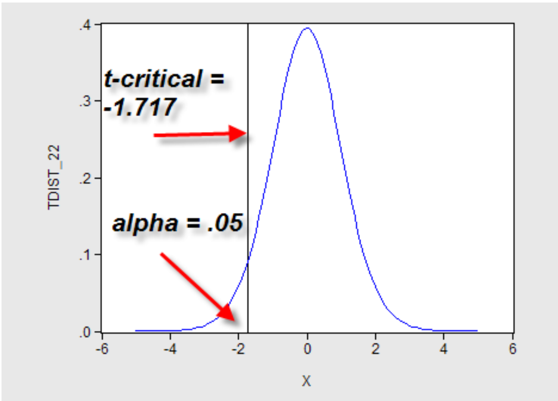


Figure xr3.3 One tail rejection region

(d) Reject H_0 because $t = -2.32 < -t_c = -2.074$.

(e) A 99% interval estimate of the slope is given by (0.079, 0.541)

EXERCISE 3.6

(a) We reject the null hypothesis because the test statistic value $t = 4.265 > t_c = 2.500$. The *p*-value is 0.000145

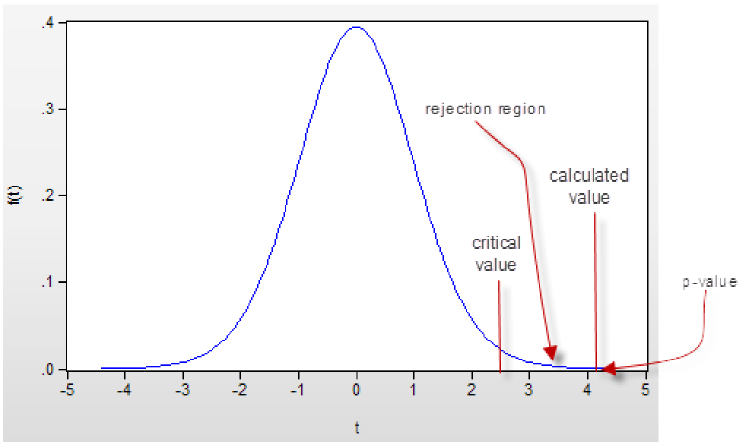


Figure xr3.6(a) Rejection region and *p*-value

(b) We do not reject the null hypothesis because the test statistic value $t = -2.093 > t_c = -2.500$. The *p*-value is 0.0238

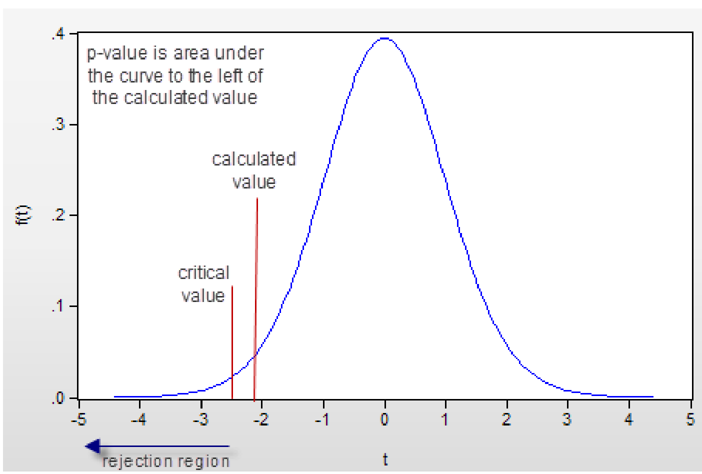


Figure xr3.6(b) Rejection region and *p*-value

(c) Since $t = -2.221 < t_c = -1.714$, we reject H_0 at a 5% significance level.

(d) A 95% interval estimate for δ_2 is given by $(-25.57, -0.91)$.

(e) Since $t = -3.542 < t_c = -2.500$, we reject H_0 at a 5% significance level.

(f) A 95% interval estimate for γ_2 is given by $(-22.36, -5.87)$.

EXERCISE 3.9

(a) We set up the hypotheses H_0: $\beta_2 = 0$ versus H_1: $\beta_2 > 0$. Since $t = 4.870 > 1.717$, we reject the null hypothesis.

(b) A 95% interval estimate for β_2 from the regression in part (a) is (0.509, 1.263).

(c) We set up the hypotheses H_0: $\beta_2 = 0$ versus H_1: $\beta_2 < 0$. Since $t = -0.741 > -1.717$, we do not reject the null hypothesis.

(d) A 95% interval estimate for β_2 from the regression in part (c) is $(-1.688, 0.800)$.

(e) We test $H_0 : \beta_1 \geq 50$ against the alternative $H_1 : \beta_1 < 50$. Since $t = 1.515 > -1.717$, we do not reject the null hypothesis.

(f) The 95% interval estimate is $(49.40, 55.64)$.

EXERCISE 3.13

(a)

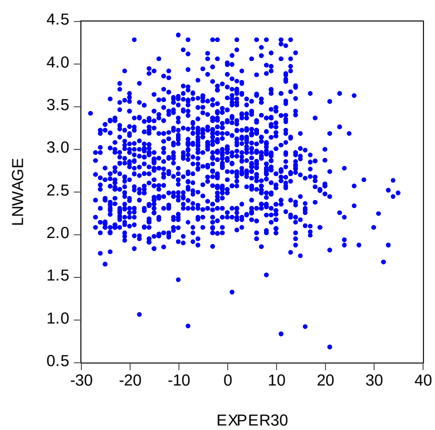


Figure xr3.13(a) Scatter plot of ln(*WAGE*) against *EXPER30*

(b) The estimated log-polynomial model is $\ln(WAGE) = 2.9826 - 0.0007088 EXPER30^2$.

We test $H_0 : \gamma_2 \geq 0$ against the alternative $H_1 : \gamma_2 < 0$. Because $t = -8.067 < -1.646$, we reject $H_0 : \gamma_2 \geq 0$.

Exercise 3.13 (continued)

(c)

$$\text{me}_{10} = \left.\frac{\widehat{d(WAGE)}}{d(EXPER)}\right|_{EXPER=10} = 0.4215$$

$$\text{me}_{30} = \left.\frac{\widehat{d(WAGE)}}{d(EXPER)}\right|_{EXPER=30} = 0.0$$

$$\text{me}_{50} = \left.\frac{\widehat{d(WAGE)}}{d(EXPER)}\right|_{EXPER=50} = -0.4215$$

(d)

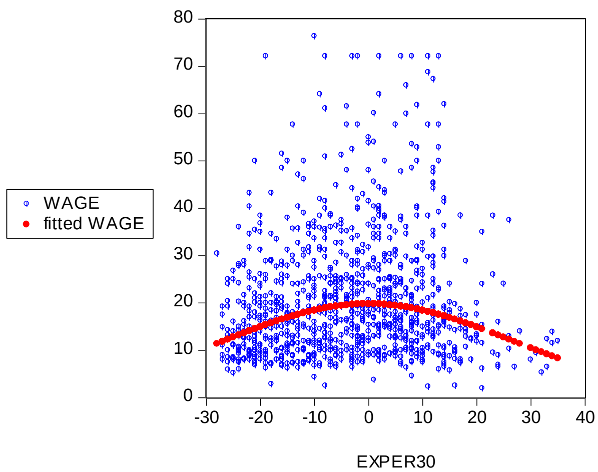


Figure xr3.13(d) Plot of fitted and actual values of *WAGE*

CHAPTER 4

Exercise Answers

EXERCISE 4.1

(a) $R^2 = 0.71051$

(b) $R^2 = 0.8455$

(c) $\hat{\sigma}^2 = 6.4104$

EXERCISE 4.2

(a) $\hat{y} = 5.83 + 17.38x^*$ where $x^* = \dfrac{x}{20}$
$(1.23)\ (2.34)$

(b) $\hat{y}^* = 0.1166 + 0.01738x$ where $\hat{y}^* = \dfrac{\hat{y}}{50}$
$(0.0246)\ (0.00234)$

(c) $\hat{y}^* = 0.2915 + 0.869x^*$ where $\hat{y}^* = \dfrac{\hat{y}}{20}$ and $x^* = \dfrac{x}{20}$
$(0.0615)\ (0.117)$

EXERCISE 4.9

(a) Equation 1: $\hat{y}_0 = 0.69538 + 0.015025 \times 48 = 1.417$

$\hat{y}_0 \pm t_{(0.975,45)} se(f) = 1.4166 \pm 2.0141 \times 0.25293 = (0.907, 1.926)$

Equation 2: $\hat{y}_0 = 0.56231 + 0.16961 \times \ln(48) = 1.219$

$\hat{y}_0 \pm t_{(0.975,45)} se(f) = 1.2189 \pm 2.0141 \times 0.28787 = (0.639, 1.799)$

Equation 3: $\hat{y}_0 = 0.79945 + 0.000337543 \times (48)^2 = 1.577$

$\hat{y}_0 \pm t_{(0.975,45)} se(f) = 1.577145 \pm 2.0141 \times 0.234544 = (1.105, 2.050)$

The actual yield in Chapman was 1.844.

Exercise 4.9 (continued)

(b) Equation 1: $\widehat{\frac{dy_t}{dt}} = 0.0150$

Equation 2: $\widehat{\frac{dy_t}{dt}} = 0.0035$

Equation 3: $\widehat{\frac{dy_t}{dt}} = 0.0324$

(c) Equation 1: $\widehat{\frac{dy_t}{dt}\frac{t}{y_t}} = 0.509$

Equation 2: $\widehat{\frac{dy_t}{dt}\frac{t}{y_t}} = 0.139$

Equation 3: $\widehat{\frac{dy_t}{dt}\frac{t}{y_t}} = 0.986$

(d) The slopes dy/dt and the elasticities $(dy/dt)\times(t/y)$ give the marginal change in yield and the percentage change in yield, respectively, that can be expected from technological change in the next year. The results show that the predicted effect of technological change is very sensitive to the choice of functional form.

EXERCISE 4.11

(a) The estimated regression model for the years 1916 to 2008 is:

$$\widehat{VOTE} = 50.8484 + 0.8859GROWTH \qquad R^2 = 0.5189$$
$$\text{(se)} \qquad (1.0125)\,(0.1819)$$

$$\widehat{VOTE}_{2008} = 51.043 \qquad\qquad VOTE_{2008} - \widehat{VOTE}_{2008} = -4.443$$

(b) The estimated regression model for the years 1916 to 2004 is:

$$\widehat{VOTE} = 51.0533 + 0.8780GROWTH \qquad R^2 = 0.5243$$
$$\text{(se)} \qquad (1.0379)\ (0.1825)$$

$$\widehat{VOTE}_{2008} = 51.246 \qquad\qquad f = VOTE_{2008} - \widehat{VOTE}_{2008} = -4.646$$

This prediction error is larger in magnitude than the least squares residual. This result is expected because the estimated regression in part (b) does not contain information about *VOTE* in the year 2008.

Exercise 4.11 (continued)

(c) $\widehat{VOTE}_{2008} \pm t_{(0.975,21)} \times \text{se}(f) = 51.2464 \pm 2.0796 \times 4.9185 = (41.018, 61.475)$

The actual 2008 outcome $VOTE_{2008} = 46.6$ falls within this prediction interval.

(d) $GROWTH = -1.086$

EXERCISE 4.13

(a) The regression results are:

$$\begin{array}{lll} \ln(PRICE) = 10.5938 + 0.000596\,SQFT & & \\ (\text{se}) & (0.0219) & (0.000013) \\ (t) & (484.84) & (46.30) \end{array}$$

The coefficient 0.000596 suggests an increase of one square foot is associated with a 0.06% increase in the price of the house.

$$\frac{dPRICE}{dSQFT} = 67.23$$

$$\text{elasticity} = \beta_2 \times \overline{SQFT} = 0.00059596 \times 1611.9682 = 0.9607$$

(b) The regression results are:

$$\begin{array}{lll} \ln(PRICE) = 4.1707 + 1.0066\ln(SQFT) & & \\ (\text{se}) & (0.1655) & (0.0225) \\ (t) & (25.20) & (44.65) \end{array}$$

The coefficient 1.0066 says that an increase in living area of 1% is associated with a 1% increase in house price.

The coefficient 1.0066 is the elasticity.

$$\frac{dPRICE}{dSQFT} = 70.444$$

(c) From the linear function, $R^2 = 0.672$.

From the log-linear function in part (a), $R_g^2 = 0.715$.

From the log-log function in part (b), $R_g^2 = 0.673$.

Exercise 4.13 (continued)

(d)

Jarque-Bera = 78.85

p -value = 0.0000

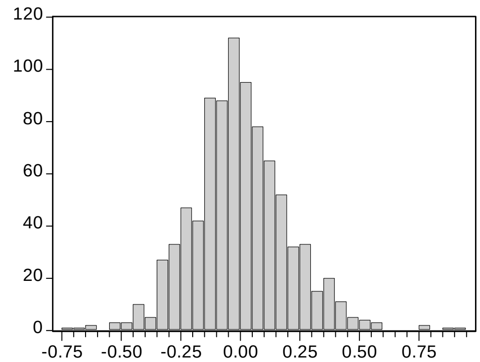


Figure xr4.13(d) Histogram of residuals for log-linear model

Jarque-Bera = 52.74

p -value = 0.0000

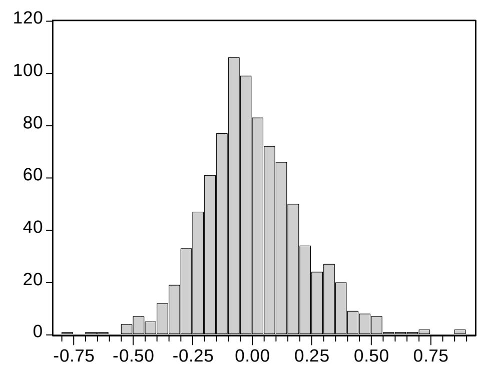


Figure xr4.13(d) Histogram of residuals for log-log model

Jarque-Bera = 2456

p -value = 0.0000

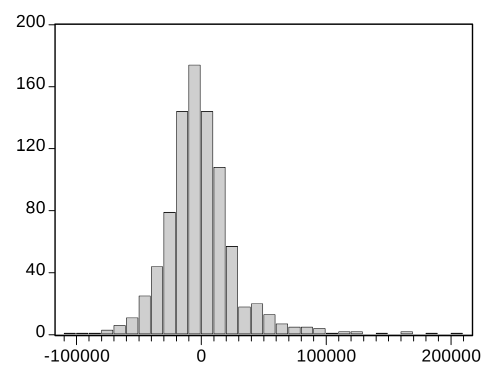
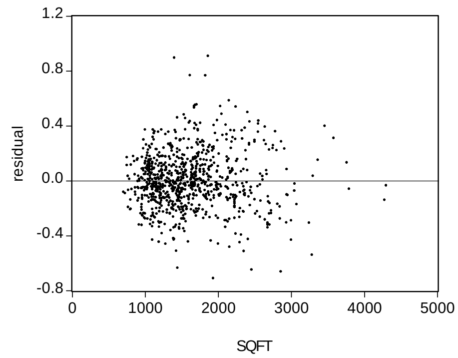


Figure xr4.13(d) Histogram of residuals for simple linear model

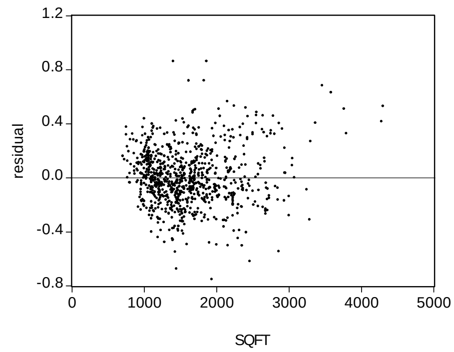
All Jarque-Bera values are significantly different from 0 at the 1% level of significance. We can conclude that the residuals are not compatible with an assumption of normality, particularly in the simple linear model.

