Chapter 16 Regression Models for Nonlinear Relationships

Solutions

1. When , with the linear model, .



With the quadratic model,



When , with the linear model, .



With the quadratic model,



1. When , with the linear model, .



With the quadratic model,



With the cubic model,



When , with the linear model, .



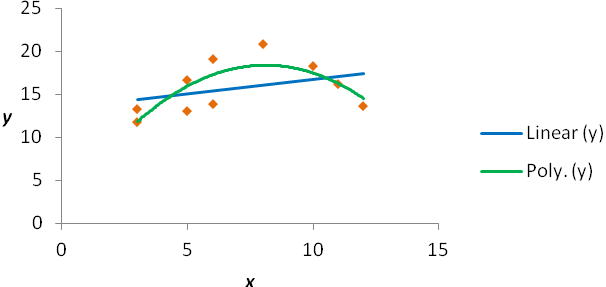
With the quadratic model,



With the cubic model,







The above plot suggests that the quadratic model (green) provides a better fit. The following table shows the relevant output of both models. Note that in order to estimate

the quadratic model, the variable data must all be squared and the model to estimate is:



|  |  |  |
| --- | --- | --- |
| Variable | Linear Model | Quadratic Model |
| Intercept | 13.3087\*  (0.00) | 1.7656  (0.71) |
|  | 0.3392  (0.30) | 4.0966\*  (0.02) |

|  |  |  |
| --- | --- | --- |
|  | NA | ‒0.2528\*  (0.03) |
|  | | |
| Adjusted | 0.0232 | 0.4657 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5 % level. The last row presents adjusted *R*2 for model comparison | | |

* 1. Since the quadratic model has the higher adjusted  (0.4657 > 0.0232), the quadratic model fits the data best. In addition, notice that the variable in the linear model is not significant (*p*-value = 0.30), even at the 10% significance level, whereas both and are significant in the quadratic model.



* 1. For



For



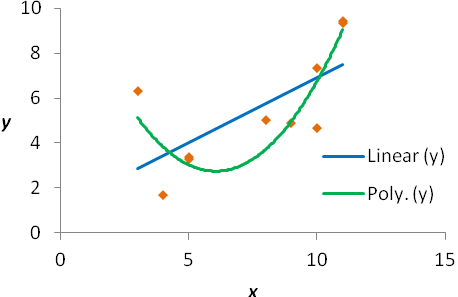
For



* 1. The maximum of the quadratic equation is achieved at units.







The above plot doesn’t give an obvious suggestion on which model is better. The following table shows the relevant results of both models. Note that in order to estimate the quadratic model, the variable data must all be squared and the model to estimate is:



|  |  |  |
| --- | --- | --- |
| Variable | Linear Model | Quadratic Model |

|  |  |  |
| --- | --- | --- |
| Intercept | 1.1006  (0.54) | 12.1338\*  (0.01) |
|  | 0.5828\*  (0.03) | ‒3.1034\*  (0.03) |
|  | NA | 0.2565\*  (0.02) |
|  | | |
| Adjusted | 0.4166 | 0.7843 |

|  |
| --- |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last row presents adjusted *R*2 for model comparison |

* 1. Since the quadratic model has the higher adjusted R2, (0.7843 > 0.4166), the quadratic model fits the data best.
  2. For



For



For



* 1. The minimum of the quadratic equation is achieved at units.



1. 1. Linear Model:

For



For



Quadratic Model:

For



For



Cubic Model:

For



For



* 1. Since the cubic model has the highest adjusted *R2*(0.895 > 0.833 > 0.809), the cubic model is the most appropriate model.

1. 1. Linear Model:

For



For



Quadratic Model:

For



For



Cubic Model:

For



For



* 1. Since the quadratic model has the highest adjusted *R*2 (0.691 > 0.689 > 0.636), the quadratic model is the most appropriate model.

