Chapter 8
Nonlinear Regression Functions

■ Solutions to Exercises

1. (a) The percentage increase in sales is $100 \times \frac{198-196}{196} = 1.0204\%$. The approximation is $100 \times [\ln(198) - \ln(196)] = 1.0152\%$.

   (b) When $Sales_{2002} = 205$, the percentage increase is $100 \times \frac{205-196}{196} = 4.5918\%$ and the approximation is $100 \times [\ln(205) - \ln(196)] = 4.4895\%$. When $Sales_{2002} = 250$, the percentage increase is $100 \times \frac{250-196}{196} = 27.551\%$ and the approximation is $100 \times [\ln(250) - \ln(196)] = 24.335\%$. When $Sales_{2002} = 500$, the percentage increase is $100 \times \frac{500-196}{196} = 155.1\%$ and the approximation is $100 \times [\ln(500) - \ln(196)] = 93.649\%$.

   (c) The approximation works well when the change is small. The quality of the approximation deteriorates as the percentage change increases.

2. (a) According to the regression results in column (1), the house price is expected to increase by 21% (= $100\% \times 0.00042 \times 500$) with an additional 500 square feet and other factors held constant. The 95% confidence interval for the percentage change is $100\% \times 500 \times (0.00042 \pm 1.96 \times 0.000038) = [17.276\% \text{ to } 24.724\%]$.

   (b) Because the regressions in columns (1) and (2) have the same dependent variable, $\bar{R}^2$ can be used to compare the fit of these two regressions. The log-log regression in column (2) has the higher $\bar{R}^2$, so it is better so use ln(*Size*) to explain house prices.

   (c) The house price is expected to increase by 7.1% (= $100\% \times 0.071 \times 1$). The 95% confidence interval for this effect is $100\% \times (0.071 \pm 1.96 \times 0.034) = [0.436\% \text{ to } 13.764\%]$.

   (d) The house price is expected to increase by 0.36% ($100\% \times 0.0036 \times 1 = 0.36\%$) with an additional bedroom while other factors are held constant. The effect is not statistically significant at a 5% significance level: $|t| = \frac{0.0036}{0.037} = 0.09730 < 1.96$. Note that this coefficient measures the effect of an additional bedroom holding the size of the house constant.

   (e) The quadratic term $\ln(Size)^2$ is not important. The coefficient estimate is not statistically significant at a 5% significance level: $|t| = \frac{0.0078}{0.14} = 0.05571 < 1.96$.

   (f) The house price is expected to increase by 7.1% (= $100\% \times 0.071 \times 1$) when a swimming pool is added to a house without a view and other factors are held constant. The house price is expected to increase by 7.32% (= $100\% \times (0.071 \times 1 + 0.0022 \times 1)$) when a swimming pool is added to a house with a view and other factors are held constant. The difference in the expected percentage change in price is 0.22%. The difference is not statistically significant at a 5% significance level: $|t| = \frac{0.0022}{0.10} = 0.022 < 1.96$.

3 (a) The regression functions for hypothetical values of the regression coefficients that are consistent with the educator's statement are: $\beta_1 > 0$ and $\beta_2 < 0$. When *TestScore* is plotted against *STR* the regression will show three horizontal segments. The first segment will be for values of $STR < 20$; the next segment for $20 \le STR \le 25$; the final segment for $STR > 25$. The first segment will be higher than the second, and the second segment will be higher than the third.

(b) It happens because of perfect multicollinearity. With all three class size binary variables included in the regression, it is impossible to compute the OLS estimates because the intercept is a perfect linear function of the three class size regressors.

4. (a) With 2 years of experience, the man's expected *AHE* is

$$\widehat{\ln(AHE)} = (0.0899\times 16) - (0.521\times 0) + (0.0207\times 0\times 16) + (0.232\times 2) - 0.000368\times 2^2)$$
$$-(0.058\times 0) - (0.078\times 0) - (0.030\times 1) + 1.215 = 2.578$$

With 3 years of experience, the man's expected *AHE* is

$$\widehat{\ln(AHE)} = (0.0899\times 16) - (0.521\times 0) + (0.0207\times 0\times 16) + (0.232\times 3) - (0.000368\times 3^2)$$
$$-(0.058\times 0) - (0.078\times 0) - (0.030\times 1) + 1.215 = 2.600$$

Difference = 2.600 − 2.578 = 0.022 (or 2.2%)