Exercise 4.13 (continued)

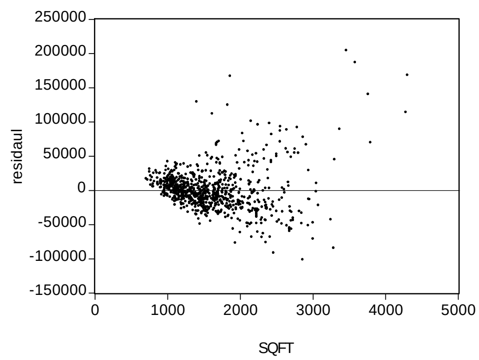
(e)



Residuals of log-linear model



Residuals of log-log model



Residuals of simple linear model

The residuals appear to increase in magnitude as *SQFT* increases. This is most evident in the residuals of the simple linear functional form. Furthermore, the residuals for the simple linear model in the area less than 1000 square feet are all positive indicating that perhaps the functional form does not fit well in this region.

(f) Prediction for log-linear model: $\widehat{PRICE} = 203{,}516$

Prediction for log-log model: $\widehat{PRICE} = 188{,}221$

Prediction for simple linear model: $\widehat{PRICE} = 201{,}365$

(g) The standard error of forecast for the log-linear model is $\text{se}(f) = 0.20363$.

The 95% confidence interval is: $(133{,}683;\ 297{,}316)$.

The standard error of forecast for the log-log model is $\text{se}(f) = 0.20876$.

The 95% confidence interval is $(122{,}267;\ 277{,}454)$.

The standard error of forecast for the simple linear model is $\text{se}(f) = 30348.26$.

The 95% confidence interval is $\left(141{,}801;\ 260{,}928\right)$.

Exercise 4.13 (continued)

(h) The simple linear model is not a good choice because the residuals are heavily skewed to the right and hence far from being normally distributed. It is difficult to choose between the other two models – the log-linear and log-log models. Their residuals have similar patterns and they both lead to a plausible elasticity of price with respect to changes in square feet, namely, a 1% change in square feet leads to a 1% change in price. The log-linear model is favored on the basis of its higher R_g^2 value, and its smaller standard deviation of the error, characteristics that suggest it is the model that best fits the data.

CHAPTER 5

Exercise Answers

EXERCISE 5.1

(a) $\bar{y} = 1, \quad \bar{x}_2 = 0, \quad \bar{x}_3 = 0$

x_{i2}^*	x_{i3}^*	y_i^*
0	1	0
1	−2	1
2	1	2
−2	0	−2
1	−1	−1
−2	−1	−2
0	1	1
−1	1	0
1	0	1

(b) $\sum y_i^* x_{i2}^* = 13, \quad \sum x_{i2}^{*2} = 16, \quad \sum y_i^* x_{i3}^* = 4, \quad \sum x_{i3}^{*2} = 10$

(c) $b_2 = 0.8125 \quad b_3 = 0.4 \quad b_1 = 1$

(d) $\hat{e} = \left(-0.4,\ 0.9875,\ -0.025,\ -0.375,\ -1.4125,\ 0.025,\ 0.6,\ 0.4125,\ 0.1875\right)$

(e) $\hat{\sigma}^2 = 0.6396$

(f) $r_{23} = 0$

(g) $\text{se}(b_2) = 0.1999$

(h) $SSE = 3.8375 \quad SST = 16 \quad SSR = 12.1625 \quad R^2 = 0.7602$

EXERCISE 5.2

(a) $b_2 \pm t_{(0.975,6)}\,\text{se}(b_2) = (0.3233, 1.3017)$

(b) We do not reject H_0 because $t = -0.9377$ and $|-0.9377| < 2.447 = t_{(0.975,\,6)}$.

EXERCISE 5.4

(a) The regression results are:

$$\widehat{WTRANS} = -0.0315 + 0.0414\ln(TOTEXP) - 0.0001AGE - 0.0130\,NK \qquad R^2 = 0.0247$$
$$(\text{se}) \qquad (0.0322) \quad (0.0071) \qquad\qquad (0.0004) \qquad (0.0055)$$

(b) The value $b_2 = 0.0414$ suggests that as $\ln(TOTEXP)$ increases by 1 unit the budget proportion for transport increases by 0.0414. Alternatively, one can say that a 10% increase in total expenditure will increase the budget proportion for transportation by 0.004. (See Chapter 4.3.3.) The positive sign of b_2 is according to our expectation because as households become richer they tend to use more luxurious forms of transport and the proportion of the budget for transport increases.

The value $b_3 = -0.0001$ implies that as the age of the head of the household increases by 1 year the budget share for transport decreases by 0.0001. The expected sign for b_3 is not clear. For a given level of total expenditure and a given number of children, it is difficult to predict the effect of age on transport share.

The value $b_4 = -0.0130$ implies that an additional child decreases the budget share for transport by 0.013. The negative sign means that adding children to a household increases expenditure on other items (such as food and clothing) more than it does on transportation. Alternatively, having more children may lead a household to turn to cheaper forms of transport.

(c) The *p*-value for testing $H_0: \beta_3 = 0$ against the alternative $H_1: \beta_3 \neq 0$ where β_3 is the coefficient of *AGE* is 0.869, suggesting that *AGE* could be excluded from the equation. Similar tests for the coefficients of the other two variables yield *p*-values less than 0.05.

(d) $R^2 = 0.0247$

(e) For a one-child household: $\widehat{WTRANS}_0 = 0.1420$

For a two-child household: $\widehat{WTRANS}_0 = 0.1290$

EXERCISE 5.8

(a) Equations describing the marginal effects of nitrogen and phosphorus on yield are

$$\frac{\partial E(YIELD)}{\partial(NITRO)} = 8.011 - 3.888NITRO - 0.567PHOS$$

$$\frac{\partial E(YIELD)}{\partial(PHOS)} = 4.800 - 1.556PHOS - 0.567NITRO$$

The marginal effect of both fertilizers declines – we have diminishing marginal products – and these marginal effects eventually become negative. Also, the marginal effect of one fertilizer is smaller, the larger is the amount of the other fertilizer that is applied.

(b) (i) The marginal effects when $NITRO = 1$ and $PHOS = 1$ are

$$\frac{\partial E(YIELD)}{\partial(NITRO)} = 3.556 \qquad \frac{\partial E(YIELD)}{\partial(PHOS)} = 2.677$$

(ii) The marginal effects when $NITRO = 2$ and $PHOS = 2$ are

$$\frac{\partial E(YIELD)}{\partial(NITRO)} = -0.899 \qquad \frac{\partial E(YIELD)}{\partial(PHOS)} = 0.554$$

When $NITRO = 1$ and $PHOS = 1$, the marginal products of both fertilizers are positive. Increasing the fertilizer applications to $NITRO = 2$ and $PHOS = 2$ reduces the marginal effects of both fertilizers, with that for nitrogen becoming negative.

(c) To test these hypotheses, the coefficients are defined according to the following equation

$$YIELD = \beta_1 + \beta_2 NITRO + \beta_3 PHOS + \beta_4 NITRO^2 + \beta_5 PHOS^2 + \beta_6 NITRO \times PHOS + e$$

(i) Testing $H_0: \beta_2 + 2\beta_4 + \beta_6 = 0$ against the alternative $H_1: \beta_2 + 2\beta_4 + \beta_6 \neq 0$, the *t*-value is $t = 7.367$. Since t > $t_c = t_{(0.975,\,21)} = 2.080$, we reject the null hypothesis and conclude that the marginal effect of nitrogen on yield is not zero when *NITRO* = 1 and *PHOS* = 1.

(ii) Testing $H_0: \beta_2 + 4\beta_4 + \beta_6 = 0$ against $H_1: \beta_2 + 4\beta_4 + \beta_6 \neq 0$, the *t*-value is $t = -1.660$. Since $|t| < 2.080 = t_{(0.975,\,21)}$, we do not reject the null hypothesis. A zero marginal yield with respect to nitrogen cannot be rejected when *NITRO* = 1 and *PHOS* = 2.

(iii) Testing $H_0: \beta_2 + 6\beta_4 + \beta_6 = 0$ against $H_1: \beta_2 + 6\beta_4 + \beta_6 \neq 0$, the *t*-value is $t = -8.742$. Since $|t| > 2.080 = t_{(0.975,\,21)}$, we reject the null hypothesis and conclude that the marginal product of yield to nitrogen is not zero when *NITRO* = 3 and *PHOS* = 1.

(d) The maximizing levels are $NITRO^* = 1.701$ and $PHOS^* = 2.465$. The yield maximizing levels of fertilizer are not necessarily the optimal levels. The optimal levels are those where the marginal cost of the inputs is equal to their marginal value product.

EXERCISE 5.15

(a) The estimated regression model is:

$$\widehat{VOTE} = 52.16 + 0.6434 GROWTH - 0.1721 INFLATION$$
$$\text{(se)} \quad (1.46) \quad (0.1656) \qquad\qquad (0.4290)$$

The hypothesis test results on the significance of the coefficients are:

$H_0: \beta_2 = 0 \quad H_1: \beta_2 > 0$ *p*-value = 0.0003 significant at 10% level

$H_0: \beta_3 = 0 \quad H_1: \beta_3 < 0$ *p*-value = 0.3456 not significant at 10% level

One-tail tests were used because more growth is considered favorable, and more inflation is considered not favorable, for re-election of the incumbent party.

(b) (i) For $INFLATION = 4$ and $GROWTH = -3$, $\widehat{VOTE}_0 = 49.54$.

(ii) For $INFLATION = 4$ and $GROWTH = 0$, $\widehat{VOTE}_0 = 51.47$.

(iii) For $INFLATION = 4$ and $GROWTH = 3$, $\widehat{VOTE}_0 = 53.40$.

(c) (i) When $INFLATION = 4$ and $GROWTH = -3$, the hypotheses are

$$H_0: \beta_1 - 3\beta_2 + 4\beta_3 \le 50 \qquad H_1: \beta_1 - 3\beta_2 + 4\beta_3 > 50$$

The calculated *t*-value is $t = -0.399$. Since $-0.399 < 2.457 = t_{(0.99,30)}$, we do not reject H_0. There is no evidence to suggest that the incumbent part will get the majority of the vote when $INFLATION = 4$ and $GROWTH = -3$.

(ii) When $INFLATION = 4$ and $GROWTH = 0$, the hypotheses are

$$H_0: \beta_1 + 4\beta_3 \le 50 \qquad H_1: \beta_1 + 4\beta_3 > 50$$

The calculated *t*-value is $t = 1.408$. Since $1.408 < 2.457 = t_{(0.99,30)}$, we do not reject H_0. There is insufficient evidence to suggest that the incumbent part will get the majority of the vote when $INFLATION = 4$ and $GROWTH = 0$.

(iii) When $INFLATION = 4$ and $GROWTH = 3$, the hypotheses are

$$H_0: \beta_1 + 3\beta_2 + 4\beta_3 \le 50 \qquad H_1: \beta_1 + 3\beta_2 + 4\beta_3 > 50$$

The calculated *t*-value is $t = 2.950$. Since $2.950 > 2.457 = t_{(0.99,30)}$, we reject H_0. We conclude that the incumbent part will get the majority of the vote when $INFLATION = 4$ and $GROWTH = 3$.

As a president seeking re-election, you would not want to conclude that you would be re-elected without strong evidence to support such a conclusion. Setting up re-election as the alternative hypothesis with a 1% significance level reflects this scenario.

EXERCISE 5.23

The estimated model is

$$\widehat{SCORE} = -39.594 + 47.024 \times AGE - 20.222 \times AGE^2 + 2.749 \times AGE^3$$
$$\text{(se)} \quad (28.153) \quad (27.810) \qquad (8.901) \qquad (0.925)$$

The within sample predictions, with age expressed in terms of years (not units of 10 years) are graphed in the following figure. They are also given in a table on page 27.

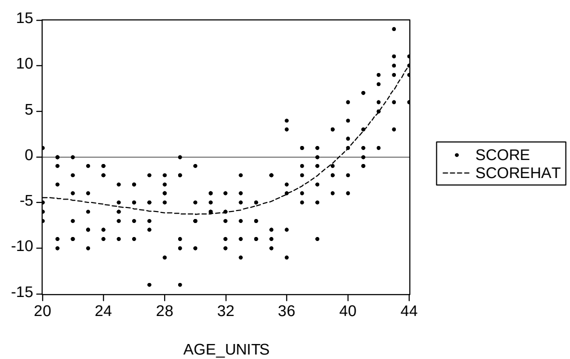


Figure xr5.23 Fitted line and observations

(a) We test $H_0 : \beta_4 = 0$. The *t*-value is 2.972, with corresponding *p*-value 0.0035. We therefore reject H_0 and conclude that the quadratic function is not adequate. For suitable values of β_2, β_3 and β_4, the cubic function can decrease at an increasing rate, then go past a point of inflection after which it decreases at a decreasing rate, and then it can reach a minimum and increase. These are characteristics worth considering for a golfer. That is, the golfer improves at an increasing rate, then at a decreasing rate, and then declines in ability. These characteristics are displayed in Figure xr5.23.