1. 1. In order to estimate the quadratic model, first square the “Hours TV” data. The model to estimate is ε. The regression results are in the table below:



|  |  |  |
| --- | --- | --- |
| Variable | Linear Model | Quadratic Model |
| Intercept | 3.7433\*  (0.00) | 3.0945\*  (0.00) |

|  |  |  |
| --- | --- | --- |
|  | ‒0.0439\*  (0.00) | 0.0410  (0.08) |
|  | NA | ‒0.0022\*  (0.00) |
|  | | |
| Adjusted | 0.6502 | 0.7737 |
| Notes: Parameter estimates are in the top part of the table with the p-values in parentheses; \* represents significance at the 5% level. The last row presents adjusted *R*2 for model comparison | | |

* 1. The quadratic term is justified since the quadratic variable is statistically significant (*p*-value = 0.00) at any reasonable significance level. Further, as shown above, the adjusted *R*2 is higher for the quadratic model than the linear model (0.7737 > 0.6502).



* 1. The maximum GPA is reached when the number of TV hours is hours per week. Therefore, this study suggests that 9.32 is the optimal number of hours of TV for middle school students.







|  |  |  |
| --- | --- | --- |
| Variable | Linear Model | Quadratic Model |
| Intercept | 4.3315\*  (0.00) | 3.8129\*  (0.00) |
|  | ‒1.2623  (0.06) | ‒1.0077\*  (0.00) |

|  |  |  |
| --- | --- | --- |
|  | 0.0875\*  (0.03) | 0.3747\*  (0.00) |
|  | NA | ‒0.0159\*  (0.01) |
|  | | |
| Adjusted | 0.0490 | 0.0959 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last row presents adjusted *R*2 for model comparison. | | |

The estimated quadratic model is



* 1. The quadratic term is justified since the quadratic variable is statistically significant (*p*-value = 0.01) at the 5% significance level. Further, as shown above, the adjusted *R*2 is higher for the quadratic model than the linear model (0.0959 > 0.0490).



* 1. The predicted number of bids for a firm that has a bid premium of 1.2.

With a firm size of $4 billion, .



With a firm size of $8 billion,



With a firm size of $12 billion, .



With a firm size of $16 billion, .



The maximum number of bids is reached when the firm size is billion.







|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Linear Model | Quadratic Model | Restricted model |

|  |  |  |  |
| --- | --- | --- | --- |
| Intercept | 198.9956  (0.22) | ‒264.623  (0.47) | 210.2977\*  (0.03) |
|  | 10.5122\*  (0.00) | 11.4820\*  (0.00) | 10.44\*  (0.00) |
|  | 0.6186  (0.93) | 73.9527  (0.17) | NA |
|  | NA | ‒3.2393  (0.17) | NA |
|  | | |  |

|  |  |  |  |
| --- | --- | --- | --- |
| *SSE* | 96011.69 | 88080.01 | 96045.55 |
| Notes: Parameter estimates are in the top part of the table with the p-values in parentheses; \* represents significance at 5%. The last row presents error sum of squares *SSE*. | | | |

The estimated linear model is



Unemployment is not significant at the 5% significance level (*p*-value = 0.93 > 0.05)

* 1. Given that the response variable is average monthly debt payments, one possible explanation for a quadratic relationship is that as unemployment increases, debt payments may also increase as people make purchases on credit. However, if unemployment increases beyond a point (say due to recession), people may get pessimistic and adjust their expenditures, thus decreasing debt.
  2. The regression results are shown above. The estimated quadratic model is .



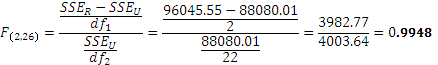
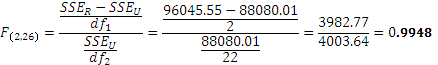
To determine if Unemp and Unemp2 are jointly significant, we must conduct the partial *F* test. The restricted model is . The unrestricted model is . As shown in the above table, *SSER*= 96045.55, *SSEU* = 88080.01.



The competing hypotheses are:



because there are 2 restrictions,



With α = 0.05, , . Therefore, we reject if > 3.44. Since 0.99 < 3.44, we do not reject . At the 5% significance level, we cannot conclude that are jointly significant.



1. Model 1:



Model 2:



Model 3:



Model 4:



1. The slope coefficients are interpreted as:

Model 1: As increases by one unit, decreases by 4.2 units.



Model 2: As increases by one percent, decreases by about 2.8 units .



Model 3: As increases by one unit, decreases by about 4% .



Model 4: As increases by one percent, decreases by about 0.8%.