(b) With 10 years of experience, the man's expected *AHE* is

$$\widehat{\ln(AHE)} = (0.0899\times 16) - (0.521\times 0) + (0.0207\times 0\times 16) + (0.232\times 10) - (0.000368\times 10^2)$$
$$-(0.058\times 0) - (0.078\times 0) - (0.030\times 1) + 1.215 = 2.729$$

With 11 years of experience, the man's expected *AHE* is

$$\widehat{\ln(AHE)} = (0.0899\times 16) - (0.521\times 0) + (0.0207\times 0\times 16) + (0.232\times 11) - (0.000368\times 11^2)$$
$$-(0.058\times 0) - (0.078\times 0) - (0.030\times 1) + 1.215 = 2.744$$

Difference = 2.744 − 2.729 = 0.015 (or 1.5%)

(c) The regression in nonlinear in experience (it includes *Potential experience*^{2}).

(d) Yes, the coefficient on *Potential experience*^{2} is significant at the 1% level.

(e) No. This would affect the level of ln(*AHE*), but not the change associated with another year of experience.

(f) Include interaction terms *Female*× *Potential experience* and *Female* × (*Potential experience*)^2.

5. (a) (1) The demand for older journals is less elastic than for younger journals because the interaction term between the log of journal age and price per citation is positive. (2) There is a linear relationship between log price and log of quantity follows because the estimated coefficients on log price squared and log price cubed are both insignificant. (3) The demand is greater for journals with more characters follows from the positive and statistically significant coefficient estimate on the log of characters.

(b) (i) The effect of ln(*Price per citation*) is given by [–0.899 + 0.141 × ln(*Age*)] × ln(*Price per citation*). Using *Age* = 80, the elasticity is [–0.899 + 0.141 × ln(80)] = –0.28.

(ii) As described in equation (8.8) and the footnote on page 263, the standard error can be found by dividing 0.28, the absolute value of the estimate, by the square root of the *F*-statistic testing $\beta_{ln(Price\ per\ citation)} + \ln(80) \times \beta_{ln(Age)\times ln(Price\ per\ citation)} = 0$.

(c) $\ln\left(\frac{Characters}{a}\right) = \ln(Characters) - \ln(a)$ for any constant *a*. Thus, estimated parameter on *Characters* will not change and the constant (intercept) will change.

6. (a) (i) There are several ways to do this. Here is one. Create an indicator variable, say *DV*1, that equals one if *%Eligible* is greater than 20% and less than 50%. Create another indicator, say *DV*2, that equals one if *%Eligible* is greater than 50%. Run the regression:

$$TestScore = \beta_0 + \beta_1 \%Eligible + \beta_2 DV1 \times \%Eligible + \beta_3 DV2 \times \%Eligible + \text{other regressors}$$

The coefficient β_1 shows the marginal effect of *%Eligible* on *TestScores* for values of *%Eligible* > 20%, $\beta_1 + \beta_2$ shows the marginal effect for values of *%Eligible* between 20% and 50% and $\beta_1 + \beta_3$ shows the marginal effect for values of *%Eligible* greater than 50%.

(ii) The linear model implies that $\beta_2 = \beta_3 = 0$, which can be tested using an *F*-test.

(b) (i) There are several ways to do this, perhaps the easiest is to include an interaction term *STR* × ln(*Income*) to the regression in column (7).

(ii) Estimate the regression in part (b.i) and test the null hypothesis that the coefficient on the interaction term is equal to zero.

7. (a) (i) ln(*Earnings*) for females are, on average, 0.44 lower for men than for women.

(ii) The error term has a standard deviation of 2.65 (measured in log-points).

(iii) Yes. But the regression does not control for many factors (size of firm, industry, profitability, experience and so forth).