(b) (i) Age = 30

(ii) Between the ages of 20 and 25.

(iii) Between the ages of 25 and 30.

(iv) Age = 36.

(v) Age = 40.

(c) No. At the age of 70, the predicted score (relative to par) for Lion Forrest is 241.71. To break 100 it would need to be less than 28 $(= 100 - 72)$.

Exercise 5.23 (continued)

Predicted scores at different ages

age	predicted scores
20	– 4.4403
21	– 4.5621
22	– 4.7420
23	– 4.9633
24	– 5.2097
25	– 5.4646
26	– 5.7116
27	– 5.9341
28	– 6.1157
29	– 6.2398
30	– 6.2900
31	– 6.2497
32	– 6.1025
33	– 5.8319
34	– 5.4213
35	– 4.8544
36	– 4.1145
37	– 3.1852
38	– 2.0500
39	– 0.6923
40	0.9042
41	2.7561
42	4.8799
43	7.2921
44	10.0092

EXERCISE 5.24

(a) The coefficient estimates, standard errors, *t*-values and *p*-values are in the following table.

Dependent Variable: ln(*PROD*)

	Coeff	Std. Error	*t*-value	*p*-value
C	-1.5468	0.2557	-6.0503	0.0000
ln(*AREA*)	0.3617	0.0640	5.6550	0.0000
ln(*LABOR*)	0.4328	0.0669	6.4718	0.0000
ln(*FERT*)	0.2095	0.0383	5.4750	0.0000

All estimates have elasticity interpretations. For example, a 1% increase in labor will lead to a 0.4328% increase in rice output. A 1% increase in fertilizer will lead to a 0.2095% increase in rice output. All *p*-values are less than 0.0001 implying all estimates are significantly different from zero at conventional significance levels.

(b) Testing $H_0: \beta_2 = 0.5$ against $H_1: \beta_2 \neq 0.5$, the *t*-value is $t = -2.16$.
Since $-2.59 < -2.16 < 2.59 = t_{(0.995,348)}$, we do not reject H_0. The data are compatible with the hypothesis that the elasticity of production with respect to land is 0.5.

(c) A 95% interval estimate of the elasticity of production with respect to fertilizer is given by

$$b_4 \pm t_{(0.975,348)} \times \text{se}(b_4) = (0.134,\ 0.285)$$

This relatively narrow interval implies the fertilizer elasticity has been precisely measured.

(d) Testing $H_0: \beta_3 \leq 0.3$ against $H_1: \beta_3 > 0.3$, the *t*-value is $t = 1.99$. We reject H_0 because $1.99 > 1.649 = t_{(0.95,348)}$. There is evidence to conclude that the elasticity of production with respect to labor is greater than 0.3. Reversing the hypotheses and testing $H_0: \beta_3 \geq 0.3$ against $H_1: \beta_3 < 0.3$, leads to a rejection region of $t \leq -1.649$. The calculated *t*-value is $t = 1.99$. The null hypothesis is not rejected because $1.99 > -1.649$.

CHAPTER 6

Exercise Answers

EXERCISE 6.3

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

$$SSE = 42.8281 \qquad SST = 802.0243 \qquad SSR = 759.1962$$

(b) A 95% confidence interval for β_2 is

$$b_2 \pm t_{(0.975,17)}\text{se}(b_2) = (0.2343, 1.1639)$$

A 95% confidence interval for β_3 is

$$b_2 \pm t_{(0.975,17)}\text{se}(b_3) = (1.3704, 2.1834)$$

(c) To test H_0: $\beta_2 \geq 1$ against the alternative H_1: $\beta_2 < 1$, we calculate $t = -1.3658$. Since $-1.3658 > -1.740 = t_{(0.05,17)}$, 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 $F = 151$. Since $151 > 3.59 = F_{(0.95,2,17)}$, we reject H_0 and conclude that the hypothesis $\beta_2 = \beta_3 = 0$ is not compatible with the data.

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

$$t = \frac{(2b_2 - b_3)}{\text{se}(2b_2 - b_3)} = \frac{-0.37862}{0.59675} = -0.634$$

Since $-2.11 < -0.634 < 2.11 = t_{(0.025,17)}$, we do not reject H_0. There is no evidence to suggest that $2\beta_2 \neq \beta_3$.

EXERCISE 6.5

(a) 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 $F = 70.32$ .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, we reject the null hypothesis and conclude that education and experience have different effects on $\ln(WAGE)$.

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			

Exercise 6.10 (continued)

(c) Testing $H_0:\beta_2+\beta_3+\beta_4+\beta_5=0$ against $H_1:\beta_2+\beta_3+\beta_4+\beta_5\neq 0$, the value of the test statistic is F = 2.50, with a p-value of 0.127. The critical value is $F_{(0.95,1,25)}=4.24$. 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.

(d)(e) The results for parts (d) and (e) appear in the following table.

					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 6.12

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 6.20

(a) Testing $H_0:\beta_2=\beta_3$ against $H_1:\beta_2\neq\beta_3$, the calculated F-value is 0.342. We do not reject H_0 because $0.342<3.868=F_{(0.95,1,348)}$. 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.

Using a t-test, we fail to reject H_0 because $t=-0.585$ and the critical values are $t_{(0.025,348)}=-1.967$ and $t_{(0.975,348)}=1.967$. The p-value of the test is 0.559.

Exercise 6.20 (continued)

(b) Testing $H_0: \beta_2 + \beta_3 + \beta_4 = 1$ against $H_1: \beta_2 + \beta_3 + \beta_4 \neq 1$, the *F*-value is 0.0295. The *t*-value is $t = 0.172$. The critical values are $F_{(0.90,1,348)} = 2.72$ or $t_{(0.95,348)} = 1.649$ and $t_{(0.05,348)} = -1.649$. The *p*-value of the test is 0.864. The hypothesis of constant returns to scale cannot be rejected at a 10% significance level.

(c) The null and alternative hypotheses are

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

The critical value is $F_{(0.95,2,348)} = 3.02$. The calculated *F*-value is 0.183. The *p*-value of the test is 0.833. The joint null hypothesis of constant returns to scale and equality of land and labor elasticities cannot be rejected at a 5% significance level.

(d) The estimates and (standard errors) from the 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

	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

(a) $F = 7.40$ $F_c = 3.26$ $p\text{-value} = 0.002$

We reject H_0 and conclude that age does affect pizza expenditure.

(b) Point estimates, standard errors and 95% interval estimates for the marginal propensity to spend on pizza for different ages are given in the following table.

Age	Point Estimate	Standard Error	Confidence Interval	
			Lower	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

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) Point estimates, standard errors and 95% interval estimates for the marginal propensity to spend on pizza for different ages are given in the following table.

Age	Point Estimate	Standard Error	Confidence Interval	
			Lower	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

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.

For the joint test, $F = 5.136$. The corresponding *p*-value is 0.0048. The critical value at the 5% significance level is $F_{(0.95,3,35)} = 2.874$. 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.

Exercise 6.22 (continued)

(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.

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

CHAPTER 7

Exercise Answers

EXERCISE 7.2

(a) *Intercept*: At the beginning of the time period over which observations were taken, on a day which is not Friday, Saturday or a holiday, and a day which has neither a full moon nor a half moon, the estimated average number of emergency room cases was 93.69.

T: We estimate that the average number of emergency room cases has been increasing by 0.0338 per day, other factors held constant. The *t*-value is 3.06 and *p*-value = 0.003 < 0.01.

HOLIDAY: The average number of emergency room cases is estimated to go up by 13.86 on holidays, holding all else constant. The "holiday effect" is significant at the 0.05 level.

FRI and *SAT*: The average number of emergency room cases is estimated to go up by 6.9 and 10.6 on Fridays and Saturdays, respectively, holding all else constant. These estimated coefficients are both significant at the 0.01 level.

FULLMOON: The average number of emergency room cases is estimated to go up by 2.45 on days when there is a full moon, all else constant. However, a null hypothesis stating that a full moon has no influence on the number of emergency room cases would not be rejected at any reasonable level of significance.

NEWMOON: The average number of emergency room cases is estimated to go up by 6.4 on days when there is a new moon, all else held constant. However, a null hypothesis stating that a new moon has no influence on the number of emergency room cases would not be rejected at the usual 10% level, or smaller.

(b) There are very small changes in the remaining coefficients, and their standard errors, when *FULLMOON* and *NEWMOON* are omitted.

(c) Testing $H_0:\beta_6=\beta_7=0$ against $H_1:\beta_6 \text{ or } \beta_7 \text{ is nonzero}$, we find $F=1.29$. The 0.05 critical value is $F_{(0.95,\,2,\,222)}=3.307$, and corresponding *p*-value is 0.277. Thus, we do not reject the null hypothesis that new and full moons have no impact on the number of emergency room cases.

EXERCISE 7.5

(a) The estimated equation, with standard errors in parentheses, is

$$\widehat{\ln(PRICE)} = \underset{(0.0264)}{4.4638} + \underset{(0.0359)}{0.3334}UTOWN + \underset{(0.00104)}{0.03596}SQFT - \underset{(0.001414)}{0.003428}(SQFT \times UTOWN)$$

$$-\underset{(0.000218)}{0.000904}AGE + \underset{(0.00510)}{0.01899}POOL + \underset{(0.004140)}{0.006556}FPLACE \qquad R^2 = 0.8619$$

(b) Using this result for the coefficients of *SQFT* and *AGE*, we estimate that an additional 100 square feet of floor space is estimated to increase price by 3.6% for a house not in University town and 3.25% for a house in University town, holding all else fixed. A house which is a year older is estimated to sell for 0.0904% less, holding all else constant. The estimated coefficients of *UTOWN*, *AGE*, and the slope-indicator variable *SQFT_UTOWN* are significantly different from zero at the 5% level of significance.

(c) An approximation of the percentage change in price due to the presence of a pool is 1.90%. The exact percentage change in price due to the presence of a pool is estimated to be 1.92%.

(d) An approximation of the percentage change in price due to the presence of a fireplace is 0.66%. The exact percentage change in price due to the presence of a fireplace is also 0.66%.

(e) The percentage change in price attributable to being near the university, for a 2500 square-feet home, is 28.11%.

EXERCISE 7.9

(a) The estimated average test scores are:

regular sized class with no aide = 918.0429
regular sized class with aide = 918.3568
small class = 931.9419

From the above figures, the average scores are higher with the small class than the regular class. The effect of having a teacher aide is negligible.

The results of the estimated models for parts (b)-(g) are summarized in the table on page 38.

(b) The coefficient of *SMALL* is the difference between the average of the scores in the regular sized classes (918.36) and the average of the scores in small classes (931.94). That is b_2 = 931.9419 − 918.0429 = 13.899. Similarly the coefficient of *AIDE* is the difference between the average score in classes with an aide and regular classes. The *t*-value for the significance of β_3 is $t = 0.136$. The critical value at the 5% significance level is 1.96. We cannot conclude that there is a significant difference between test scores in a regular class and a class with an aide.

Exercise 7.9 (continued)

Exercise 7-9

	(1) (b)	(2) (c)	(3) (d)	(4) (e)	(5) (g)
C	918.043***	904.721***	923.250***	931.755***	918.272***
	(1.641)	(2.228)	(3.121)	(3.940)	(4.357)
SMALL	13.899***	14.006***	13.896***	13.980***	15.746***
	(2.409)	(2.395)	(2.294)	(2.302)	(2.096)
AIDE	0.314	-0.601	0.698	1.002	1.782
	(2.310)	(2.306)	(2.209)	(2.217)	(2.025)
TCHEXPER		1.469***	1.114***	1.156***	0.720***
		(0.167)	(0.161)	(0.166)	(0.167)
BOY			-14.045***	-14.008***	-12.121***
			(1.846)	(1.843)	(1.662)
FREELUNCH			-34.117***	-32.532***	-34.481***
			(2.064)	(2.126)	(2.011)
WHITE_ASIAN			11.837***	16.233***	25.315***
			(2.211)	(2.780)	(3.510)
TCHWHITE				-7.668***	-1.538
				(2.842)	(3.284)
TCHMASTERS				-3.560*	-2.621
				(2.019)	(2.184)
SCHURBAN				-5.750**	.
				(2.858)	.
SCHRURAL				-7.006***	.
				(2.559)	.
N	5786	5766	5766	5766	5766
adj. R-sq	0.007	0.020	0.101	0.104	0.280
BIC	66169.500	65884.807	65407.272	65418.626	64062.970
SSE	31232400.314	30777099.287	28203498.965	28089837.947	22271314.955

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

(c) The *t*-statistic for the significance of the coefficient of *TCHEXPER* is 8.78 and we reject the null hypothesis that a teacher's experience has no effect on total test scores. The inclusion of this variable has a small impact on the coefficient of *SMALL*, and the coefficient of *AIDE* has gone from positive to negative. However *AIDE*'s coefficient is not significantly different from zero and this change is of negligible magnitude, so the sign change is not important.

(d) The inclusion of *BOY*, *FREELUNCH* and *WHITE_ASIAN* has little impact on the coefficients of *SMALL* and *AIDE*. The variables themselves are statistically significant at the $\alpha = 0.01$ level of significance.

Exercise 7.9 (continued)

(e) The regression result suggests that *TCHWHITE*, *SCHRURAL* and *SCHURBAN* are significant at the 5% level and *TCHMASTERS* is significant at the 10% level. The inclusion of these variables has only a very small and negligible effect on the estimated coefficients of *AIDE* and *SMALL*.

(f) The results found in parts (c), (d) and (e) suggest that while some additional variables were found to have a significant impact on total scores, the estimated advantage of being in small classes, and the insignificance of the presence of a teacher aide, is unaffected. The fact that the estimates of the key coefficients did not change is support for the randomization of student assignments to the different class sizes. The addition or deletion of uncorrelated factors does not affect the estimated effect of the key variables.