1. Model 1: When When . Therefore, as increases by 1, decreases by 4.2



Model 2: When When . Therefore, as increases by 1%, decreases by 2.78



Model 3: When When . Therefore, as increases by 1, decreases by3.22 (78.92 ‒ 82.14), or by 3.92%



Model 4: When When . Therefore, as increases by 1%, decreases by 0.59 (74.74 ‒ 75.33), or by 0.78%



1. Model 1:



Model 2:



Model 3:



Model 4:



1. The slope coefficients are interpreted as:

Model 1: As increases by one unit, increases by 4.4 units.



Model 2: As increases by one percent, increases by about 0.23 units.



Model 3: As increases by one unit, increases by about 10 percent.



Model 4: As increases by one percent, increases by about 0.6 percent.



1. Model 1: When When . Therefore, as increases by half of a unit (0.5), increases by 2.2 units .



Model 2: When When . Therefore, as increases by 5%, increases by 1.12



Model 3: When When . Therefore, as increases by 5%, or half of a unit, increases by 5.13%



Model 4: When When . Therefore, as increases by 5%, increases by 2.96%



1. Model 1:



Model 2:



Model 3:



Model 4:



1. Model 1:



Model 2:



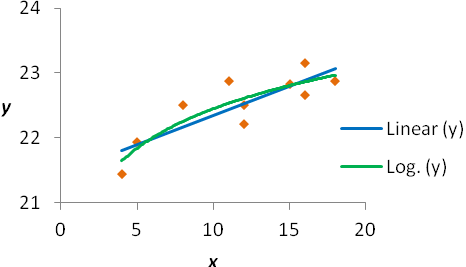
Model 3:



Model 4:







It is difficult to tell by the graph alone whether the line or the logarithmic curve fit the data the best, although the curve shows a slightly better fit. Therefore the logarithmic model should be evaluated with the linear model.

* 1. The 2 models to evaluate are:

Model 1:



Model 2:



In order to estimate the 2nd model, we must log-transform . The model estimates are:



|  |  |  |
| --- | --- | --- |
| Variable | Model 1 | Model 2 |
| Intercept | 21.45\*  (0.00) | 20.42\*  (0.00) |
|  | 0.09\*  (0.00) | NA |
|  | NA | 0.88\*  (0.0008) |
|  | 0.7029 | 0.7723 |

|  |
| --- |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last row presents the computer generated *R*2. |

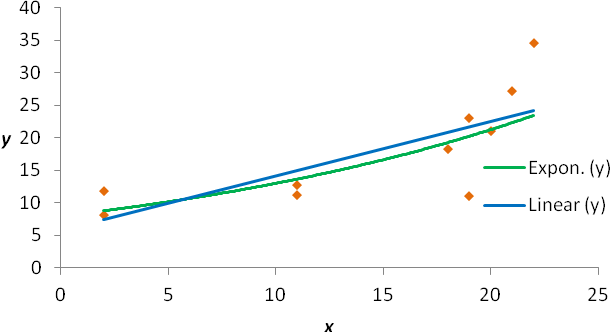
Since Models 1 and 2 are both specified in terms of we can simply use the computer generated to compare them. Model 2 is preferred since it has a higher (0.7723 > 0.7029).



1. Using Model 2, when







The exponential curve seems to be a better fit compared to the linear fit.

* 1. The 2 models to evaluate are:

Model 1:



Model 2:



We must log-transform in order to estimate Model 2.



The model estimates are:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | Response variable: |
| Variable | Model 1 | Model 2 |

|  |  |  |
| --- | --- | --- |
| Intercept | 5.6665  (0.2203) | 2.0695\*  (0.00) |
|  | 0.8433\*  (0.0125) | 0.0493\*  (0.0057) |
|  | | |
|  | 5.9943 | 0.3005 |
|  | 0.5619 | 0.6360 |

|  |
| --- |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last two rows present the computer generated *s*e and *R*2. |

In order to compare the models, we must compute for model 2 in terms of . When running Model 2 regression in Excel, we check *Residuals* to get the predicted value . We transform this value into ; see the table below.



|  |  |
| --- | --- |
|  |  |
| 3.1548 | /2)=24.53 |
| 2.1682 | 9.15 |
| .  .  . | .  .  . |
| 3.0562 | 22.23 |

We use the correl function in Excel to compute Therefore, . Since this is greater than the value for Model 1, (0.6366 > 0.5619), Model 2 provides a better fit, as was suggested by the scatterplot above.