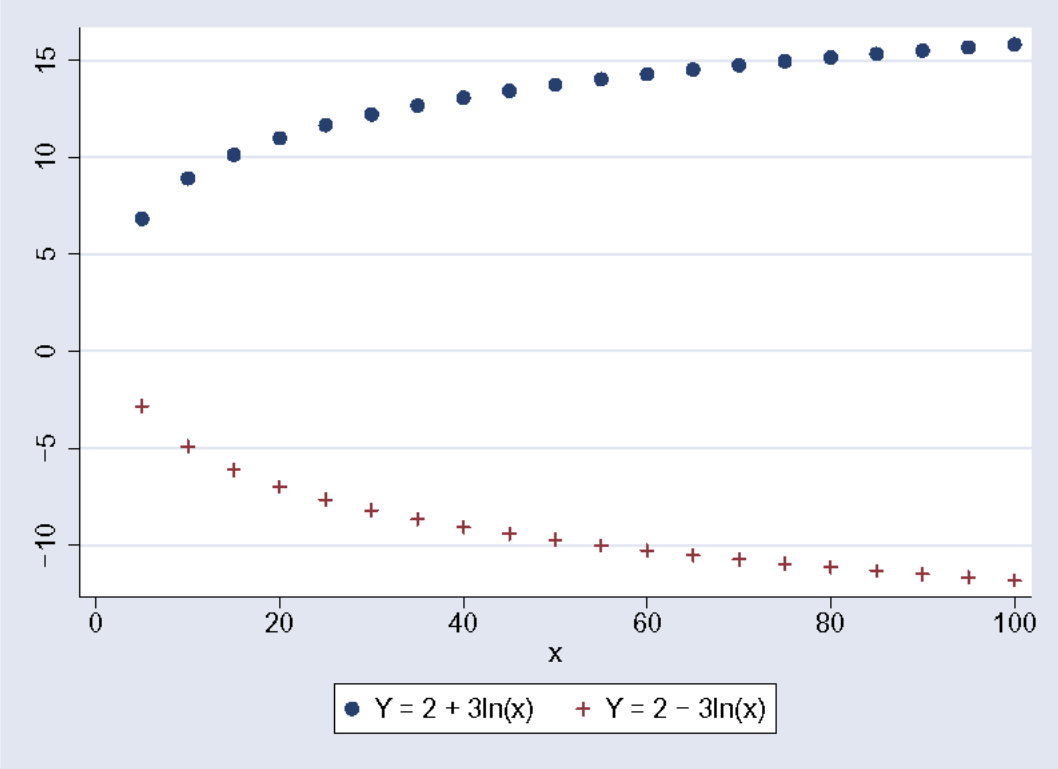
(iv) No. In isolation, these results do imply gender discrimination. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. Thus, it is also important to control for characteristics of the workers that may affect their productivity (education, years of experience, etc.) If these characteristics are systematically different between men and women, then they may be responsible for the difference in mean wages. (If this were true, it would raise an interesting and important question of why women tend to have less education or less experience than men, but that is a question about something other than gender discrimination.) These are potentially important omitted variables in the regression that will lead to bias in the OLS coefficient estimator for *Female*. Since these characteristics were not controlled for in the statistical analysis, it is premature to reach a conclusion about gender discrimination.

(b) (i) If *MarketValue* increases by 1%, earnings increase by 0.37%

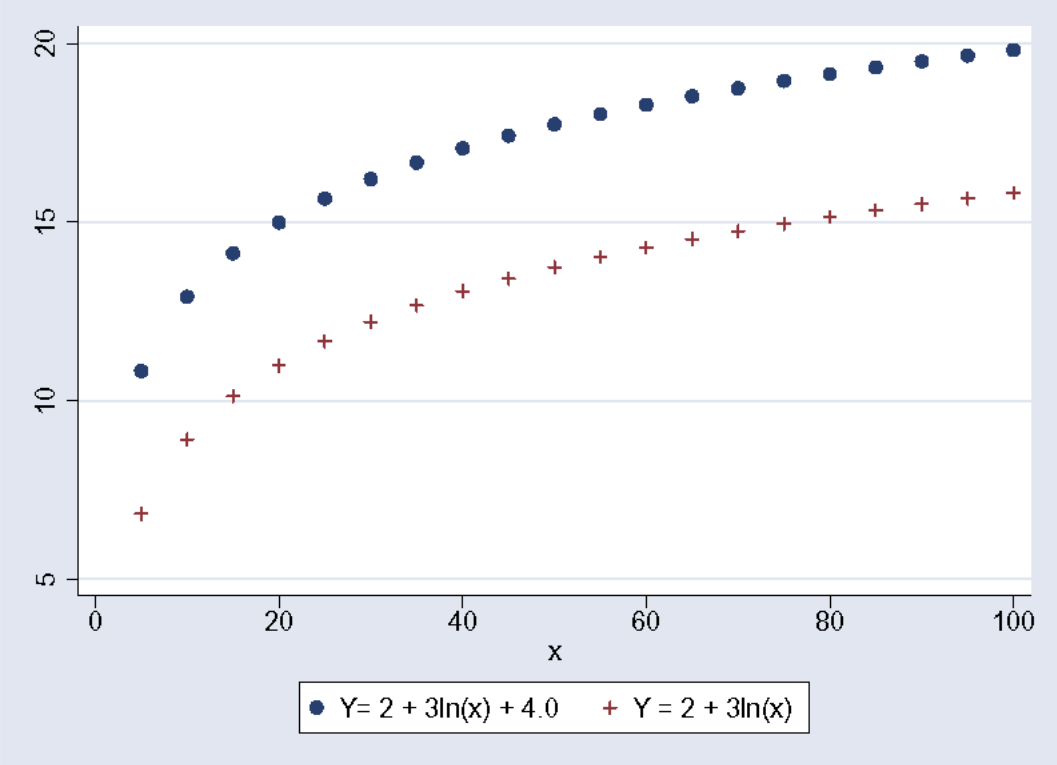
(ii) *Female* is correlated with the two new included variables and at least one of the variables is important for explaining ln(*Earnings*). Thus the regression in part (a) suffered from omitted variable bias.