(g) We find that inclusion of the school effects increases the estimates of the benefits of small classes and the presence of a teacher aide, although the latter effect is still insignificant statistically. The *F*-test of the joint significance of the school indicators is 19.15. The 5% *F*-critical value for 78 numerator and 5679 denominator degrees of freedom is 1.28, thus we reject the null hypothesis that all the school effects are zero, and conclude that at least some are not zero.

The variables *SCHURBAN* and *SCHRURAL* drop out of this model because they are exactly collinear with the included 78 indicator variables.

EXERCISE 7.14

(a) We expect the parameter estimate for the dummy variable *PERSON* to be positive because of reputation and knowledge of the incumbent. However, it could be negative if the incumbent was, on average, unpopular and/or ineffective. We expect the parameter estimate for *WAR* to be positive reflecting national feeling during and immediately after first and second world wars.

(b) The regression functions for each value of *PARTY* are:

$$E(VOTE \mid PARTY = 1) = (\beta_1 + \beta_7) + \beta_2 GROWTH + \beta_3 INFLATION + \beta_4 GOODNEWS + \beta_5 PERSON + \beta_6 DURATION + \beta_8 WAR$$

$$E(VOTE \mid PARTY = -1) = (\beta_1 - \beta_7) + \beta_2 GROWTH + \beta_3 INFLATION + \beta_4 GOODNEWS + \beta_5 PERSON + \beta_6 DURATION + \beta_8 WAR$$

The intercept when there is a Democrat incumbent is $\beta_1 + \beta_7$. When there is a Republican incumbent it is $\beta_1 - \beta_7$. Thus, the effect of *PARTY* on the vote is $2\beta_7$ with the sign of β_7 indicating whether incumbency favors Democrats $(\beta_7 > 0)$ or Republicans $(\beta_7 < 0)$.

Exercise 7.14 (continued)

(c) The estimated regression using observations for 1916-2004 is

$$\widehat{VOTE} = 47.2628 + 0.6797GROWTH - 0.6572INFLATION + 1.0749GOODNEWS$$
$$(\text{se}) \quad (2.5384)\ (0.1107) \qquad (0.2914) \qquad (0.2493)$$

$$+3.2983PERSON - 3.3300DURATION - 2.6763PARTY + 5.6149WAR$$
$$(1.4081) \qquad (1.2124) \qquad (0.6264) \qquad (2.6879)$$

The signs are as expected. Can you explain why? All the estimates are statistically significant at a 1% level of significance except for *INFLATION, PERSON, DURATION* and *WAR*. The coefficients of *INFLATION, DURATION* and *PERSON* are statistically significant at a 5% level, however. The coefficient of *WAR* is statistically insignificant at a level of 5%. Lastly, an R^2 of 0.9052 suggests that the model fits the data very well.

(d) Using the data for 2008, and based on the estimates from part (c), we summarize the actual and predicted vote as follows, along with a listing of the values of the explanatory variables.

vote	growth	inflation	goodnews	person	duration	party	war	votehat
46.6	.22	2.88	3	0	1	-1	0	48.09079

Thus, we predict that the Republicans, as the incumbent party, will lose the 2008 election with 48.091% of the vote. This prediction was correct, with Democrat Barack Obama defeating Republican John McCain with 52.9% of the popular vote to 45.7%.

(e) A 95% confidence interval for the vote in the 2008 election is

$$\widehat{VOTE}_{2012} \pm t_{(0.975,15)} \times \text{se}(f) = (42.09, 54.09)$$

(f) For the 2012 election the Democratic party will have been in power for one term and so we set *DURATION* = 1 and *PARTY* = 1. Also, the incumbent, Barack Obama, is running for election and so we set *PERSON* = 1. *WAR* = 0. We use the value of inflation 3.0% anticipating higher rates of inflation after the policy stimulus. We consider 3 scenarios for *GROWTH* and *GOODNEWS* representing good economic outcomes, moderate and poor, if there is a "double-dip" recession. The values and the prediction intervals based on regression estimates with data from 1916-2008, are

GROWTH	*INFLATION*	*GOODNEWS*	lb	vote	ub
3.5	3	6	45.6	51.5	57.3
1	3	3	40.4	46.5	52.5
-3	3	1	35.0	41.5	48.0

We see that if there is good economic performance, then President Obama can expect to be re-elected. If there is poor economic performance, then we predict he will lose the election with the upper bound of the 95% prediction interval for a vote in his favor being only 48%. In the intermediate case, with only modest growth and less good news, then we predict he will lose the election, though the interval estimate upper bound is greater than 50%, meaning that anything could happen.

EXERCISE 7.16

(a) The histogram for *PRICE* is positively skewed. On the other hand, the logarithm of *PRICE* is much less skewed and is more symmetrical. Thus, the histogram of the logarithm of *PRICE* is closer in shape to a normal distribution than the histogram of *PRICE*.

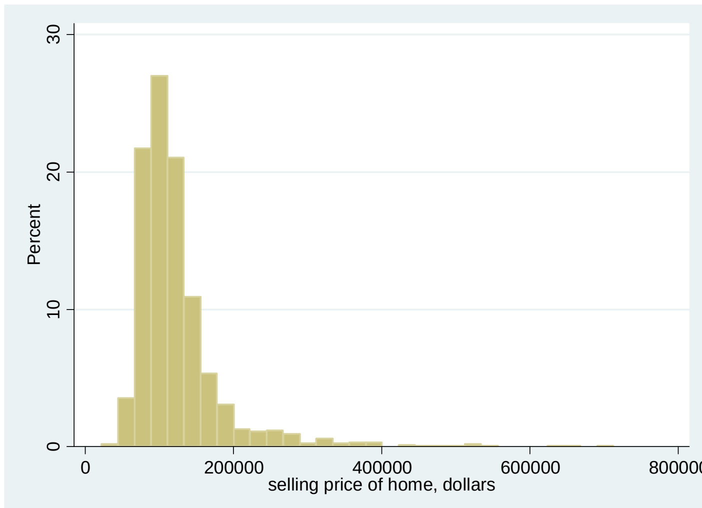


Figure xr7.16(a) Histogram of *PRICE*

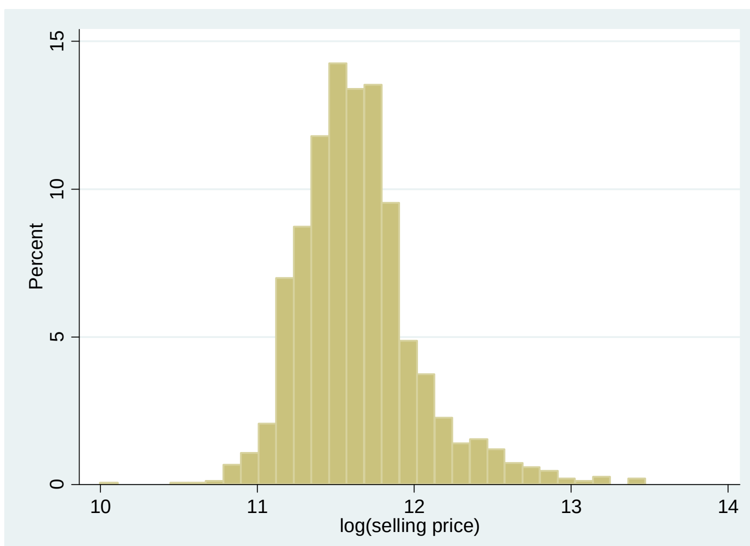


Figure xr7.16(b) Histogram of ln(*PRICE*)

Exercise 7.16 (continued)

(b) The estimated equation is

$$\widehat{\ln\left(PRICE/1000\right)} = \underset{(0.0373)}{3.9860} + \underset{(0.0017)}{0.0539}LIVAREA - \underset{(0.0114)}{0.0382}BEDS - \underset{(0.0165)}{0.0103}BATHS$$
(se)

$$+ \underset{(0.0255)}{0.2531}LGELOT - \underset{(0.0005)}{0.0013}AGE + \underset{(0.0231)}{0.0787}POOL$$

All coefficients are significant with the exception of that for *BATHS*. All signs are reasonable: increases in living area, larger lot sizes and the presence of a pool are associated with higher selling prices. Older homes depreciate and have lower prices. Increases in the number of bedrooms, holding all else fixed, implies smaller bedrooms which are less valued by the market. The number of baths is statistically insignificant, so its negative sign cannot be reliably interpreted.

(c) The price of houses on lot sizes greater than 0.5 acres is approximately $100\left(\exp(-0.2531)-1\right) = 28.8\%$ larger than the price of houses on lot sizes less than 0.5 acres.

(d) The estimated regression after including the interaction term is:

$$\widehat{\ln\left(PRICE/1000\right)} = \underset{(0.0370)}{3.9649} + \underset{(0.0019)}{0.0589}LIVAREA - \underset{(0.0113)}{0.0480}BEDS - \underset{(0.0164)}{0.0201}BATHS$$
(se)

$$+ \underset{(0.0632)}{0.6134}LGELOT - \underset{(0.0005)}{0.0016}AGE + \underset{(0.0228)}{0.0853}POOL$$

$$- \underset{(0.0026)}{0.0161}LGELOT \times LIVAREA$$

Interpretation of the coefficient of $LGELOT \times LIVAREA$:

The estimated marginal effect of an increase in living area of 100 square feet in a house on a lot of less than 0.5 acres is 5.89%, all else constant. The same increase for a house on a large lot is estimated to increase the house selling price by 1.61% less, or 4.27%. However, note that by adding this interaction variable into the model, the coefficient of *LGELOT* increases dramatically. The inclusion of the interaction variable separates the effect of the larger lot from the fact that larger lots usually contain larger homes.

(e) The value of the *F*-statistic is

$$F = \frac{\left(SSE_R - SSE_U\right)/J}{SSE_U/(N-K)/} = \frac{\left(72.0633\text{-}65.4712\right)/6}{65.4712/\left(1488\right)} = 24.97$$

The 5% critical *F* value is $F_{(0.95,6,1488)} = 2.10$. Thus, we conclude that the pricing structure for houses on large lots is not the same as that on smaller lots.

Exercise 7.16 (continued)

A summary of the alternative model estimations follows.

Exercise 7-16

	(1) *LGELOT*=1	(2) *LGELOT*=0	(3) Rest	(4) Unrest
C	4.4121***	3.9828***	3.9794***	3.9828***
	(0.183)	(0.037)	(0.039)	(0.038)
LIVAREA	0.0337***	0.0604***	0.0607***	0.0604***
	(0.005)	(0.002)	(0.002)	(0.002)
BEDS	-0.0088	-0.0522***	-0.0594***	-0.0522***
	(0.048)	(0.012)	(0.012)	(0.012)
BATHS	0.0827	-0.0334**	-0.0262	-0.0334*
	(0.066)	(0.017)	(0.017)	(0.017)
AGE	-0.0018	-0.0016***	-0.0008*	-0.0016***
	(0.002)	(0.000)	(0.000)	(0.000)
POOL	0.1259*	0.0697***	0.0989***	0.0697***
	(0.074)	(0.024)	(0.024)	(0.025)
LGELOT				0.4293***
				(0.141)
LOT_AREA				-0.0266***
				(0.004)
LOT_BEDS				0.0434
				(0.037)
LOT_BATHS				0.1161**
				(0.052)
LOT_AGE				-0.0002
				(0.001)
LOT_POOL				0.0562
				(0.060)
N	95	1405	1500	1500
adj. R-sq	0.676	0.608	0.667	0.696
BIC	50.8699	-439.2028	-252.8181	-352.8402
SSE	7.1268	58.3445	72.0633	65.4712

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
** *LOT_X* indicates interaction between *LGELOT* and *X*

CHAPTER 8

Exercise Answers

EXERCISE 8.7

(a) $\sum x_i = 0 \qquad \sum y_i = 31.1 \qquad \sum x_i y_i = 89.35 \qquad \sum x_i^2 = 52.34$

$\bar{x} = 0 \qquad \bar{y} = 3.8875$

$$b_2 = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{8\times 89.35 - 0\times 31.1}{8\times 52.34 - (0)^2} = 1.7071$$

$$b_1 = \bar{y} - b_2\bar{x} = 3.8875 - 1.7071\times 0 = 3.8875$$

(b)

observation	$\hat{e}$	$\ln(\hat{e}^2)$	$z\times\ln(\hat{e}^2)$
1	– 1.933946	1.319125	4.353113
2	0.733822	– 0.618977	– 0.185693
3	9.549756	4.513031	31.591219
4	– 1.714707	1.078484	5.068875
5	– 3.291665	2.382787	4.527295
6	3.887376	2.715469	18.465187
7	– 3.484558	2.496682	5.742369
8	– 3.746079	2.641419	16.905082

(c) We use the estimating equation

$$\ln(\hat{e}_i^2) = \alpha z_i + v_i$$

Using least squares to estimate α from this model is equivalent to a simple linear regression without a constant term. The least squares estimate for α is

Exercise 8.7(c) (continued)

$$\hat{\alpha} = \frac{\sum_{i=1}^{8}\left(z_i \ln(\hat{e}_i^2)\right)}{\sum_{i=1}^{8} z_i^2} = \frac{86.4674}{178.17} = 0.4853$$

(d) Variance estimates are given by the predictions $\hat{\sigma}_i^2 = \exp(\hat{\alpha} z_i) = \exp(0.4853 \times z_i)$. These values and those for the transformed variables

$$y_i^* = \left(\frac{y_i}{\hat{\sigma}_i}\right), \qquad x_i^* = \left(\frac{x_i}{\hat{\sigma}_i}\right)$$

are given in the following table.

observation	$\hat{\sigma}_i^2$	y_i^*	x_i^*
1	4.960560	0.493887	− 0.224494
2	1.156725	− 0.464895	− 2.789371
3	29.879147	3.457624	0.585418
4	9.785981	− 0.287700	− 0.575401
5	2.514531	4.036003	2.144126
6	27.115325	0.345673	− 0.672141
7	3.053260	2.575316	1.373502
8	22.330994	− 0.042323	− 0.042323

(e) From Exercise 8.2, the generalized least squares estimate for β_2 is

$$\hat{\beta}_2 = \frac{\dfrac{\sum y_i^* x_i^*}{\sum \sigma_i^{-2}} - \left(\dfrac{\sum \sigma_i^{-2} y_i}{\sum \sigma_i^{-2}}\right)\left(\dfrac{\sum \sigma_i^{-2} x_i}{\sum \sigma_i^{-2}}\right)}{\dfrac{\sum x_i^{*2}}{\sum \sigma_i^{-2}} - \left(\dfrac{\sum \sigma_i^{-2} x_i}{\sum \sigma_i^{-2}}\right)^2}$$

$$= \frac{\dfrac{15.33594}{2.008623} - 2.193812 \times (-0.383851)}{\dfrac{15.442137}{2.008623} - (-0.383851)^2}$$

$$= \frac{8.477148}{7.540580}$$

$$= 1.1242$$

The generalized least squares estimate for β_1 is

$$\hat{\beta}_1 = \frac{\sum \sigma_i^{-2} y_i}{\sum \sigma_i^{-2}} - \left(\frac{\sum \sigma_i^{-2} x_i}{\sum \sigma_i^{-2}}\right)\hat{\beta}_2 = 2.193812 - (-0.383851) \times 1.1242 = 2.6253$$

EXERCISE 8.10

(a) The transformed model corresponding to the variance assumption $\sigma_i^2 = \sigma^2 x_i$ is

$$\frac{y_i}{\sqrt{x_i}} = \beta_1\left(\frac{1}{\sqrt{x_i}}\right) + \beta_2\sqrt{x_i} + e_i^* \qquad \text{where } e_i^* = \left(\frac{e_i}{\sqrt{x_i}}\right)$$

Squaring the residuals and regressing them on x_i gives

$$\hat{e}^{*2} = -123.79 + 23.35x \qquad R^2 = 0.13977$$

$$\chi^2 = N \times R^2 = 40 \times 0.13977 = 5.59$$

A null hypothesis of no heteroskedasticity is rejected. The variance assumption $\sigma_i^2 = \sigma^2 x_i$ was not adequate to eliminate heteroskedasticity.