* 1. Using Model 2 with



1. The regression results for both models are:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | Response variable: |

|  |  |  |
| --- | --- | --- |
| Variable | Model 1 | Model 2 |
| Intercept | 66.1852\*  (0.0000) | 4.3277\*  (0.0000) |
|  | ‒0.9931\*  (0.0000) | ‒0.0248\*  (0.0000) |
|  | | |
|  | 4.1702 | 0.1111 |
|  | 0.6260 | 0.5944 |

|  |
| --- |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. |

* 1. For a man with a BMI of 25, , or $41,358.



For a man with a BMI of 30, , or $36,392.



* 1. For a man with a BMI of 25, or $41,012.



For a man with a BMI of 30, or $36,229.



Since the exponential model (Model 2) has a response variable in terms of ln(, we must compute *R*2 in terms of *y* forcomparison with Model 1. When running Model 2 in Excel, we check *Residuals* to get the predicted value . We transform this value into ; see the table below.



|  |  |
| --- | --- |
|  |  |
| 3.5108 | /2)=33.68 |
| 3.6841 | 40.05 |
| 3.8078 | 45.33 |

We use the correl function in Excel to compute . Therefore, . We compare this with the computer generated value for the linear model to conclude that the exponential model (Model 2) is better since



0.6358 > 0.6260. Note that the difference between the two models is minimal, which also explains why their predicted values are similar.

1. The regression results for both models are:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | Response variable: |
| Variable | Model 1 | Model 2 |

|  |  |  |
| --- | --- | --- |
| Intercept | ‒4.2174\* (0.0035) | ‒3.1716\*  (0.0001) |
|  | 0.3023\*  (0.0004) | 0.1903\*  (0.0000) |
|  | 0.0010\*  (0.0105) | 0.0006\*  (0.0002) |
|  | ‒0.0016\*  (0.0197) | ‒0.0010\*  (0.0049) |

|  |  |  |
| --- | --- | --- |
|  | 0.0123\*  (0.0346) | 0.0081\*  (0.0120) |
|  | | |
|  | 0.2188 | 0.1176 |
|  | 0.6692 | 0.7356 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last two rows present the computer generated *s*e and *R*2. | | |

* 1. For : .



* 1. For :.



* 1. Since the exponential model (Model 2) has a response variable in terms of ln(, we must compute *R*2 in terms of *y* to compare with Model 1. When running Model 2



in Excel, we check *Residuals* to get the predicted value . We transform this value into ; see the table below.



|  |  |
| --- | --- |
|  |  |
| 0.5087 | 1.6747 |
| 0.5600 | 1.7628 |
| .  .  . | .  .  . |
| 0.1467 | 1.1661 |

We use the correl function in Excel to compute . Therefore, . We compare this with the computer generated results for



the linear model to conclude that the exponential model, Model 2, is better since 0.7569 > 0.6692.

* 1. The regression results are below:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | Response variable: |
| Variable | Model 1 | Model 2 |

|  |  |  |
| --- | --- | --- |
| Intercept | 65.8255  (0.8076) | 6.2925\*  (0.0000) |
|  | 237.8506  (0.2346) | 0.1326  (0.3585) |
|  | 389.3673\*  (0.0010) | 0.2191\*  (0.0077) |
|  | 0.1831  (0.7651) | 0.0002  (0.5789) |
|  | | |

|  |  |  |
| --- | --- | --- |
|  | 284.6173 | 0.2067 |
|  | 0.7936 | 0.7408 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last two rows present the computer generated *s*e and *R*2. | | |

* 1. For the linear model with



For the exponential model with .



* 1. Since the exponential model (Model 2) has a response variable in terms of ln(, we must compute *R*2 in terms of *y* forcomparison with Model 1. When running Model 2 in Excel, we check *Residuals* to get the predicted value . We transform this value into ; see the table below.