(c) Forgetting about the effect or *Return*, whose effects seems small and statistically insignificant, the omitted variable bias formula (see equation (6.1)) suggests that *Female* is negatively correlated with ln(*MarketValue*).

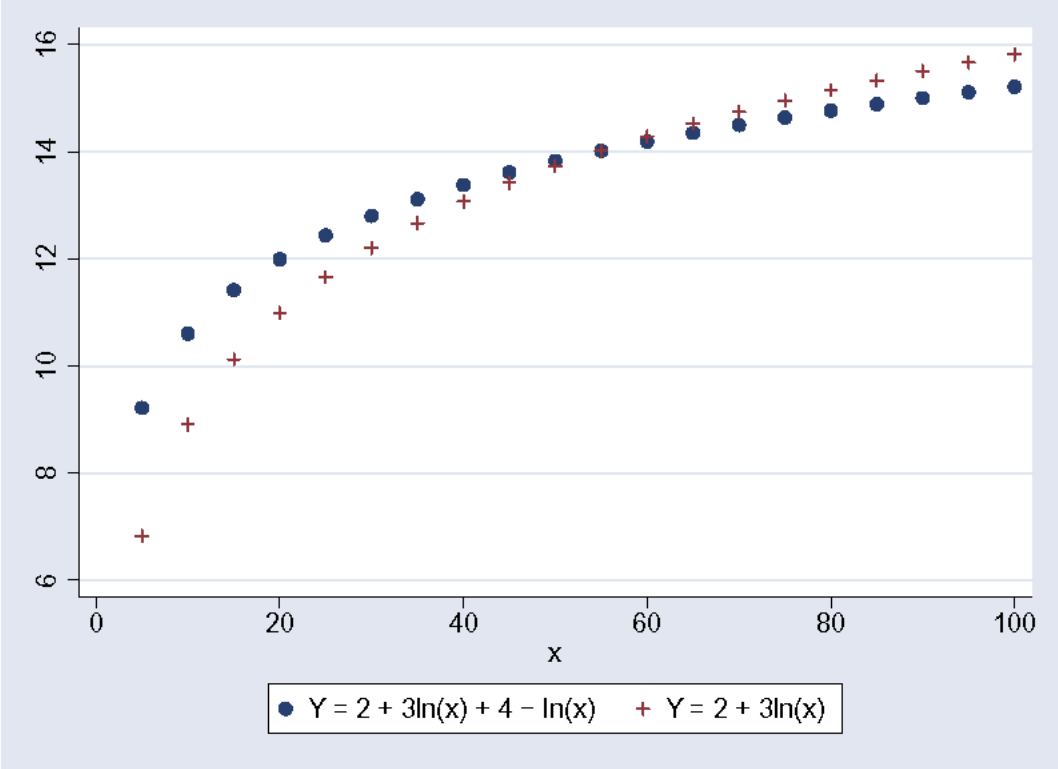
8. (a) and (b)


15
10
5
0
−5
−10
0
20
40
60
80
100
x
Y = 2 + 3ln(x)
Y = 2 − 3ln(x)