(b) The transformed model used to obtain the estimates in (8.27) is

$$\frac{y_i}{\hat{\sigma}_i} = \beta_1\left(\frac{1}{\hat{\sigma}_i}\right) + \beta_2\frac{x_i}{\hat{\sigma}_i} + e_i^* \qquad \text{where } e_i^* = \left(\frac{e_i}{\hat{\sigma}_i}\right)$$

$$\hat{\sigma}_i = \sqrt{\exp(0.93779596 + 2.32923872 \times \ln(x_i)}$$

Squaring the residuals and regressing them on x_i gives

$$\hat{e}^{*2} = 1.117 + 0.05896x \qquad R^2 = 0.02724$$

$$\chi^2 = N \times R^2 = 40 \times 0.02724 = 1.09$$

A null hypothesis of no heteroskedasticity is not rejected. The variance assumption $\sigma_i^2 = \sigma^2 x_i^\gamma$ is adequate to eliminate heteroskedasticity.

EXERCISE 8.13

(a) For the model $C_{1t} = \beta_1 + \beta_2 Q_{1t} + \beta_3 Q_{1t}^2 + \beta_4 Q_{1t}^3 + e_{1t}$, where $\text{var}(e_{1t}) = \sigma^2 Q_{1t}$, the generalized least squares estimates of β_1, β_2, β_3 and β_4 are:

	estimated coefficient	standard error
β_1	93.595	23.422
β_2	68.592	17.484
β_3	–10.744	3.774
β_4	1.0086	0.2425

(b) The calculated F value for testing the hypothesis that $\beta_1 = \beta_4 = 0$ is 108.4. The 5% critical value from the $F_{(2,24)}$ distribution is 3.40. Since the calculated F is greater than the critical F, we reject the null hypothesis that $\beta_1 = \beta_4 = 0$.

Exercise 8.13 (continued)

(c) The average cost function is given by

$$\frac{C_{1t}}{Q_{1t}} = \beta_1\left(\frac{1}{Q_{1t}}\right) + \beta_2 + \beta_3 Q_{1t} + \beta_4 Q_{1t}^2 + \frac{e_t}{Q_{1t}}$$

Thus, if $\beta_1 = \beta_4 = 0$, average cost is a linear function of output.

(d) The average cost function is an appropriate transformed model for estimation when heteroskedasticity is of the form $\text{var}(e_{1t}) = \sigma^2 Q_{1t}^2$.

EXERCISE 8.14

(a) The least squares estimated equations are

$$\begin{array}{lllll} \hat{C}_1 = 72.774 + 83.659Q_1 & -13.796Q_1^2 & +1.1911Q_1^3 & & \hat{\sigma}_1^2 = 324.85 \\ \text{(se)} \qquad\qquad (23.655) & (4.597) & (0.2721) & & SSE_1 = 7796.49 \end{array}$$

$$\begin{array}{lllll} \hat{C}_2 = 51.185 + 108.29Q_2 & -20.015Q_2^2 & +1.6131Q_2^3 & & \hat{\sigma}_2^2 = 847.66 \\ \text{(se)} \qquad\qquad (28.933) & (6.156) & (0.3802) & & SSE_2 = 20343.83 \end{array}$$

To see whether the estimated coefficients have the expected signs consider the marginal cost function

$$MC = \frac{dC}{dQ} = \beta_2 + 2\beta_3 Q + 3\beta_4 Q^2$$

We expect $MC > 0$ when $Q = 0$; thus, we expect $\beta_2 > 0$. Also, we expect the quadratic MC function to have a minimum, for which we require $\beta_4 > 0$. The slope of the MC function is $d(MC)/dQ = 2\beta_3 + 6\beta_4 Q$. For this slope to be negative for small Q (decreasing MC), and positive for large Q (increasing MC), we require $\beta_3 < 0$. Both our least-squares estimated equations have these expected signs. Furthermore, the standard errors of all the coefficients except the constants are quite small indicating reliable estimates. Comparing the two estimated equations, we see that the estimated coefficients and their standard errors are of similar magnitudes, but the estimated error variances are quite different.

(b) Testing $H_0: \sigma_1^2 = \sigma_2^2$ against $H_1: \sigma_1^2 \neq \sigma_2^2$ is a two-tail test. The critical values for performing a two-tail test at the 10% significance level are $F_{(0.05,24,24)} = 0.0504$ and $F_{(0.95,24,24)} = 1.984$. The value of the F statistic is

$$F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{847.66}{324.85} = 2.61$$

Since $F > F_{(0.95,24,24)}$, we reject H_0 and conclude that the data do not support the proposition that $\sigma_1^2 = \sigma_2^2$.

Exercise 8.14 (continued)

(c) Since the test outcome in (b) suggests $\sigma_1^2 \neq \sigma_2^2$, but we are assuming both firms have the same coefficients, we apply generalized least squares to the combined set of data, with the observations transformed using $\hat{\sigma}_1$ and $\hat{\sigma}_2$. The estimated equation is

$$\hat{C} = 67.270 + 89.920Q - 15.408Q^2 + 1.3026Q^3$$
$$\text{(se)} \qquad (16.973) \quad (3.415) \quad (0.2065)$$

<u>Remark</u>: Some automatic software commands will produce slightly different results if the transformed error variance is restricted to be unity or if the variables are transformed using variance estimates from a pooled regression instead of those from part (a).

(d) Although we have established that $\sigma_1^2 \neq \sigma_2^2$, it is instructive to first carry out the test for

$$H_0: \beta_1 = \delta_1, \quad \beta_2 = \delta_2, \quad \beta_3 = \delta_3, \quad \beta_4 = \delta_4$$

under the assumption that $\sigma_1^2 = \sigma_2^2$, and then under the assumption that $\sigma_1^2 \neq \sigma_2^2$.

Assuming that $\sigma_1^2 = \sigma_2^2$, the test is equivalent to the Chow test discussed on pages 268-270 of the text. The test statistic is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)}$$

where SSE_U is the sum of squared errors from the full dummy variable model. The dummy variable model does not have to be estimated, however. We can also calculate SSE_U as the sum of the SSE from separate least squares estimation of each equation. In this case

$$SSE_U = SSE_1 + SSE_2 = 7796.49 + 20343.83 = 28140.32$$

The restricted model has not yet been estimated under the assumption that $\sigma_1^2 = \sigma_2^2$. Doing so by combining all 56 observations yields $SSE_R = 28874.34$. The *F*-value is given by

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(28874.34 - 28140.32)/4}{28140.32/(56-8)} = 0.313$$

The corresponding χ^2-value is $\chi^2 = 4 \times F = 1.252$. These values are both much less than their respective 5% critical values $F_{(0.95,4,48)} = 2.565$ and $\chi^2_{(0.95,4)} = 9.488$. There is no evidence to suggest that the firms have different coefficients. In the formula for *F*, note that the number of observations *N* is the total number from both firms, and *K* is the number of coefficients from both firms.

The above test is not valid in the presence of heteroskedasticity. It could give misleading results. To perform the test under the assumption that $\sigma_1^2 \neq \sigma_2^2$, we follow the same steps, but we use values for *SSE* computed from transformed residuals. For restricted estimation from part (c) the result is $SSE_R^* = 49.2412$. For unrestricted estimation, we have the interesting result

Exercise 8.14(d) (continued)

$$SSE_U^* = \frac{SSE_1}{\hat{\sigma}_1^2} + \frac{SSE_2}{\hat{\sigma}_2^2} = \frac{(N_1 - K_1)\times\hat{\sigma}_1^2}{\hat{\sigma}_1^2} + \frac{(N_2 - K_2)\times\hat{\sigma}_2^2}{\hat{\sigma}_2^2} = N_1 - K_1 + N_2 - K_2 = 48$$

Thus,

$$F = \frac{(49.2412 - 48)/4}{48/48} = 0.3103 \qquad \text{and} \qquad \chi^2 = 1.241$$

The same conclusion is reached. There is no evidence to suggest that the firms have different coefficients.

The χ^2 and F test values can also be conveniently calculated by performing a Wald test on the coefficients after running weighted least squares on a pooled model that includes dummy variables to accommodate the different coefficients.

EXERCISE 8.15

(a) To estimate the two variances using the variance model specified, we first estimate the equation

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 METRO_i + e_i$$

From this equation we use the squared residuals to estimate the equation

$$\ln(\hat{e}_i^2) = \alpha_1 + \alpha_2 METRO_i + v_i$$

The estimated parameters from this regression are $\hat{\alpha}_1 = 1.508448$ and $\hat{\alpha}_2 = 0.338041$. Using these estimates, we have

$$METRO = 0 \quad \Rightarrow \quad \hat{\sigma}_R^2 = \exp(1.508448 + 0.338041\times 0) = 4.519711$$

$$METRO = 1, \quad \Rightarrow \quad \hat{\sigma}_M^2 = \exp(1.508448 + 0.338041\times 1) = 6.337529$$

These error variance estimates are much smaller than those obtained from separate sub-samples ($\hat{\sigma}_M^2 = 31.824$ and $\hat{\sigma}_R^2 = 15.243$). One reason is the bias factor from the exponential function – see page 317 of the text. Multiplying $\hat{\sigma}_M^2 = 6.3375$ and $\hat{\sigma}_R^2 = 4.5197$ by the bias factor $\exp(1.2704)$ yields $\hat{\sigma}_M^2 = 22.576$ and $\hat{\sigma}_R^2 = 16.100$. These values are closer, but still different from those obtained using separate sub-samples. The differences occur because the residuals from the combined model are different from those from the separate sub-samples.

(b) To use generalized least squares, we use the estimated variances above to transform the model in the same way as in (8.35). After doing so the regression results are, with standard errors in parentheses

$$\widehat{WAGE}_i = -9.7052 + 1.2185 EDUC_i + 0.1328 EDUC_i + 1.5301 METRO_i$$

$$\text{(se)} \quad (1.0485)\ (0.0694) \qquad (0.0150) \qquad (0.3858)$$

Exercise 8.15(b) (continued)

The magnitudes of these estimates and their standard errors are almost identical to those in equation (8.36). Thus, although the variance estimates can be sensitive to the estimation technique, the resulting generalized least squares estimates of the mean function are much less sensitive.

(c) The regression output using White standard errors is

$$\widehat{WAGE}_i = -9.9140 + 1.2340 EDUC_i + 0.1332 EDUC_i + 1.5241 METRO_i$$
$$\text{(se)} \quad (1.2124) \quad (0.0835) \qquad (0.0158) \qquad (0.3445)$$

With the exception of that for *METRO*, these standard errors are larger than those in part (b), reflecting the lower precision of least squares estimation.

CHAPTER 9

Exercise Answers

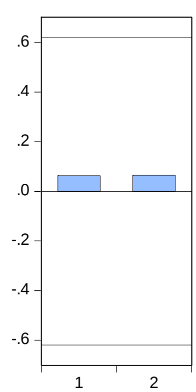
EXERCISE 9.4

(a) Using hand calculations

$$r_1 = \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2} = \frac{0.0979}{1.5436} = 0.0634, \qquad r_2 = \frac{\sum_{t=3}^{T} \hat{e}_t \hat{e}_{t-2}}{\sum_{t=1}^{T} \hat{e}_t^2} = \frac{0.1008}{1.5436} = 0.0653$$

(b) (i) For testing $H_0 : \rho_1 = 0$ against $H_1 : \rho_1 \neq 0$, $Z = \sqrt{T} r_1 = \sqrt{10} \times 0.0634 = 0.201$. Critical values are $Z_{(0.025)} = -1.96$ and $Z_{(0.975)} = 1.96$. We do not reject the null hypothesis and conclude that r_1 is not significantly different from zero.

(ii) For testing $H_0 : \rho_2 = 0$ against $H_1 : \rho_2 \neq 0$, $Z = \sqrt{T} r_2 = \sqrt{10} \times 0.0653 = 0.207$. Critical values are $Z_{(0.025)} = -1.96$ and $Z_{(0.975)} = 1.96$. We do not reject the null hypothesis and conclude that r_2 is not significantly different from zero.



The significance bounds are drawn at $\pm 1.96/\sqrt{10} = \pm 0.62$. With this small sample, the autocorrelations are a long way from being significantly different from zero.