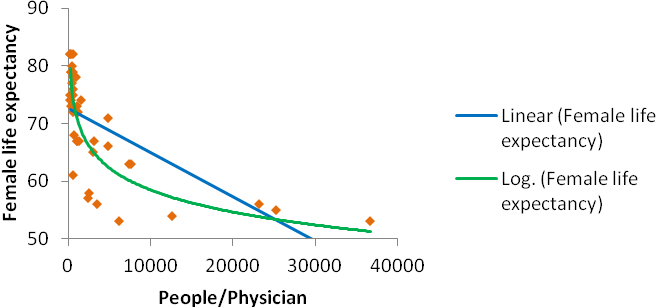


|  |  |
| --- | --- |
|  |  |
| 8.0589 | 3230.114 |
| 7.6265 | 2096.098 |
| .  .  . | .  .  . |
| 7.0070 | 1128.231 |

We use the correl function in Excel to compute . Therefore, . We compare this with the computer generated results for the linear model to conclude that the exponential model, Model 2, is better since 0.8102 > 0.7936.







The logarithmic model appears to have a better fit than the linear model.

* 1. The regression results for both the linear and logarithmic models are below:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | |
| Variable | Model 1 | Model 2 |
| Intercept | 72.6441\*  (0.0000) | 109.4952\*  (0.0000) |
|  | ‒0.0008\*  (0.0000) | NA |

|  |  |  |
| --- | --- | --- |
|  | NA | ‒5.5414\*  (0.0000) |
|  | | |
|  | 0.4126 | 0.6967 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5percent level. The last row presents the computer generated *R*2. | | |

When the people-to-physician ratio is 1000, female life expectancy is 71.84.



When the people-to-physician ratio is 500, 72.24.



Therefore, as decreases by 500, increases by 72.24 – 71.84 = 0.40 year.



* 1. When the people-to-physician ratio is 1000, female life expectancy is



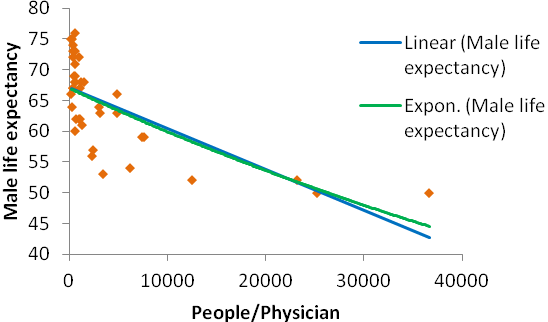
When the people-to-physician ratio is 500, .



Therefore, as decreases by 500, increases by 75.06 – 71.22 = 3.84 years.



* 1. Since both models have the same response variable, we can compare the computer generated *R*2 values directly. Since 0.6967 > 0.4126, the logarithmic model provides a better fit.



Although it is more difficult to tell from this plot, it looks like the exponential curve is more appropriate than the line.

* 1. The regression results for both the linear and logarithmic models are below:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | |
| Variable | Model 1 | Model 2 |
| Intercept | 67.1605\*  (0.0000) | 96.2515\*  (0.0000) |
|  | ‒0.0006\*  (0.0000) | NA |

|  |  |  |
| --- | --- | --- |
|  | NA | ‒4.4075\*  (0.0000) |
|  | | |
|  | 5.4600 | 4.3079 |
|  | 0.4726 | 0.6717 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last row presents the computer generated *R*2. | | |

When the people-to-physician ratio is 1000, male life expectancy is 66.56.



When the people-to-physician ratio is 500, 66.86.



Therefore, as decreases by 500, increases by year.



* 1. When the people-to-physician ration is 1000, .



When the people-to-physician ration is 500, .



Therefore, as decreases by 500, increases by 68.86 – 65.81 = 3.05 years.



* 1. Since both models have the same response variable, we can compare the computer generated values directly. Since 0.6717 > 0.4726, the logarithmic model provides a better fit.



* 1. Since life expectancy increases proportionately more for females with a decrease in the people-to-physician ratio, females are more likely to benefit from an increase in physicians in the population.

1. The relevant regression output is:

|  |  |
| --- | --- |
| Variable | Response Variable: |
| Intercept | ‒0.1741  (0.1526) |
|  | 0.7089\*  (0.1068) |
|  | 0.3431\*  (0.1322) |

|  |
| --- |
| Notes: Parameter estimates are in the top part of the table with the standard errors in parentheses; \* represents significance at 5%. |

* 1. For the above log-log model, if *L* increases by 1%, *Q* is predicted to change by about 0.7089% , holding capital constant.