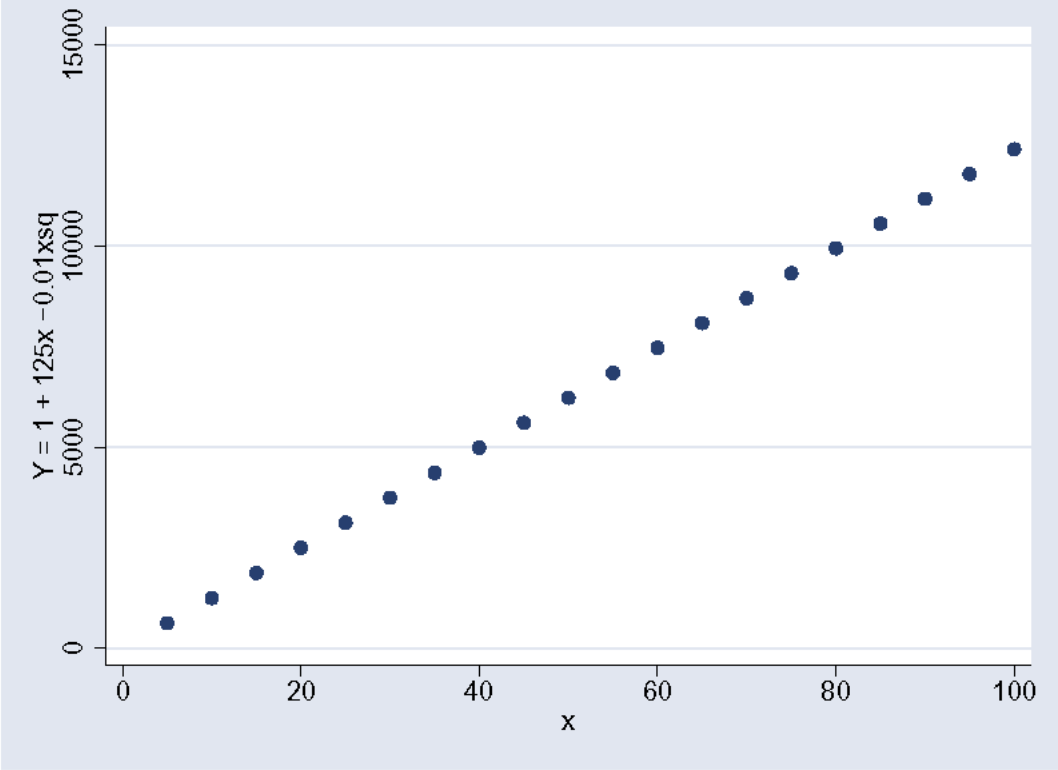
(c)


20
15
10
5
0
20
40
60
80
100
x
Y= 2 + 3ln(x) + 4.0
Y = 2 + 3ln(x)

(d)


16
14
12
10
8
6
0
20
40
60
80
100
x
Y = 2 + 3ln(x) + 4 − ln(x)
Y = 2 + 3ln(x)

(e)


15000
10000
5000
0
Y = 1 + 125x −0.01xsq
0
20
40
60
80
100
x

9. Note that

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \beta_2 X^2 \\ &= \beta_0 + (\beta_1 + 21\beta_2)X + \beta_2(X^2 - 21X). \end{aligned}$$

Define a new independent variable $Z = X^2 - 21X$, and estimate

$$Y = \beta_0 + \gamma X + \beta_2 Z + u_i.$$

The confidence interval is $\hat{\gamma} \pm 1.96 \times \mathrm{SE}(\hat{\gamma})$.

10. (a) $\Delta Y = f(X_1 + \Delta X_1, X_2) - f(X_1, X_2) = \beta_1 \Delta X_1 + \beta_3 \Delta X_1 \times X_2$, so $\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$.

(b) $\Delta Y = f(X_1, X_2 + \Delta X_2) - f(X_1, X_2) = \beta_2 \Delta X_2 + \beta_3 X_1 \times \Delta X_2$, so $\frac{\Delta Y}{\Delta X_2} = \beta_2 + \beta_3 X_1$.

(c)

$$\begin{aligned} \Delta Y &= f(X_1 + \Delta X_1, X_2 + \Delta X_2) - f(X_1, X_2) \\ &= \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2(X_2 + \Delta X_2) + \beta_3(X_1 + \Delta X_1)(X_2 + \Delta X_2) \\ &\quad - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2) \\ &= (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2. \end{aligned}$$