EXERCISE 9.7

(a) Under the assumptions of the AR(1) model, $\text{corr}(e_t, e_{t-k}) = \rho^k$. Thus,

(i) $\text{corr}(e_t, e_{t-1}) = \rho = 0.9$

(ii) $\text{corr}(e_t, e_{t-4}) = \rho^4 = 0.9^4 = 0.6561$

(iii) $\sigma_e^2 = \dfrac{\sigma_v^2}{1-\rho^2} = \dfrac{1}{1-0.9^2} = 5.263$

(b) (i) $\text{corr}(e_t, e_{t-1}) = \rho = 0.4$

(ii) $\text{corr}(e_t, e_{t-4}) = \rho^4 = 0.4^4 = 0.0256$

(iii) $\sigma_e^2 = \dfrac{\sigma_v^2}{1-\rho^2} = \dfrac{1}{1-0.4^2} = 1.190$

When the correlation between the current and previous period error is weaker, the correlations between the current error and the errors at more distant lags die out relatively quickly, as is illustrated by a comparison of $\rho_4 = 0.6561$ in part (a)(ii) with $\rho_4 = 0.0256$ in part (b)(ii). Also, the larger the correlation ρ, the greater the variance σ_e^2, as is illustrated by a comparison of $\sigma_e^2 = 5.263$ in part (a)(iii) with $\sigma_e^2 = 1.190$ in part (b)(iii).

EXERCISE 9.10

(a) The forecasts are $\widehat{DURGWTH}_{2010Q1} = 0.7524$ and $\widehat{DURGWTH}_{2010Q2} = 0.6901$.

(b) The lag weights for up to 12 quarters are as follows.

Lag	Estimate
0	0.7422
1	0.2268
2	−0.0370
3	0.0060
4	-9.8×10^{-4}
5	1.6×10^{-4}
6	-2.6×10^{-5}
7	4.3×10^{-6}
8	-6.9×10^{-7}
9	1.1×10^{-7}
10	-1.9×10^{-8}
11	3.0×10^{-9}
12	-4.9×10^{-10}

Exercise 9.10 (continued)

(c) The one and two-quarter delay multipliers are

$$\hat{\beta}_1 = \frac{\partial DURGWTH_t}{\partial INGRWTH_{t-1}} = 0.2268 \qquad \hat{\beta}_2 = \frac{\partial DURGWTH_t}{\partial INGRWTH_{t-2}} = -0.0370$$

These values suggest that if income growth increases by 1% and then returns to its original level in the next quarter, then growth in the consumption of durables will increase by 0.227% in the next quarter and decrease by 0.037% two quarters later.

The one and two-quarter interim multipliers are

$$\hat{\beta}_0 + \hat{\beta}_1 = 0.7422 + 0.2268 = 0.969$$

$$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 = 0.969 - 0.0370 = 0.932$$

These values suggest that if income growth increases by 1% and is maintained at its new level, then growth in the consumption of durables will increase by 0.969% in the next quarter and increase by 0.932% two quarters later.

The total multiplier is $\sum_{j=0}^{\infty} \hat{\beta}_j = 0.9373$. This value suggests that if income growth increases by 1% and is maintained at its new level, then, at the new equilibrium, growth in the consumption of durables will increase by 0.937%.

EXERCISE 9.12

(a)

Coefficient Estimates and AIC and SC Values for Finite Distributed Lag Model

	$q=0$	$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$
$\hat{\alpha}$	0.4229	0.5472	0.5843	0.5828	0.6002	0.5990	0.5239
$\hat{\beta}_0$	–0.3119	–0.2135	–0.1974	–0.1972	–0.1940	–0.1940	–0.1830
$\hat{\beta}_1$		–0.1954	–0.1693	–0.1699	–0.1726	–0.1728	–0.1768
$\hat{\beta}_2$			–0.0707	–0.0713	–0.0664	–0.0662	–0.0828
$\hat{\beta}_3$				0.0021	0.0065	0.0062	0.0192
$\hat{\beta}_4$					–0.0222	–0.0225	–0.0475
$\hat{\beta}_5$						0.0015	–0.0169
$\hat{\beta}_6$							0.0944
AIC	–3.1132	–3.4314	–3.4587	–3.4370	–3.4188	–3.3971	–3.4416
AIC*	–0.2753	–0.5935	–0.6208	–0.5991	–0.5809	–0.5592	–0.6037
SC	–3.0584	–3.3492	–3.3490	–3.2999	–3.2543	–3.2052	–3.2223
SC*	–0.2205	–0.5113	–0.5111	–0.4620	–0.4165	–0.3673	–0.3844

Note: $\text{AIC}^* = \text{AIC} - 1 - \ln(2\pi)$ and $\text{SC}^* = \text{SC} - 1 - \ln(2\pi)$

The AIC is minimized at $q = 2$ while the SC is minimized at $q = 1$.

Exercise 9.12 (continued)

(b) (i) A 95% confidence interval for β_0 is given by

$$\hat{\beta}_0 \pm t_{(0.975,88)}\text{se}\left(\hat{\beta}_0\right) = -0.1974 \pm 1.987 \times 0.0328 = (-0.2626, -0.1322)$$

(ii) The null and alternative hypotheses are

$$H_0: \beta_0 + \beta_1 + \beta_2 = -0.5 \qquad H_1: \beta_0 + \beta_1 + \beta_2 > -0.5$$

The test statistic is

$$t = \frac{b_0 + b_1 + b_2 - (-0.5)}{\text{se}\left(b_0 + b_1 + b_2\right)} = \frac{0.062656}{0.034526} = 1.815$$

The critical value is $t_{(0.95,88)} = 1.662$. Since $t = 1.815 > 1.662$, we reject the null hypothesis and conclude that the total multiplier is greater than –0.5. The *p*-value is 0.0365.

(iii) The estimated normal growth rate is $\hat{G}_N = 0.58427/0.437344 = 1.336$. The 95% confidence interval for the normal growth rate is

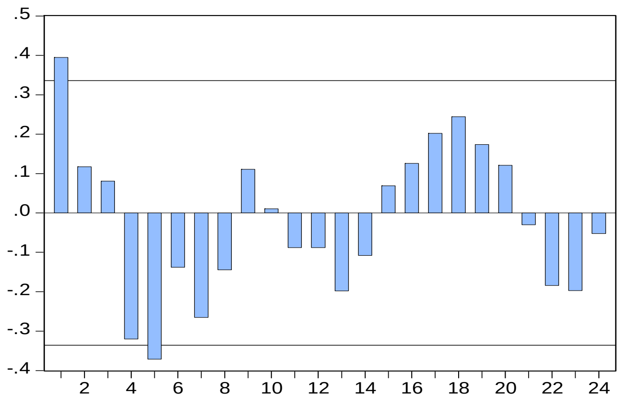
$$\hat{G}_N \pm t_{(0.975,88)}\text{se}\left(\hat{G}_N\right) = 1.336 \pm 1.987 \times 0.0417 = (1.253, 1.419)$$

EXERCISE 9.15

$$\widehat{\ln(AREA)} = 3.8933 + 0.7761\ln(PRICE)$$

(0.0613)	(0.2771)	least squares se's
(0.0624)	(0.3782)	HAC se's

(a) The correlogram for the residuals is



The significant bounds used are $\pm 1.96/\sqrt{34} = \pm 0.336$. Autocorrelations 1 and 5 are significantly different from zero.

Exercise 9.15 (continued)

(b) The null and alternative hypotheses are $H_0: \rho = 0$ and $H_0: \rho \neq 0$, and the test statistic is $LM = 5.4743$, yielding a *p*-value of 0.0193. Since the *p*-value is less than 0.05, we reject the null hypothesis and conclude that there is evidence of autocorrelation at the 5 percent significance level.

(c) The 95% confidence intervals are:

(i) Using least square standard errors

$$b_2 \pm t_{(0.975,32)} \times \text{se}(b_2) = 0.7761 \pm 2.0369 \times 0.2775 = (0.2109, 1.3413)$$

(ii) Using HAC standard errors

$$b_2 \pm t_{(0.975,32)} \times \text{se}(b_2) = 0.7761 \pm 2.0369 \times 0.3782 = (0.0057, 1.5465)$$

The wider interval under HAC standard errors shows that ignoring serially correlated errors gives an exaggerated impression about the precision of the least-squares estimated elasticity of supply.

(d) The estimated equation under the assumption of AR(1) errors is

$$\widehat{\ln(AREA_t)} = \underset{(0.0922)}{3.8988} + \underset{(0.2593)}{0.8884}\ln(PRICE_t) \qquad e_t = \underset{(0.1660)}{0.4221}e_{t-1} + v_t$$

(se)

The *t*-value for testing whether the estimate for ρ is significantly different from zero is $t = 0.4221/0.1660 = 2.542$, with a *p*-value of 0.0164. We conclude that $\hat{\rho}$ is significantly different from zero at a 5% level. A 95% confidence interval for the elasticity of supply is

$$b_2 \pm t_{(0.975,30)} \times \text{se}(b_2) = 0.8884 \pm 2.0423 \times 0.2593 = (0.3588, 1.4179)$$

This confidence interval is narrower than the one from HAC standard errors in part (c), reflecting the increased precision from recognizing the AR(1) error. It is also slightly narrower than the one from least squares, although we cannot infer much from this difference because the least squares standard errors are incorrect.

(e) We write the ARDL(1,1) model as

$$\ln(AREA_t) = \delta + \theta_1 \ln(AREA_{t-1}) + \delta_0 \ln(PRICE_t) + \delta_1 \ln(PRICE_{t-1}) + e_t$$

The estimated model is

$$\widehat{\ln(AREA_t)} = \underset{(0.6557)}{2.3662} + \underset{(0.1666)}{0.4043}\ln(AREA_{t-1}) + \underset{(0.2798)}{0.7766}\ln(PRICE_t) - \underset{(0.2966)}{0.6109}\ln(PRICE_{t-1})$$

For this ARDL(1,1) model to be equal to the AR(1) model in part (d), we need to impose the restriction $\delta_1 = -\theta_1\delta_0$. Thus, we test $H_0: \delta_1 = -\theta_1\delta_0$ against $H_1: \delta_1 \neq -\theta_1\delta_0$.

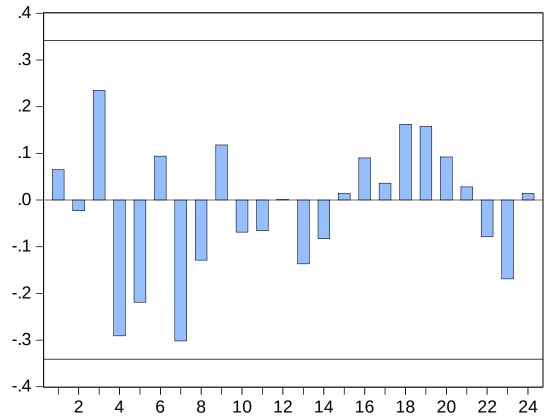
Exercise 9.15(e) (continued)

The test value is

$$t = \frac{\hat{\delta}_1 - (-\hat{\theta}_1\hat{\delta}_0)}{\text{se}\left(\hat{\delta}_1 + \hat{\theta}_1\hat{\delta}_0\right)} = \frac{-0.6109 - (-0.4043 \times 0.7766)}{0.2812} = -1.0559$$

with p-value of 0.300. Thus, we fail to reject the null hypothesis and conclude that the two models are equivalent.

The correlogram presented below suggests the errors are not serially correlated. The significance bounds used are $\pm 1.96/\sqrt{33} = 0.3412$. The *LM* test with a p-value of 0.423 confirms this decision.



EXERCISE 9.16

(a) The forecast values for $\ln(AREA_t)$ in years $T+1$ and $T+2$ are 4.04899 and 3.82981, respectively. The corresponding forecasts for *AREA* using the natural predictor are

$$\widehat{AREA}^n_{T+1} = \exp(4.04899) = 57.34$$

$$\widehat{AREA}^n_{T+2} = \exp(3.82981) = 46.05$$

Using the corrected predictor, they are

$$\widehat{AREA}^c_{T+1} = \widehat{AREA}^n_{T+1} \exp\left(\hat{\sigma}^2/2\right) = 57.3395 \times \exp\left(0.284899^2/2\right) = 59.71$$

$$\widehat{AREA}^c_{T+2} = \widehat{AREA}^n_{T+2} \exp\left(\hat{\sigma}^2/2\right) = 46.0539 \times \exp\left(0.284899^2/2\right) = 47.96$$

(b) The standard errors of the forecast errors for $\ln(AREA)$ are

$$\text{se}(u_1) = \hat{\sigma} = 0.28490$$

$$\text{se}(u_2) = \hat{\sigma}\sqrt{1+\hat{\theta}_1^2} = 0.28490\sqrt{1+0.40428^2} = 0.3073$$

Exercise 9.16(b) (continued)

The 95% interval forecasts for $\ln(AREA)$ are:

$$\widehat{\ln(AREA)}_{T+1} \pm t_{(0.975,29)} \times \text{se}(u_1) = 4.04899 \pm 2.0452 \times 0.28490 = (3.4663, 4.63167)$$

$$\widehat{\ln(AREA)}_{T+2} \pm t_{(0.975,29)} \times \text{se}(u_2) = 3.82981 \pm 2.0452 \times 0.3073 = (3.20132, 4.45830)$$

The corresponding intervals for *AREA* obtained by taking the exponential of these results are:

For $T+1$: $(e^{3.46630}, e^{4.63167}) = (32.02, 102.69)$

For $T+2$: $(e^{3.20132}, e^{4.45830}) = (24.56, 86.34)$

(c) The lag and interim elasticities are reported in the table below:

Lag	β_s	Lag Elasticities	Interim Elasticities
0	$\beta_0 = \delta_0$	0.7766	0.7766
1	$\beta_1 = \delta_1 + \theta_1\beta_0$	–0.2969	0.4797
2	$\beta_2 = \theta_1\beta_1$	–0.1200	0.3597
3	$\beta_3 = \theta_1\beta_2$	–0.0485	0.3112
4	$\beta_4 = \theta_1\beta_3$	–0.0196	0.2916

The lag elasticities show the percentage change in area sown in the current and future periods when price increases by 1% and then returns to its original level. The interim elasticities show the percentage change in area sown in the current and future periods when price increases by 1% and is maintained at the new level.