* 1. In order to test this hypothesis, we conduct a *t* test. If a 1% increase in labor increases output by more than 0.5%, then would have to be greater than 0.5. Therefore, the competing hypotheses are:



(Note: is taken from Excel’s output.) The critical value is . Thus, we reject if . Since 1.956 > 1.679,



we reject At the 5% significance level, we can conclude that a 1% increase in labor will increase output by more than 0.5%, holding capital constant.



1. The regression results for all three models are:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Response Variable: | Response variable: | Response Variable: |
| Variable | Model 1 | Model 2 | Model 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| Intercept | 32.7687  (0.1250) | 4.8008  (0.1392) | 5.0960  (0.8888) |
|  | ‒0.8265  (0.0698) | ‒0.1201  (0.0825) | ‒0.9251  (0.0504) |
|  | 0.7850\*  (0.0039) | 0.1009\*  (0.0128) | 0.8200\*  (0.0031) |
|  | 0.3862  (0.1385) | 0.0825\*  (0.0411) | 2.6280  (0.2838) |

|  |  |  |  |
| --- | --- | --- | --- |
|  | NA | NA | ‒0.0371  (0.3565) |
|  | | | |
|  | 5.4989 | 0.8637 | 5.5108 |
|  | 0.3931 | 0.3523 | 0.4122 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last two rows present the computer generated *s*e and *R*2. | | | |

* 1. The estimated linear model (Model 1) is . For a one unit increase in Pass Completion Rate, Salary is predicted to decrease by $0.83 million, or $830,000, holding everything else constant. For a one unit increase in touch-downs scored, Salary is predicted to increase by 0.785 million, or $785,000, holding everything else constant. And for an increase in age by one year, salary is predicted to increase by 0.39 million, or $390,000.



Comparing the linear model with the exponential model (Model 2), we must compute for model 2 in terms of *y*. When running Model 2 in Excel, we check *Residuals* to get the predicted value . We transform this value into . Using Excel’s correl function, we compute Therefore, = 0.51472 = 0.2649. Therefore, since Model 1 has an *R*2 of 0.3931 > 0.2649, Model 1 is better.



* 1. The regression results are reported in the table above as Model 3. The hypotheses for the partial *F* test are:

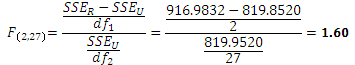
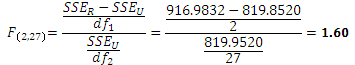


Model 3 is the restricted model and the restricted model is . In order to implement the test, we use the error sum of squares of the restricted and the unrestricted models as:



*SSER*= 916.9832, *SSEU* = 819.952 (both of these are reported in the regression results from Excel, in the ANOVA table).

because there are 2 restrictions,



With α = 0.05, , . Therefore, we reject *H*0 if > 3.35. Since 1.60 < 3.35, we do not to reject *H*0.



At the 5% significance level, we cannot conclude that are jointly significant.



1. 1. The regression results for both models are below:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | Response variable: |
| Variable | Model 1 | Model 2 |

|  |  |  |
| --- | --- | --- |
| Intercept | 153348.27\*  (0.0114) | 12.4137\*  (0.0000) |
|  | 95.86\*  (0.0108) | 0.0002\*  (0.0057) |
|  | 556.89  (0.9783) | 0.0002  (0.9955) |
|  | 92022.91\*  (0.0009) | 0.1686\*  (0.0012) |

|  |  |  |
| --- | --- | --- |
|  | | |
|  | 74984.98 | 0.1419 |
|  | 0.7237 | 0.7331 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last two rows present the computer generated *s*e and *R*2. | | |

Model 1:

* a house’s square footage increases by one square foot, the price of the home increases by $95.86.



* For each extra bedroom as house has, the price of the home increases by $556.89.
* or each extra bathroom a house has, the price of the home increases by $92,022.91.



Model 2:

* s a house’s square footage increases by one square foot, the price of the home increases by 0.02% (= 0.0002 100).



* or each extra bedroom as house has, the price of the home increases by 0.02% (= 0.0002 100).



* or each extra bathroom a house has, the price of the home increases by 16.86% (= 0.1686 100).



* 1. Comparing the linear model with the exponential model (Model 2), we must compute *R*2 for model 2 in terms of *y*. When running Model 2 in Excel, we check *Residuals* to get the predicted value . We transform this value into . Using Excel’s correl function, we compute to get   
      = 0.84272 = 0.7101. Therefore, we can conclude that Model 1 is the preferred Model since it has a higherof 0.7237.



1. 1. The regression results for both models are in the following table:

|  |  |  |
| --- | --- | --- |
|  | Response variable: | |
| Variable | Model 1 | Model 2 |
| Intercept | 27.4116\*  (0.0000) | 87.8631\*  (0.0017) |

|  |  |  |
| --- | --- | --- |
|  | ‒0.3547\*  (0.0012) | ‒0.3484\*  (0.0014) |
|  | ‒0.0001\*  (0.0272) | NA |
|  | NA | ‒6.1512\*  (0.0210) |
|  | | |
|  | 0.4613 | 0.4663 |

|  |
| --- |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5percent level. The last row presents the computer generated *R*2. |

* 1. Since Models 1 and 2 have the same response variable , we can directly compare them. Given that Model 2 has the higher , the logarithmic model (Model 2), is a better fit.



1. 1. The regression results for both models are:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | Response variable: |
| Variable | Model 1 | Model 2 |

|  |  |  |
| --- | --- | --- |
| Intercept | ‒40.8632\*  (0.0000) | ‒112.0910\*  (0.0000) |
|  | 0.0041\*  (0.0000) | NA |
|  | NA | 12.2033\*  (0.0000) |
|  | | |
|  | 1.0032 | 0.2446 |

|  |  |  |
| --- | --- | --- |
|  | 0.7001 | 0.7430 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last two rows present the computer generated *s*e and *R*2. | | |

* 1. For Model 1, for a one unit increase in income, the savings rate is predicted to increase by 0.0041%.

For Model 2, for a one percent increase in Income, the savings rate is predicted to increase by 12.20%.

* 1. Comparing the linear model with the log-log model (Model 2), we must compute for model 2 in terms of *y*. When running Model 2 in Excel, we check *Residuals* to get the predicted value . We transform this value into . Using Excel’s correl function, we compute to get = 0.80612 = 0.6498. Therefore, one can conclude that Model 1 is the preferred model with a higher of 0.7001.



**Case Study 16.1**

1. We estimate the following two models:

Model 1:



Model 2:



The regression output is summarized in the following table:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | |
| Variable | Model 1 | Model 2 |
| Intercept | 8.0669\*  (0.0000) | ‒25.4418\*  (0.0000) |
| Adj ROA | 0.5985  (0.1633) | 1.2937\*  (0.0012) |
| Adj Return | 0.4921  (0.1640) | 0.5923  (0.0668) |

|  |  |  |
| --- | --- | --- |
| Total Assets | 0.0000\*  (0.0000) | NA |
|  | NA | 3.6666\*  (0.0000) |
|  | 0.1640 | 0.3035 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last row presents the computer generated *R*2. | | |

1. Since Models 1 and 2 have the same response variable, we use the computer generated to infer that Model 2 provides a better fit relative to Model 1 (0.3035 > 0.1640). Model 2 is estimated as:



.



We compute the average values of each explanatory variable as: , . For these average values, the predicted compensation is



.



**Case Study 16.2**

1. The relevant portion of the regression output for the quadratic model is presented in the second column of the following table.

|  |  |  |
| --- | --- | --- |
| Variable | Quadratic Model | Linear Model |

|  |  |  |
| --- | --- | --- |
| Intercept | ‒5609.37  (0.08) | ‒4769.40  (0.13) |
|  | 3.79  (0.76) | 4.76  (0.70) |
|  | 81.95\*  (0.00) | 80.44\*  (0.00) |
|  | 897.27\*  (0.00) | 539.67\*  (0.00) |

|  |  |  |
| --- | --- | --- |
|  | ‒22.55  (0.11) | NA |
|  | | |
| Adjusted *R*2 | 0.5749 | 0.5700 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. Adjusted *R*2 is provided for model comparison. | | |

1. The variable, with a *p*-value of 0.11, is not statistically significant at the 5% (or 10%) level. However, based on adjusted *R*2, the quadratic model is a better fit than the linear model (0.5749 > 0.5700). Overall, we can be indifferent between the two models.