(d) The total elasticity is given by

$$\sum_{j=0}^{\infty} \beta_j = \frac{\hat{\delta}_0 + \hat{\delta}_1}{1 - \hat{\theta}_1} = \frac{0.77663 - 0.61086}{1 - 0.40428} = 0.2783$$

If price is increased by 1% and then maintained at its new level, then area sown will be 0.28% higher when the new equilibrium is reached.

CHAPTER 10

Exercise Answers

EXERCISE 10.5

(a) The least-squares estimated equation is

$$\widehat{SAVINGS} = 4.3428 - 0.0052 INCOME$$
$$\text{(se)} \quad (0.8561) \ (0.0112)$$

(b) The estimated equation using the instrumental variables estimator, with instrument z = *AVERAGE_INCOME* is

$$\widehat{SAVINGS} = 0.9883 + 0.0392 INCOME$$
$$\text{(se)} \quad (1.5240) \ (0.0200)$$

(c) To perform the Hausman test we estimate the artificial regression as

$$\widehat{SAVINGS} = 0.9883 + 0.3918 INCOME - 0.0755\hat{v}_t$$
$$\text{(se)} \quad (1.1720)(0.0154) \qquad (0.0201)$$

(d) The first stage estimation yields

$$\widehat{INCOME} = -35.0220 + 1.6417 AVERAGE\_INCOME$$

The second stage regression is

$$\widehat{SAVINGS} = 0.9883 + 0.0392 INCOME$$
$$\text{(se)} \quad (1.2530) \ (0.0165)$$

EXERCISE 10.7

(a) The least squares estimated equation is

$$\hat{Q} = 1.7623 + 0.1468 XPER + 0.4380 CAP + 0.2392 LAB$$
$$(\text{se})\ (1.0550)\quad (0.0634)\qquad (0.1176)\qquad (0.0998)$$

(b) (i) $\hat{Q}_0 = 9.0647$ and $\widehat{\text{var}(f)}$ = 7.756. The 95% interval prediction is $\hat{Q}_0 \pm t_c \text{se}(f) = 9.0647 \pm 1.9939 \times 2.785 = (3.51,\ 14.62)$

(ii) $\hat{Q}_0 = 10.533$ and $\text{se}(f) = 2.802$. A 95% interval prediction is $10.533 \pm 1.9939 \times 2.802 = (4.95, 16.12)$.

(iii) $\hat{Q}_0 = 12.001$ and $\text{se}(f) = 2.957$. The interval prediction is $12.001 \pm 1.9939 \times 2.957 = (6.11, 17.90)$.

(c) The estimated artificial regression is

$$\hat{Q} = -2.4867 + 0.5121 XPER + 0.3321 CAP + 0.2400 LAB - 0.4158 \hat{v}$$
$$(t)\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (-2.1978)$$

The *p*-value of the test is 0.031 so at a 5% level of significance we can conclude that there is correlation between *XPER* and the error term.

(d) The IV estimated equation is

$$\hat{Q} = -2.4867 + 0.5121 XPER + 0.3321 CAP + 0.2400 LAB$$
$$(\text{se})\quad (2.7230)\ (0.2205)\qquad (0.1545)\qquad (0.1209)$$
$$(t)\quad (-0.91)\quad (2.32)\qquad (2.15)\qquad (1.99)$$

(e) (i) $\hat{Q}_0 = 7.6475$ and $\text{se}(f) = 3.468$.

The interval prediction is $7.6475 \pm 1.9939 \times 3.468 = (0.73, 14.56)$

(ii) $\hat{Q}_0 = 12.768$ and $\text{se}(f) = 3.621$.
The interval prediction is $12.768 \pm 1.9939 \times 3.621 = (5.55, 19.99)$.

(iii) $\hat{Q}_0 = 17.890$ and $\text{se}(f) = 4.891$.
The interval prediction is $17.89 \pm 1.9939 \times 4.891 = (8.14,\ 27.64)$

CHAPTER 11

Exercise Answers

EXERCISE 11.7

(a) Rearranging the demand equation, $Q = \alpha_1 + \alpha_2 P + \alpha_3 PS + \alpha_4 DI + e^d$, yields

$$P = \frac{1}{\alpha_2}\left(Q - \alpha_1 + \alpha_3 PS + \alpha_4 DI + e^d\right)$$
$$= \delta_1 + \delta_2 Q + \delta_3 PS + \delta_4 DI + u^d$$

We expect $\delta_2 < 0$, $\delta_3 > 0$, $\delta_4 > 0$.

Rearranging the supply equation, $Q = \beta_1 + \beta_2 P + \beta_3 PF + e^s$, yields

$$P = \frac{1}{\beta_2}\left(Q - \beta_1 + \beta_3 PF + e^s\right)$$
$$= \phi_1 + \phi_2 Q + \phi_3 PF + u^s$$

We expect $\phi_2 > 0$, $\phi_3 > 0$.

(b) The estimated demand equation is

$$\hat{P} = -11.4284 - 2.6705Q + 3.4611PS + 13.3899DI$$
$$(\text{se}) \quad (13.5916) \quad (1.1750) \quad (1.1156) \quad (2.7467)$$

The estimated supply equation is

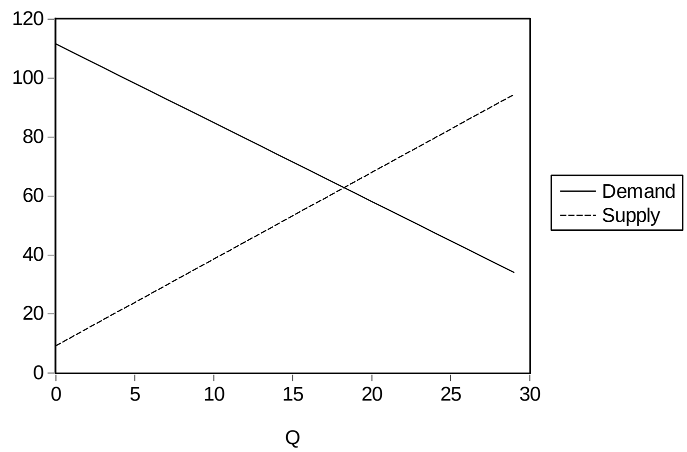
$$\hat{P} = -58.7982 + 2.9367Q + 2.9585PF$$
$$(\text{se}) \quad (5.8592) \quad (0.2158) \quad (0.1560)$$

Exercise 11.7 (continued)

(c) The estimated price elasticity of demand at the mean is $\hat{\varepsilon}_D = -1.2725$

(d) The figure below is a sketch of the supply and demand equations using the estimates from part (b) and the given exogenous variable values. The lines are given by linear equations:

Demand: $\hat{P} = 111.5801 - 2.6705Q$; Supply: $\hat{P} = 9.2470 + 2.9367Q$



(e) The estimated equilibrium values from part (d) are

$$Q_{EQM} = 18.2503 \qquad P_{EQM} = 62.8427$$

Using the reduced form estimates in Tables 11.2a and 11.2b, the predicted equilibrium values are

$$Q_{EQM\_RF} = 18.2604 \qquad P_{EQM\_RF} = 62.8154.$$

(f) The estimated least-squares estimated demand equation is

$$\hat{P} = -13.6195 + 0.1512Q + 1.3607PS + 12.3582DI$$
$$(\text{se}) \quad (9.0872) \quad (0.4988) \quad (0.5940) \quad (1.8254)$$

The sign for the coefficient of Q is incorrect because it suggests that there is a positive relationship between price and quantity demanded.

The estimated supply equation is

$$\hat{P} = -52.8763 + 2.6613Q + 2.9217PF$$
$$(\text{se}) \quad (5.0238) \quad (0.1712) \quad (0.1482)$$

All estimates in this supply equation are significantly different from zero. All coefficient signs are correct, and the coefficient values do not differ much from the estimates in part (b).

EXERCISE 11.8

(a) The summary statistics are presented in the following table

	Mean		Standard Deviation	
Variable	*LFP* = 1	*LFP* = 0	*LFP* = 1	*LFP* = 0
AGE	41.9720	43.2831	7.7211	8.4678
KIDSL6	0.1402	0.3662	0.3919	0.6369
FAMINC	24130	21698	11671	12728

On average, women who work are younger, have fewer children under the age of 6 and have a higher family income. Also, the standard deviation across all variables is smaller for working women.

(b) $\beta_2 > 0$: A higher wage leads to an increased quantity of labor supplied.

β_3: The effect of an increase in education is unclear.

β_4: This sample has been taken for working women between the ages of 30 and 60. It is not certain whether hours worked increases or decreases over this age group.

$\beta_5 < 0, \quad \beta_6 < 0$: The presence of children in the household reduces the number of hours worked because they demand time from their mother.

$\beta_7 < 0$: As income from other sources increases, it becomes less necessary for the woman to work.

NWIFEINC measures the sum of all family income excluding the wife's income.

(c) The least squares estimated equation is

$$\widehat{HOURS} = 2115 - 17.41\ln(WAGE) - 14.44EDUC - 7.730AGE$$
$$(\text{se}) \quad (340.1)(54.22) \qquad (17.97) \qquad (5.530)$$
$$-342.5KIDSL6 - 115.0KIDS618 - 0.00425NWIFEINC$$
$$(100.0) \qquad (30.83) \qquad (0.00366)$$

The negative coefficient for ln(*WAGE*) is unexpected; we expected this coefficient to be positive.

(d) An additional year of education increases wage by $0.0999 \times 100\%$ = 9.99%.

(e) The presence of *EXPER* and $EXPER^2$ in the reduced form equation and their absence in the supply equation serves to identify the supply equation. The *F*-test of their joint significance yields an *F* value of 8.25, which gives a *p*-value of 0.0003. The *F* value is less than the rule of thumb value for strong instrumental variables of 10.

Exercise 11.8 (continued)

(f) The two-stage least squares estimated equation is

$$\widehat{HOURS} = 2432 + 1545\ln(WAGE) - 177EDUC - 10.78AGE$$
$$\begin{matrix}\text{(se)} & (594.2) & (480.7) & (58.1) & (9.577)\end{matrix}$$

$$-211KIDSL6 - 47.56KIDS618 - 0.00925NWIFEINC$$
$$\begin{matrix}(177) & (56.92) & (0.00648)\end{matrix}$$

The statistically significant coefficients are the coefficients of ln(*WAGE*) and *EDUC*. The sign of ln(*WAGE*) is now in line with our expectations. The other coefficients have signs that are not contrary to our expectations.

CHAPTER 15

Exercise Answers

EXERCISE 15.5

(a) The three estimates for β_2 are:

(i) Dummy variable / fixed effects estimator $b_2 = 0.0207$ $\text{se}(b_2) = 0.0209$

(ii) Estimator from averaged data $\hat{\beta}_2^A = 0.0273$ $\text{se}(\hat{\beta}_2^A) = 0.0075$

(iii)..Random effects estimator $\hat{\beta}_2 = 0.0266$ $\text{se}(\hat{\beta}_2) = 0.0070$

The estimates from the averaged data and from the random effects model are very similar, with the standard error from the random effects model suggesting the estimate from this model is more precise. The dummy variable model estimate is noticeably different and its standard error is much bigger than that of the other two estimates.

(b) To test $H_0 : \beta_{1,1} = \beta_{1,2} = \cdots = \beta_{1,40}$ against the alternative that not all of the intercepts are equal, we use the usual *F*-test for testing a set of linear restrictions. The calculated value is $F = 3.175$, while the 5% critical value is $F_{(0.95,39,79)} = 1.551$. Thus, we reject H_0 and conclude that the household intercepts are not all equal. The *F* value can be obtained using the equation

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(NT-K)} = \frac{(195.5481 - 76.15873)/39}{76.15873/(120-41)} = 3.175$$

EXERCISE 15.14

(a),(b) Least squares and SUR estimates and standard errors for the demand system appear in the following table

	Estimates		Standard Errors	
Coefficient	LS	SUR	LS	SUR
Constant	1.017	2.501	1.354	1.092
Price-1	–0.567	–0.911	0.215	0.130
Income	1.434	1.453	0.229	0.217
Constant	2.463	3.530	1.453	1.232
Price-2	–0.648	–0.867	0.188	0.125
Income	1.144	1.136	0.261	0.248
Constant	4.870	5.021	0.546	0.468
Price-3	–0.964	–0.999	0.065	0.034
Income	0.871	0.870	0.108	0.103

All price elasticities are negative and all income elasticities are positive, agreeing with our *a priori* expectations.

For testing the null hypothesis that the errors are uncorrelated against the alternative that they are correlated, we obtain a value for the $\chi^2_{(3)}$ test statistic

$$LM = T(r_{12}^2 + r_{13}^2 + r_{23}^2) = 30 \times (0.0144 + 0.3708 + 0.2405) = 18.77$$

where

$$\hat{\sigma}_{12} = \frac{1}{30-3}\sum_{t=1}^{30} \hat{e}_{1,t}\hat{e}_{2,t} = -0.0213 \quad \Rightarrow \quad r_{12}^2 = \frac{(-0.0213)^2}{(0.3943)^2(0.4506)^2} = 0.0144$$

$$\hat{\sigma}_{13} = \frac{1}{30-3}\sum_{t=1}^{30} \hat{e}_{1,t}\hat{e}_{3,t} = -0.0448 \quad \Rightarrow \quad r_{13}^2 = \frac{(-0.0448)^2}{(0.3943)^2(0.1867)^2} = 0.3708$$

$$\hat{\sigma}_{23} = \frac{1}{30-3}\sum_{t=1}^{30} \hat{e}_{2,t}\hat{e}_{3,t} = -0.0413 \quad \Rightarrow \quad r_{23}^2 = \frac{(-0.0413)^2}{(0.4506)^2(0.1867)^2} = 0.2405$$

The 5% critical value for a χ^2 test with 3 degrees of freedom is $\chi^2_{(0.95,3)} = 7.81$. Thus, we reject the null hypothesis and conclude that the errors are contemporaneously correlated.

(c) We wish to test $H_0: \beta_{13} = 1, \beta_{23} = 1, \beta_{33} = 1$ against the alternative that at least one income elasticity is not unity. This test can be performed using an F-test or a χ^2-test. Both are large-sample approximate tests. The test values are $F = 1.895$ with a p-value of 0.14 or $\chi^2 = 5.686$ with a p-value of 0.13. Thus, we do not reject the hypothesis that all income elasticities are equal to 1.