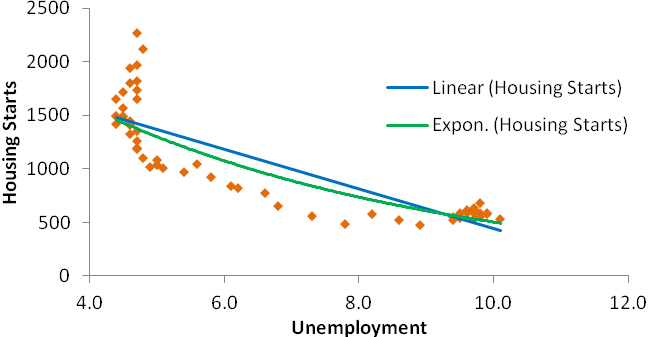
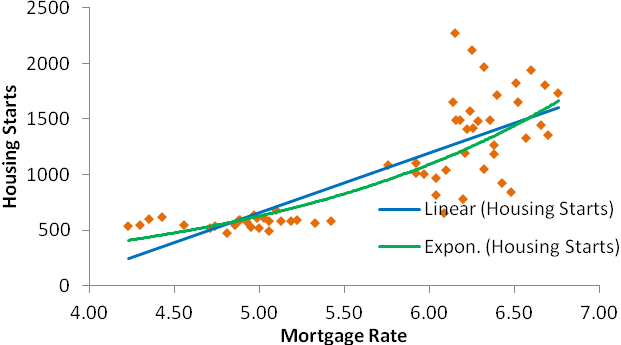


We use the estimated quadratic model to determine the optimal level of experience at which the salary is maximized. We obtain this value by solving , or approximately 20 years. Therefore, salaries increase with experience for up to 20 years before



they begin to fall. Since we have only one player in the sample with 20 years of experience, it shows why the quadratic model is not really an improvement over the linear model.

**Case Study 16.3**



The above scatterplots imply that the relationship of housing starts with mortgage rate and unemployment rate is better captured by a curve than a line.

1. Consider two models:

Model 1:



Model 2:



The estimated models are summarized below:

|  |  |  |
| --- | --- | --- |
|  | Response Variable: | Response variable: |

|  |  |  |
| --- | --- | --- |
| Variable | Model 1 | Model 2 |
| Intercept | 1539.4262  (0.1131) | 7.1573\*  (0.0000) |
|  | 98.0221  (0.4272) | 0.1242  (0.1803) |
|  | ‒156.1995\*  (0.0003) | ‒0.1524\*  (0.0000) |
|  | | |
|  | 271.6402 | 0.2028 |

|  |  |  |
| --- | --- | --- |
|  | 0.7191 | 0.8278 |
| Notes: Parameter estimates are in the top part of the table with the *p*-values in parentheses; \* represents significance at the 5% level. The last two rows present the computer generated *s*e and *R*2. | | |

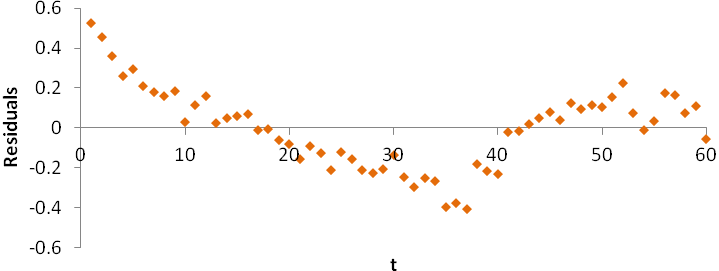
Comparing the linear model with the exponential model, we must compute for Model 2 in terms of *y*. When running Model 2 in Excel, we check *Residuals* to get the predicted value . We transform this value into . Using Excel’s correl function, we compute to get *R*2 We can compare this value with the computer-



generated value for Model 1. Since 0.7653 > 0.7191, we conclude that Model 2 is a better model for making predictions.

Serial Correlation is often a problem in time series studies. With serial correlation, the estimators are unbiased, but they are not efficient. Moreover, the standard errors are biased making the regular *t* and *F* tests invalid.

We can plot the residuals against time to look for serial correlation. The scatterplot is shown below:



The scatterplot of residuals against time shows a definite pattern of the residuals over time, suggesting a positive serial correlation. A common correction for the standard errors is the Newey-West procedure.