CHAPTER 16

Exercise Answers

EXERCISE 16.2

(a) The maximum likelihood estimates of the logit model are

$$\tilde{\beta}_1 + \tilde{\beta}_2 DTIME = -0.2376 + 0.5311 DTIME$$
$$\text{(se)} \qquad (0.7505) \quad (0.2064)$$

These estimates are quite different from the probit estimates on page 593. The logit estimate $\tilde{\beta}_1$ is much smaller than the probit estimate, whereas $\tilde{\beta}_2$ and the standard errors are larger compared to the probit model. The differences are primarily a consequence of the variance of the logistic distribution $\left(\pi^2/3\right)$ being different to that of the standard normal (1).

(b) $\dfrac{dp}{dx} = \dfrac{d\Lambda(l)}{dl} \cdot \dfrac{dl}{dx} = \lambda(\beta_1 + \beta_2 x)\beta_2$, where $l = \beta_1 + \beta_2 x$

Given that *DTIME* = 2, the marginal effect of an increase in *DTIME* using the logit estimates is

$$\widehat{\frac{dp}{dDTIME}} = 0.1125$$

(c) Using the logit estimates, the probability of a person choosing automobile transportation given that *DTIME* = 3 is 0.7951

Exercise 16.2 (continued)

(d) The predicted probabilities (*PHAT*) are

	dtime	auto	phat
1.	-4.85	0	.0566042
2.	2.44	0	.7423664
3.	8.28	1	.9846311
4.	-2.46	0	.1759433
5.	-3.16	0	.1283255
6.	9.1	1	.9900029
7.	5.21	1	.9261805
8.	-8.77	0	.0074261
9.	-1.7	0	.2422391
10.	-5.15	0	.0486731
11.	-9.07	0	.0063392
12.	6.55	1	.9623526
13.	-4.4	1	.0708038
14.	-.7	0	.3522088
15.	5.16	1	.9243443
16.	3.24	1	.8150529
17.	-6.18	0	.0287551
18.	3.4	1	.827521
19.	2.79	1	.7762923
20.	-7.29	0	.0161543
21.	4.99	1	.9177834

Using the logit model, 90.48% of the predictions are correct.

EXERCISE 16.3

(a) The least squares estimated model is

$$\hat{p} = -0.0708 + 0.160 FIXRATE - 0.132 MARGIN - 0.793 YIELD$$
$$\text{(se)} \; (1.288) \quad (0.0822) \qquad (0.0498) \qquad (0.323)$$
$$-0.0341 MATURITY - 0.0887 POINTS + 0.0289 NETWORTH$$
$$(0.191) \qquad (0.0711) \qquad (0.0118)$$

All the signs of the estimates are consistent with expectations. The predicted values are between zero and one except those for observations 29 and 48 which are negative.

Exercise 16.3 (continued)

(b) The estimated probit model is

$$\hat{p} = \Phi(\underset{(4.121)}{-1.877} + \underset{(0.262)}{0.499}FIXRATE - \underset{(0.174)}{0.431}MARGIN - \underset{(1.083)}{2.384}YIELD$$
(se)

$$\underset{(0.623)}{-0.0591}MATURITY - \underset{(0.241)}{0.300}POINTS + \underset{(0.0379)}{0.0838}NETWORTH)$$

All the estimates have the expected signs. Ignoring the intercept and using a 5% level of significance and one-tail tests, we find that all coefficients are statistically significant with the exception of those for *MATURITY* and *POINTS*.

(c) The percentage of correct predictions using the probit model to estimate the probabilities of choosing an adjustable rate mortgage is 75.64%.

(d) The marginal effect of an increase in *MARGIN* at the sample means is

$$\widehat{\frac{dp}{dMARGIN}} = -0.164$$

This estimate suggests that, at the sample means, a one percent increase in the difference between the variable rate and the fixed rate decreases the probability of choosing the variable-rate mortgage by 16.4 percent.

APPENDIX A

Exercise Answers

EXERCISE A.1

(a) The slope is the change in the quantity supplied per unit change in market price. The slope here is 1.5, which represents a 1.5 unit increase in the quantity supplied of a good due to a one unit increase in market price.

(b) Elasticity = 1.25. The elasticity shows the percentage change in Q^s associated with a 1 percent change in *P*. At the point $P = 10$ and $Q^s = 12$, a 1 percent change in *P* is associated with a 1.25 percent change in Q^s.

When $P = 50$, Elasticity = 1.042. At the point $P = 50$ and $Q^s = 72$, a 1 percent change in *P* is associated with a 1.04 percent change in Q^s.

EXERCISE A.3

(a) $x^{2/3}$

(b) $\frac{1}{x^{5/24}}$

(c) $\frac{1}{x^2 y^{3/2}}$

EXERCISE A.5

(a) The graph of the relationship between average wheat production (*WHEAT*) and time (*t*) is shown below. For example, when t = 49, $WHEAT = 0.5 + 0.20\ln(t) = 1.2784$.

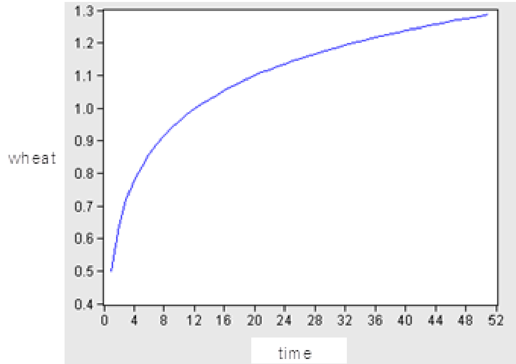


Figure xr-a.9(a) Graph of $WHEAT = 0.5 + 0.20\ln(t)$

The slope and elasticity for $t = 49$ are

$$\text{Slope} = 0.0041 \text{ when } t = 49$$

$$\text{Elasticity } = 0.1564 \text{ when } t = 49$$

(b) The graph of the relationship between average wheat production (*WHEAT*) and time (*t*) is shown below. For example, when t = 49, $WHEAT = 0.8 + 0.0004t^2 = 1.7604$.

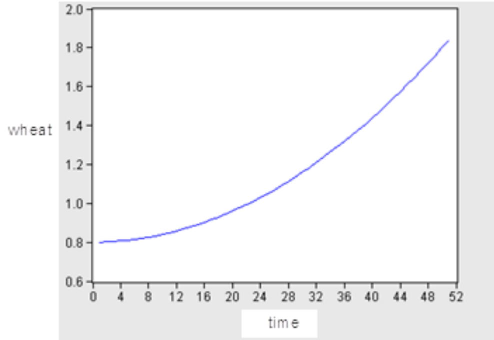


Figure xr-a.9(b) Graph of $WHEAT = 0.8 + 0.0004t^2$

The slope and elasticity for $t = 49$ are

$$\text{Slope} = 0.0392 \text{ when } t = 49$$

$$\text{Elasticity } = 1.0911 \text{ when } t = 49$$

EXERCISE A.7

(a) $x = 4.573239 \times 10^6$

$y = 5.975711 \times 10^4$

(b) $xy = 2.7328354597929 \times 10^{11}$

(c) $x / y = 7.6530458 \times 10^1$

(d) $x + y = 4.63299611 \times 10^6$

APPENDIX B

Exercise Answers

EXERCISE B.1

(a)

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right] = \frac{1}{n}\big(E(X_1) + E(X_2) + \cdots + E(X_n)\big)$$

$$= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \frac{n\mu}{n} = \mu$$

(b)

$$\operatorname{var}(\bar{X}) = \operatorname{var}\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right)$$

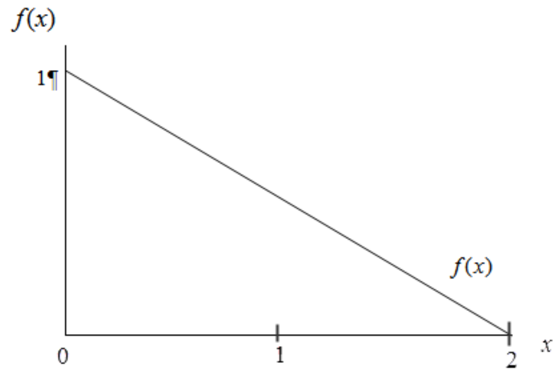
$$= \frac{1}{n^2}\big(\operatorname{var}(X_1) + \operatorname{var}(X_2) + \cdots + \operatorname{var}(X_n)\big)$$

$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Since $X_1, X_2, ..., X_n$ are independent random variables, their covariances are zero. This result was used in the second line of the equation which would contain terms like $\operatorname{cov}(X_i, X_j)$ if these terms were not zero.

EXERCISE B.3

(a) The probability density function is shown below.



(b) The total area is 1

(c) $P(X \geq 1) = \frac{1}{4}$.

(d) $P\left(X \leq \frac{1}{2}\right) = \frac{7}{16}$

(e) $P\left(X = 1\frac{1}{2}\right) = 0$.

(f) $E(X) = \frac{2}{3}$ and $\text{var}(X) = \frac{2}{9}$

(g) $F(x) = x\left(-\frac{x}{4} + 1\right)$

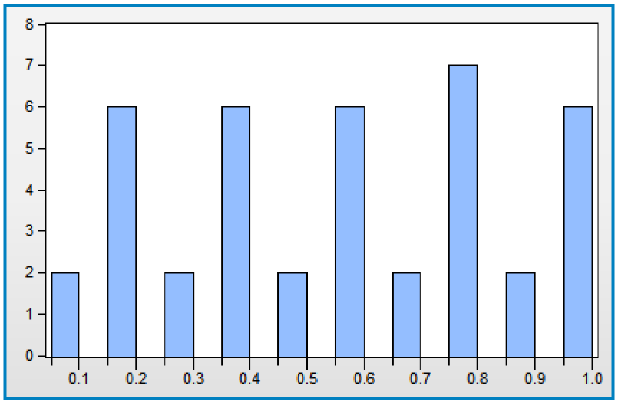
EXERCISE B.5

After setting up a workfile for 41 observations, the following EViews program can be used to generate the random numbers

```
series x
x(1)=79
scalar m=100
scalar a=263
scalar cee=71
for !i= 2 to 41
scalar q=a*x(!i-1)+cee
x(!i)=q-m*@ceiling(q/m)+m
next
series u=x/m
```

Exercise B.5 (continued)

If the random number generator has worked well, the observations in U should be independent draws of a uniform random variable on the (0,1) interval. A histogram of these numbers follows:



These numbers are far from random. There are no observations in the intervals (0.10,0.15), (0.20,0.25), (0.30,0.35), …. Moreover, the frequency of observations in the intervals (0.05,0.10), (0.25,0.30), (0.45,0.50), … is much less than it is in the intervals (0.15,0.20), (0.35,0.40), (0.55,0.60), …

The random number generator is clearly not a good one.

EXERCISE B.7

Let $E_{X,Y}$ be an expectation taken with respect to the joint density for (X,Y); E_X and E_Y are expectations taken with respect to the marginal distributions of X and Y, and $E_{Y|X}$ is an expectation taken with respect to the conditional distribution of Y given X.

Now $\text{cov}\left(Y,g(X)\right)=0$ if $E_{X,Y}\left(Y\times g(X)\right)=E_{X,Y}(Y)\times E_{X,Y}\left(g(X)\right)$. Using iterated expectations, we can write

$$
\begin{aligned}
E_{X,Y}\left(Y\times g(X)\right) &= E_X\left[E_{Y|X}\left(Y\times g(X)\right)\right] \\
&= E_X\left[g(X)E_{Y|X}\left(Y\right)\right] \\
&= E_X\left[g(X)\right]\times E_Y\left(Y\right) \\
&= E_{X,Y}\left[g(X)\right]\times E_{X,Y}\left(Y\right)
\end{aligned}
$$

EXERCISE B.9

(a) $P\left(0<X<\tfrac{1}{2}\right)=\dfrac{1}{64}$

(b) $P(1<X<2)=\dfrac{7}{8}$

EXERCISE B.12

(a) For $f(x,y)$ to be a valid *pdf*, we require $f(x,y) \geq 0$ and $\int_0^1 \int_0^1 f(x,y)\,dx\,dy = 1$. It is clear that $f(x,y) = 6x^2y \geq 0$ for all $0 \leq x \leq 1, 0 \leq y \leq 1$. To establish the second condition, we consider

$$\int_0^1 \int_0^1 6x^2 y\,dx\,dy = \int_0^1 y \int_0^1 6x^2 dx\,dy = \int_0^1 y\left[\left(2x^3\right)\Big|_0^1\right]dy = 2\int_0^1 y\,dy = 2 \times \left[\frac{y^2}{2}\Bigg|_0^1\right] = 1$$

(b) The marginal *pdf* for X is $f(x) = 3x^2$

The mean of X is $E(X) = \dfrac{3}{4}$

The variance of X is $\text{var}(X) = \dfrac{3}{80}$

(c) The marginal *pdf* for Y is $f(y) = 2y$

(d) The conditional *pdf* $f(x|y)$ is $f(x|y) = 3x^2$

and thus,

$$f\left(x \middle| Y = \tfrac{1}{2}\right) = 3x^2$$

(e) Since $f(x|y) = f(x)$, the conditional mean and variance of X given $Y = \frac{1}{2}$ are identical to the mean and variance of X found in part (b).

(f) Yes, X and Y are independent because $f(x,y) = 6x^2y = f(x)f(y) = 3x^2 \times 2y$.

APPENDIX C

Exercise Answers

EXERCISE C.3

The probability that in a 9 hour day, more than 20,000 pieces will be sold is 0.091.

EXERCISE C.5

(a) We set up the hypotheses $H_0: \mu \le 170$ versus $H_1: \mu > 170$. The alternative is $H_1: \mu > 170$ because we want to establish whether the mean monthly account balance is more than 170.

The test statistic, given H_0 is true, is:

$$t = \frac{\bar{X} - 170}{\hat{\sigma}/\sqrt{N}} \sim t_{(399)}$$

The rejection region is $t \ge 1.649$. The value of the test statistic is

$$t = \frac{178 - 170}{65/\sqrt{400}} = 2.462$$

Since $t = 2.462 > 1.649$, we reject H_0 and conclude that the new accounting system is cost effective.

(b) $$p = P\left[t_{(399)} \ge 2.462\right] = 1 - P\left[t_{(399)} < 2.462\right] = 0.007$$

EXERCISE C.8

A sample size of 424 employees is needed.